Développement de méthodes de tatouage sûres pour le traçage de contenus multimédia

Benjamin Mathon

UNIVERSITÉ DE GRENOBLE
UNIVERSITÉ CATHOLIQUE DE LOUVAIN

THÈSE EN COTUTELLE INTERNATIONALE

Pour obtenir les grades de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE et DOCTEUR DE L'UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Spécialités : **Signal, Image, Parole, Telecoms** et **Sciences de l'ingénieur**

Arrêté ministérial français : 7 août 2007

Présentée par

Benjamin MATHON

Thèse dirigée par **M. Patrick Bas** et **M. Benoît Macq**
et codirigée par **M. François Cayre**

préparée au sein du laboratoire
Grenoble Image Parole Signal Automatique
dans le cadre de l'école doctorale **E.E.A.T.S.**
et au laboratoire de **Télécommunications et Télédétection**

Développement de méthodes de tatouage sûres pour le traçage de contenus multimédia

Soutenance prévue le **07/07/2011**,
devant le jury composé de :

M. Jean-Marc Chassery

Directeur de Recherche du CNRS, Président

M. Teddy Furon

Chargé de Recherche de l'INRIA, Rapporteur

M. Luc Vandendorpe

Professeur de l'UCL, Rapporteur

M. Sviatoslav Voloshynovskiy

Professeur de l'Université de Genève, Rapporteur

M. Jean-Jacques Quisquater

Professeur de l'UCL, Examinateur

M. Patrick Bas

Chargé de Recherche du CNRS, Directeur de thèse

M. Benoît Macq

Professeur de l'UCL, Directeur de thèse

M. François Cayre

Maître de Conférences de Grenoble INP, Co-Directeur de thèse

Notre liberté est menacée par le besoin de sécurité
et la sécurité elle-même est menacée par le souci obsédant qu'on en a.
Norbert Bensaïd, *La Lumière médicale*

Remerciements

Cette thèse a été supportée par une allocation doctorale de recherche de la région Rhône-Alpes ainsi qu'une allocation ministérielle française de recherche MENRT. Les travaux réalisés pendant l'ensemble du doctorat ont été supportés par les projets NEBBIANO ANR-06-SETIN-009, ESTIVALE ANR-05-RIAM-01903, ARA SSIA TSAR et BCRYPT IAP (PAI) phase-VI.

L'ensemble des travaux réalisés lors de mon doctorat, y compris la création de ce manuscrit, résulte de l'utilisation de logiciels libres (sans exception) sous environnements Linux et MacOSX. En voici une liste non exhaustive : dvips, gcc, Grace, Inkscape, LaTeX, LibreOffice, LyX, make, MEncoder, MPlayer, ninja, pdfcrop, Perl, pgfplots, Processing, ps2pdf, Rhythmbox, VIM, xfig, xpdf.

Table des matières

Introduction

La sécurité de nos jours

L'Oxford English Dictionary définit la sécurité comme l'état d'être exempt de danger ou de menace. La sécurité est malheureusement devenue à la mode aujourd'hui et différents facteurs peuvent en être la cause : tensions géopolitiques et lutte contre le terrorisme, tensions sociales et sur-médiatisation. La sécurité est avant tout définie par la résistance d'un système face à un adversaire (ou groupement d'adversaires) cherchant à mettre à mal ce système. Il faut bien différencier les notions de sécurité et de sûreté. En effet, la sûreté est plutôt la défense d'un système face à des attaques non intentionnelles contrairement à la sécurité où l'attaque est humaine et délibérée. L'exemple récent des graves accidents survenus à la centrale nucléaire de Fukushima Dai-Ichi est un problème de sûreté car c'est un séisme qui a remis en cause la robustesse du système. Au contraire, les attentats survenus aux États-Unis d'Amérique le 11 septembre 2001 sont un exemple d'attaque de sécurité car ceux-ci ont été organisés et ciblés. La sécurité étant liée à la présence d'un adversaire, il est alors important pour le concepteur d'un système de connaître la nature même de la personne ou des personnes contre lesquelles il cherche à se défendre. Ce dernier se posera alors les questions suivantes :

- Combien de personnes cherchent à mettre à mal le système ?
- Quels sont leurs moyens (matériels, financiers, politiques) ?
- Quel est leur accès au système ?

Le concepteur peut tout d'abord chercher à être résistant au pire des adversaires, c'est-à-dire celui ayant le plus de moyens et une grande facilité d'accès au système. Toutefois, un système pouvant résister à ces attaques va être difficile à concevoir, ce pour des raisons d'ordre budgétaire ou technologique. De plus, la sécurité ne doit pas nuire à l'utilité même du système (celle pour laquelle il a été initialement prévu) ni à sa robustesse (sûreté). Prenons l'exemple de la construction d'un pont, une sécurité absolue (résistance aux attaques terroristes) nécessitera beaucoup de moyens matériels et/ou humains et gênera l'utilisation classique de ce pont par un civil. De plus, sécuriser un pont signifiera un réel contrôle des personnes entrantes et sortantes mais ne se souciera pas d'une catastrophe naturelle pouvant endommager ce pont. Ce jeu entre sécurité et robustesse apparaît également dans les communications où l'envoi d'un message d'un expéditeur à un destinataire peut soulever certains problèmes.

La sécurité dans les communications

La problématique liée à la sécurisation des communications répondait principalement à des besoins militaires et commerciaux. En effet, lorsque deux personnes géographiquement distantes veulent échanger une information confidentielle (comme un ordre martial), plusieurs questions se posent :

1. Comment être sûr qu'une personne mal intentionnée n'a pas eu accès au message pendant son transfert ?
2. Comment être sûr que le message n'a pas été modifié ?

3. Comment être sûr que la personne avec laquelle on communique est bien la bonne ?
4. Comment envoyer un message sans que l'adversaire puisse seulement se rendre compte qu'il y a eu transmission ?
5. Comment s'assurer de la persistance du message dans le canal de transmission utilisé ?

Afin d'illustrer ces différents scénarios, les cryptographes ont inventé un jeu de la sécurité qui a lieu entre trois personnages fictifs : Alice cherchera à transmettre une information à Bob en cherchant à satisfaire un des points cités plus haut et l'adversaire Ève essaiera de contourner cet objectif. Comme nous allons le voir par la suite, la *cryptographie* permet de répondre aux trois premières questions, la *stéganographie* à la quatrième, et le *tatouage*, discipline faisant partie intégrante de cette thèse, à la cinquième. Afin de mieux appréhender les contraintes propres au tatouage, nous présentons dans un premier temps celles qui sont liées à la cryptographie et celles qui sont liées à la stéganographie.

Chiffrement de messages

La *cryptographie* est une discipline permettant d'assurer que les trois premiers points énoncés ci-dessus (respectivement confidentialité, intégrité et authenticité) sont respectés lors de communications par le biais d'algorithmes de chiffrement (on parle alors de système cryptographique) alors que la *cryptanalyse* désigne l'ensemble des outils permettant de casser un tel système.

Il existe plusieurs techniques de chiffrement dans la littérature, par exemple le chiffrement de César, méthode consistant à remplacer chaque lettre du message par une autre lettre à distance fixe, la longueur du décalage est alors appelée clé secrète : Alice et Bob partagent ce secret et peuvent ainsi s'envoyer des messages de façon confidentielle car Ève, ne connaissant pas le décalage, ne peut alors comprendre les informations envoyées.

Afin de rationaliser les hypothèses nécessaires pour l'analyse de la sécurité des méthodes de chiffrement, Kerckhoffs a émis la proposition suivante [47] :

Il faut qu'il n'exige pas le secret, et qu'il puisse sans inconvénient tomber entre les mains de l'ennemi.

Cela signifie que l'algorithme de chiffrement ne doit pas être une donnée secrète, c'est-à-dire que n'importe quelle personne peut découvrir le fonctionnement de la machine ou l'algorithme de chiffrement (donnée publique). De plus Kerckhoffs ajoute :

La clef doit pouvoir en être communiquée et retenue sans le secours de notes écrites, et être changée ou modifiée au gré des correspondants.

Ce point signifie que l'unique secret partagé par Alice et Bob doit être une donnée facilement insérable dans l'algorithme ou l'appareil servant au chiffrement.

Le principe de Kerckhoffs est devenu un principe fondamental de la sécurité s'opposant au principe de "sécurité par l'obscurité" pour lequel celle-ci repose uniquement sur le secret lié à la technique de chiffrement. La sécurité par l'obscurité est encore appliquée de nos jours mais est considérée comme risquée : en effet, il suffit qu'une personne mal intentionnée dévoile le secret de l'algorithme pour que toute la sécurité s'effondre alors qu'une clé secrète est facilement modifiable si celle-ci est dévoilée au grand jour. Pour éviter les attaques de force brute (attaque consistant à tester toutes les clés possibles), beaucoup de systèmes de chiffrement actuels se reposent sur des propriétés mathématiques, c'est le cas du système RSA (Rivest Shamir Adleman [73]) où la

sécurité est ici calculatoire : si la durée de validité des données confidentielles est inférieure au temps nécessaire par plusieurs ordinateurs pour casser le système, on considère alors que la sécurité de celui-ci est acceptable en pratique. Cependant il ne s'agit pas d'une sécurité absolue car dépendante de la puissance actuelle des calculateurs.

Attention toutefois : l'utilisation d'une clé secrète ne garantit pas forcement une sécurité absolue du système ! Reprenons notre chiffrement de César. Ce dernier est peu sûr. En effet, deux techniques de cryptanalyse sont possibles : la première est une attaque de force brute qui consiste à essayer tous les décalages possibles (au nombre de 26 dans notre exemple) jusqu'à trouver un message cohérent. La deuxième technique repose sur la statistique des apparitions d'une lettre dans un texte d'une langue courante. Dans la langue française par exemple, la lettre "E" est beaucoup plus présente que n'importe quelle autre. Il suffit alors pour Ève de remplacer la lettre qui apparait le plus de fois dans le texte chiffré par la lettre "E" et de procéder ainsi sur les autres lettres en fonction des fréquences d'apparition. Comme nous le verrons dans le domaine du tatouage, l'analyse de la sécurité d'un système peut passer par le développement de méthodes permettant d'estimer la clé secrète utilisée par ce système.

Du chiffrement à l'imperceptibilité

La cryptographie permet, grâce à des méthodes de chiffrement, de permettre une certaine sécurité pour l'envoi de messages entre deux correspondants. Cependant, il n'est pas question ici de messages "cachés". C'est-à-dire qu'un adversaire, même s'il ne peut ni lire ni modifier le message transmis, ni même se faire passer pour un des deux correspondants, pourra toujours déceler la présence d'un envoi entre deux destinataires. Si la cryptographie n'a pas pour but de cacher l'existence d'une communication entre deux individus ou entités, c'est le cas de la *stéganographie* [30]. La stéganographie est l'art de camoufler une information (message secret) dans une autre. L'information hôte peut ici désigner un autre message ou un support (contenu image, audio, vidéo). Contrairement à la cryptographie qui repose sur une écriture secrète des données, la stéganographie repose sur une écriture discrète.

Respectant le principe de Kerckhoffs, toute personne n'ayant pas la clé secrète utilisée pour le camouflage des données ne pourra alors déceler la présence de ces données. Un exemple célèbre de stéganographie dans la littérature est la lettre envoyée par George Sand à Alfred de Musset, lettre d'amour emplie de poésie dont le style devient beaucoup plus cru lorsque le lecteur ne regarde qu'une ligne sur deux [1]. Nous pouvons aussi citer, dans un contexte plus récent, une lettre du gouverneur de Californie, le célèbre Arnold Schwarzenegger, adressée aux députés de son État. Dans cette lettre, où il s'oppose à une loi visant à étendre les pouvoirs financiers du port de San Francisco, nous pouvons clairement nous questionner sur le message qu'il a réellement voulu transmettre à l'assemblée en assemblant les premières lettres de chaque ligne de cette lettre [2]. Cette technique de stéganographie est appelée acrostiche.

Dans ce manuscrit de thèse, nous nous intéressons particulièrement à une discipline voisine de la stéganographie : le *tatouage*. Le tatouage qui, comme la stéganographie, est une technique permettant d'insérer une information de manière discrète dans un support. Cependant, là où la stéganographie prône une discrétion de l'information cachée au niveau statistique (une personne non autorisée n'est pas capable de déceler la présence

1. Ironiquement, l'authenticité de cette lettre est remise en cause : il s'agirait d'un canular remontant au dernier quart du XIXe siècle http://www.amisdegeorgesand.info/
2. http://www.huffingtonpost.com/2009/10/27/schwarzenegger-sends-lawm_n_336319.html

de cette information), l'art du tatouage réside dans une discrétion imperceptible (non visible à l'œil nu pour des supports photos ou vidéos et non audible pour des sons). De plus, le but du tatouage ne sera pas de cacher l'existence d'une information cachée mais plutôt de protéger son insertion contre des manipulations du support. Une autre différence entre les deux disciplines est la dépendance qui existe entre le support et le message à insérer : en stéganographie, le support est utilisé uniquement comme un canal permettant de transmettre une information de manière confidentielle et il n'y a a priori aucune relation entre l'information et ce canal. Au contraire, en tatouage, la nature de l'information est liée au contenu et devra être insérée de manière robuste. L'information cachée peut alors nous permettre de répondre aux questions suivantes : le contenu est-il authentique ? Qui en est l'auteur ? Le filigrane des billets de banque est un très bon exemple de tatouage car sa présence permet de vérifier qu'il ne s'agit pas de fausse monnaie circulant à l'intérieur d'un pays. En effet, le filigrane est conçu de manière à ne pas pouvoir être facilement reproduit. Avec l'arrivée de l'informatique et des supports numériques, les techniques de tatouage au niveau numérique peuvent permettent de protéger les droits liées aux œuvres.

L'essor de l'informatique

Si la cryptographie, la stéganographie et le tatouage ont longtemps été décrits comme des arts ne concernant qu'une infime partie de la population, les solutions qu'ils apportent (plus particulièrement pour la cryptographie et le tatouage) sont devenus pertinents pour beaucoup d'individus depuis l'arrivée de l'informatique.

La science de l'informatique, définie comme l'ensemble des moyens techniques permettant d'automatiser le traitement des informations, a pris une part de plus en plus importante dans notre société, surtout grâce à l'avènement d'Internet, réseau grâce auquel n'importe quel utilisateur peut envoyer et recevoir du courrier (emails), gérer son argent (banques en ligne), exprimer une opinion (politique, artistique, informative) sur les blogs et les réseaux sociaux, directement de chez lui à l'aide de son ordinateur personnel. La sécurité des informations échangées est alors devenue un problème quotidien qui n'est plus réservé aux militaires et commerciaux voulant protéger leurs données confidentielles. La construction de systèmes d'échange de données sûrs est alors devenue une nécessité. Avec l'informatique est aussi venue la dématérialisation des contenus. La plupart des supports multimédia utilisés actuellement comme la musique, la vidéo et plus récemment le livre s'utilisent de plus en plus de façon numérique (achats de sons au format *MP3*, films disponibles en vidéo à la demande, textes d'auteurs disponibles sur des librairies en ligne).

Cependant, la numérisation des contenus a entrainé le non-respect des lois liées à la propriété intellectuelle dans la mesure où il devient facile de dupliquer un support numérique et ce, sans perte de qualité significative. Internet et les réseaux de partage ont facilité ces échanges qui peuvent donc ne pas respecter nécessairement le droit d'auteur. Les DRM (*Digital Rights Management*) proposent des techniques afin de lutter contre la piraterie des œuvres numériques. Ceux-ci permettent de limiter le nombre de copies qu'un utilisateur peut faire de son contenu acheté sur une plateforme légale. Le fonctionnement global est le suivant : le contenu est chiffré par le serveur de contenus à l'aide d'une clé (identifiant) avant l'envoi à l'utilisateur. Le lecteur de contenus de l'utilisateur télécharge alors une licence après avoir vérifié que l'utilisateur a effectivement payé les droits. La licence contient alors la clé secrète nécessaire à la lecture du contenu. Parmi les sociétés utilisant des DRM sur leur plateforme de vente de supports culturels, nous citerons par exemple *Apple*, *Microsoft* et *Adobe*. L'utilisation des DRM sur des supports multimédia est cependant sujette à controverse. En effet, chaque plateforme de contenus utilise son propre système de DRM, ceux-ci sont souvent non inter-opérables. De plus, la valeur d'un contenu acheté dans un format propriétaire est discutable si ce format n'est pas lisible par les différents lecteurs de l'utilisateur. Ce dernier se sentira alors frustré de ne pas pouvoir le lire dans sa voiture ou sur son disque dur multimédia bon

marché. Il pourra alors préférer passer par une voie illégale pour ne pas avoir de souci quant à l'utilisation de son contenu.

L'utilisation de techniques de tatouage numérique, permettant d'insérer directement des informations dans le support multimédia [25, 78], est une solution pouvant empêcher la piraterie des œuvres numériques. Le message à insérer peut alors contenir des informations liées à l'utilisateur afin de repérer d'où provient la fuite si le contenu est échangé, le contenu est alors dit estampillé. C'est le cas de l'*iTunes Store* pour certains morceaux musicaux [31]. Le tatouage peut aussi être une marque destinée à limiter les usages du contenu. La technologie *Cinavia* [1] permet d'insérer un tatouage sur la piste audio de contenus cinématographiques détectable/décodable par les lecteurs *Blu-ray*. Nous voyons tout de suite une des principales contraintes liées à l'ajout d'un tatouage : sa robustesse face à toute transformation du support (compression, recadrage, etc). En effet, la marque ne doit pas pouvoir être effacée (même de façon accidentelle par un utilisateur). De plus, cette marque ne doit pas détériorer le support, elle ne doit pas être perceptible par l'utilisateur. La dernière contrainte qui a pris de plus en plus d'importance est la sécurité et le choix d'une clé secrète, qui, comme en cryptographie, respecte le principe de Kerckhoffs pour l'insertion et le décodage. La localisation du message caché doit uniquement être accessible aux possesseurs de la clé. Un utilisateur malveillant possédant la clé essaiera alors de modifier ou de supprimer le tatouage.

Ce manuscrit s'inscrit logiquement dans le contexte de la protection des contenus numériques et présente nos travaux en sécurité en tatouage numérique et en estampillage effectués au sein des laboratoires GIPSA-Lab et TELE pendant l'ensemble du doctorat :

- nous cherchons à protéger les contenus en insérant une information indélébile (même en présence d'un adversaire),
- nous cherchons, par l'ajout de codes d'estampillage, à identifier les utilisateurs qui mettent à disposition de manière illégale ces contenus.

Organisation du manuscrit

La première partie de cette thèse concerne la sécurité en tatouage numérique. Le premier chapitre porte sur l'étude de techniques de tatouage par étalement de spectre dites sûres dans le cadre WOA (*Watermarked Only Attack*) : un adversaire possède plusieurs contenus tatoués avec la même clé secrète mais avec des messages différents et essaie d'estimer cette clé secrète. Selon la qualité de l'estimation, l'adversaire pourra lire ou même modifier les informations cachées dans ces contenus. Nous présentons deux techniques de tatouage par étalement de spectre, le tatouage circulaire, CW (*Circular Watermarking*) et le tatouage naturel, NW (*Natural Watermarking*) permettant un tatouage sûr. Néanmoins, le principale faiblesse de ces techniques réside dans leur robustesse face à des attaques de type ajout de bruit gaussien ou compression.

La sécurité d'un schéma de tatouage étant dépendante de la distribution des contenus avant et après tatouage, nous utilisons, dans le second chapitre, l'algorithme des Hongrois qui permet de minimiser de façon globale la distance entre distributions hôte et tatouée (graphe bipartite en théorie des graphes). Cette minimisation nous a assuré une distorsion d'insertion plus faible ainsi qu'une meilleure robustesse des schémas CW et NW sans changer leur niveau de sécurité [60]. Nous avons ensuite amélioré cette technique en utilisant les résultats de la théorie du transport qui, comme l'algorithme des Hongrois, nous permet d'améliorer les schémas CW et NW en terme de robustesse et de distorsion mais avec une complexité considérablement réduite et de meilleures performances en grande dimension (c'est-à-dire la longueur en bits des messages insérés) [56, 59].

Dans le troisième chapitre, nous avons tatoué des images naturelles grâce à des méthodes par étalement de

spectre dans le domaine des ondelettes en utilisant d'une part des méthodes classiques non sûres : étalement de spectre classique, SS (*Spread-Spectrum*), et amélioré, ISS (*Improved Spread Spectrum*), et d'autre part les méthodes sûres : NW et CW. Des techniques de séparation de sources, comme l'analyse en composantes indépendantes (ACI) ou l'analyse en composantes principales (ACP) sur les images tatouées permettent d'estimer correctement la clé secrète pour SS et ISS contrairement aux méthodes NW et CW pour lesquelles l'adversaire obtient peu d'information sur ce secret [54, 55, 61].

La seconde partie de ce manuscrit concerne le cas particulier de l'estampillage qui permet de tracer les redistributeurs de copies illégales d'un contenu. Le distributeur envoie chaque version de son contenu tatoué avec l'identifiant de l'utilisateur concerné. Cependant, un groupe d'adversaires peut travailler ensemble en mixant leurs contenus et ainsi forger un contenu piraté qui contiendrait un identifiant illisible ou pire, l'identifiant d'un utilisateur innocent. Cette attaque est appelée attaque de coalition.

Le quatrième chapitre présente les codes de Tardos, offrant une solution optimale pour l'estampillage de contenus résistant aux attaques de coalition. Ces codes sont dits probabilistes car la probabilité d'accuser un innocent n'est pas nulle, cette probabilité étant dépendante de la longueur de ces codes. Nos derniers travaux présentent une attaque de coalition qui minimise l'information mutuelle entre le contenu d'un adversaire et le contenu piraté forgé par la coalition dans un contexte de technique de tatouage sécurisé. Cette attaque "au pire cas" dépend du degré d'estimation des identifiants par les adversaires et leur permet de minimiser la quantité d'information utilisée pour le traçage. De plus, nous avons quantifié le compromis qui doit être fait entre robustesse et sécurité du tatouage pour permettre de réduire les performances d'une attaque de coalition [57, 58]. Nous montrons l'intérêt d'utiliser des méthodes de tatouage sûres et moins robustes dans ce cadre applicatif.

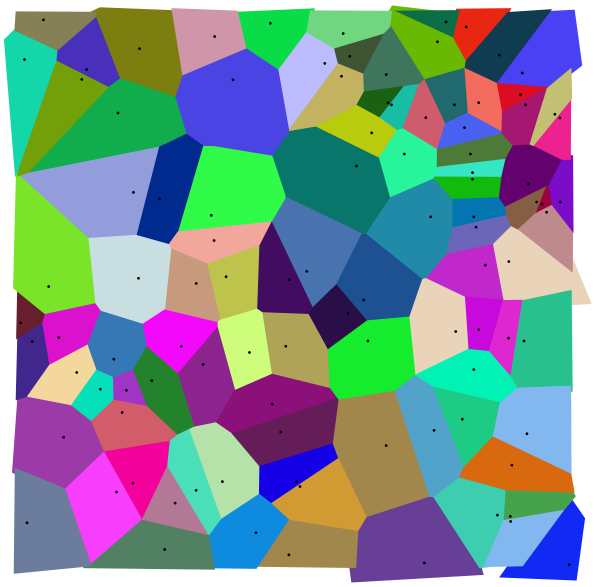
Enfin, dans le chapitre final de ce manuscrit, nous avons implémenté les codes de Tardos dans un flux vidéo en utilisant les techniques sûres par étalement de spectre. Le tatouage s'effectue image par image (à raison de 25 images/seconde) dans le domaine pixellique. L'identifiant à cacher est répété (après permutations) dans tout le flux vidéo. Nous montrons que l'utilisation de méthodes sûres ne provoque pas une perte significative de robustesse du schéma (grâce aux répétitions permutées de l'identifiant) et que l'accusation des adversaires est correcte même lorsque la qualité de réencodage de la vidéo et le nombre de répétitions de l'identifiant sont faibles.

Première partie

La sécurité en tatouage numérique

Chapitre 1

La contrainte de sécurité en tatouage



Cellules de Voronoi [1], on appelle région de Voronoï ou cellule de Voronoï associée à un élément p de $\mathcal{S}$ (ensemble des centroïdes d'un espace métrique $(\mathcal{E}, \mathrm{d})$), l'ensemble des points qui sont plus proches de p que de tout autre point de S : $\mathrm{Vor}_{\mathcal{S}}(\mathbf{p}) = \{\mathbf{x} \in \mathcal{E} \ : \ \forall \mathbf{q} \in \mathcal{S},\ \mathrm{d}(\mathbf{x}, \mathbf{p}) \leq \mathrm{d}(\mathbf{x}, \mathbf{q})\}$.

Sommaire

1. Auteur : Mysid (SVG), Cyp (original), licences GFDL (www.gnu.org/copyleft/fdl.html) et CC-BY-SA-3.0 (www.creativecommons.org/licenses/by-sa/3.0/).

NOUS avons vu précédemment que le passage de l'analogique au numérique a permis une meilleure gestion de la plupart des documents multimédia (musiques, films, images). En effet, le stockage des données est plus facile et l'indexation moins coûteuse. Cependant, son principal inconvénient réside dans le fait qu'on ne puisse pas distinguer une copie d'un original. L'évolution d'Internet et de réseaux d'échanges de données a accéléré la piraterie sur la propriété intellectuelle. Les œuvres soumises au droit d'auteur peuvent être partagées illégalement via téléchargement direct (*Megaupload*, *Rapidshare*), réseau pair à pair (*eMule*, *Torrent*), LAN dans le cadre de jeu en réseau, compression de DVD loués ou prêtés sous forme de fichiers DivX ou tout simplement par échange de clés USB entre amis ou collègues.

Le tatouage numérique est une technique permettant de résoudre certains problèmes liés aux droits d'auteur, elle consiste à insérer une marque dans un support numérique. Ce premier chapitre de ce manuscrit de thèse présente la sécurité intrinsèque des techniques de tatouage numérique, il est organisé de la manière suivante : les deux premières sections (1.1 et 1.2) définissent le tatouage numérique en présentant ses applications ainsi que la mise en œuvre dans un cadre pratique. Les deux sections suivantes (1.3 et 1.4) présentent la contrainte de sécurité en tatouage ainsi que les contextes d'attaque, la sécurité étant liée à l'utilisation d'une clé secrète symétrique. Ensuite, une grande famille parmi les techniques de tatouage existantes est présentée : le tatouage par étalement de spectre (1.5). Nous terminons ce chapitre en explicitant les attaques pratiques de sécurité utilisées par un ou plusieurs adversaires dans le cadre du tatouage par étalement de spectre ainsi que des techniques existantes permettant de parer ces attaques.

1.1 Applications et contraintes

Le tatouage numérique (*digital watermarking*) consiste à insérer une marque dans un support numérique. Dans le cadre de la diffusion de ces contenus par moyens informatiques, la marque à insérer peut être un indicateur de présence (booléen) ou un message (donnée constituée de plusieurs bits). Dans le premier cas on parle de tatouage zéro-bit, la relecture du canal caché est un problème de détection ; dans le deuxième il s'agit de tatouage multi-bits, il s'agit alors d'un problème de décodage.

L'ajout de cette marque dans un document hôte est utilisé pour répondre à certains besoins :

- **Respect des droits d'auteur** : une marque (numéro d'identifiant) propre à un auteur est caché dans son œuvre. Ce numéro peut être attribué et référencé par un tiers de confiance. L'auteur pourra ensuite légalement prouver qu'il est l'auteur de son contenu lors d'usurpations possibles effectuées par un adversaire ;
- **Protection de la copie** : une marque interdisant la copie peut être insérée dans le contenu à protéger. Avant toute opération de copie, le dispositif permettant cette action analysera le contenu ; si une telle marque est détectée, la copie est refusée ;
- **Protection des contenus** : certains lecteurs acceptent de lire uniquement les DVD qui sont cryptés et tatoués (DVD du commerce) ainsi que les DVD qui ne sont ni cryptés ni tatoués (créations personnelles, vidéos de vacances). Un DVD tatoué mais non crypté (vraisemblablement piraté) sera rejeté par le lecteur en question ;
- **Contenus augmentés** : un contenu peut proposer une autre utilisation que celle pour laquelle il a été prévu à la base. Par exemple, le projet Artus [4] vise à insérer dans les émissions classiques de télévision une information imperceptible (ne dégradant pas le contenu) permettant d'animer à la réception un clone 3D communiquant avec des personnes malentendantes. Seules les personnes disposant d'un dispositif adéquat peuvent jouir de cette fonctionnalité ;
- **Estampillage** : l'estampillage ou traçage de traitres (*fingerprinting* ou *traitor tracing*) est une technique destinée à tracer les copies légales d'un contenu en insérant un identifiant propre au possesseur d'une copie.

Si celle-ci est retrouvée sur un réseau d'échange (par exemple pair à pair), il sera alors possible d'identifier la personne responsable de sa diffusion. Le cadre de l'estampillage fait partie intégrante de ce manuscrit de thèse et sera détaillé dans la seconde partie de ce manuscrit.

Le processus de tatouage est sujet à quatre contraintes que doit respecter l'ajout d'une marque. L'importance de ces contraintes dépend du cadre applicatif défini. Ces contraintes sont :

- **Imperceptibilité** : l'insertion de la marque ne doit pas détériorer le contenu. Par exemple, elle ne doit pas être vue dans le cadre du tatouage d'images ou entendue dans le cas de données sonores ;
- **Robustesse** : la marque doit résister à des modifications du support telles une compression, des déformations géométriques (tatouage d'images), l'ajout de bruit, etc. Il faut aussi noter que dans le cadre de la diffusion ou stockage des données sur supports informatiques, ces modifications peuvent ne pas être intentionnelles. Elles sont nommées attaques de robustesse en tout généralité ;
- **Fiabilité** : contrainte principalement liée au tatouage zéro-bit, il s'agit de minimiser la probabilité de fausse alarme p_{fa}, c'est-à-dire la probabilité de détecter une marque dans un contenu non tatoué (faux positif) en compromis avec la probabilité d'erreur p_e, c'est-à-dire la probabilité de ne pas détecter une marque dans un contenu tatoué (faux négatif) ;
- **Capacité** : quelle quantité d'information est insérée dans le contenu hôte ? Il peut s'agit de quelques bits pour la protection de la copie, mais un traçage de traitres performant nécessitera beaucoup plus d'information ;
- **Sécurité** : cette dernière contrainte a pris de plus en plus d'importance au sein de la communauté des tatoueurs et fait l'objet d'une étude approfondie dans ce manuscrit de thèse. Un schéma de tatouage respecte généralement le principe de Kerckhoffs [47] : l'algorithme et les paramètres du schéma de tatouage sont publics. Une clé secrète permettant l'insertion de la marque et de son décodage (ou sa détection) est l'unique paramètre inconnu d'une personne autre que le distributeur ou auteur du contenu. Une attaque de sécurité consiste alors en une estimation de cette clé secrète.

La figure 1.1 illustre un schéma de tatouage soumis aux attaques de robustesse.

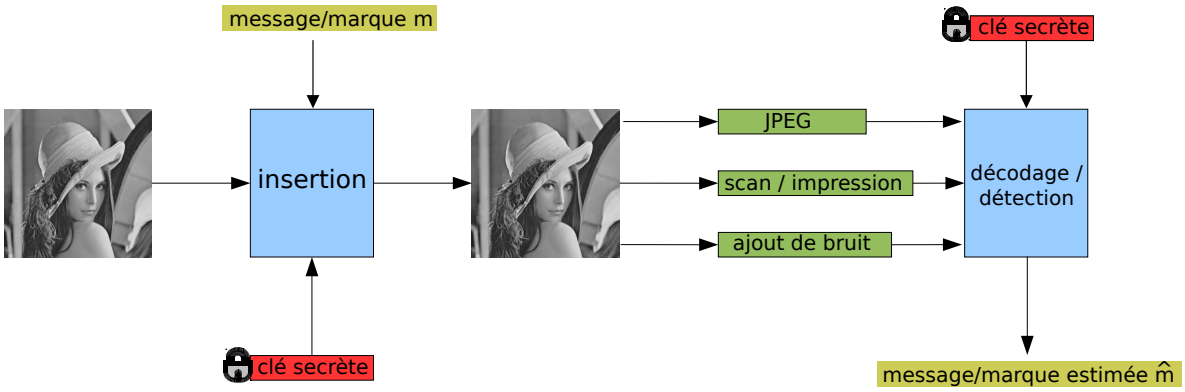


FIGURE 1.1 – Schéma de tatouage : schéma général. L'insertion et le décodage (ou la détection) sont soumis à l'utilisation d'une clé secrète. Le contenu tatoué peut subir des attaques de robustesse comme une compression, une impression suivie d'une acquisition (via un scanner) ou, en toute généralité, un ajout de bruit.

1.2 Modélisation d'une chaine de tatouage

La chaine de tatouage d'un contenu multimédia est modélisée figure 1.2. Le contenu original à tatouer $\mathbf{C}_o$

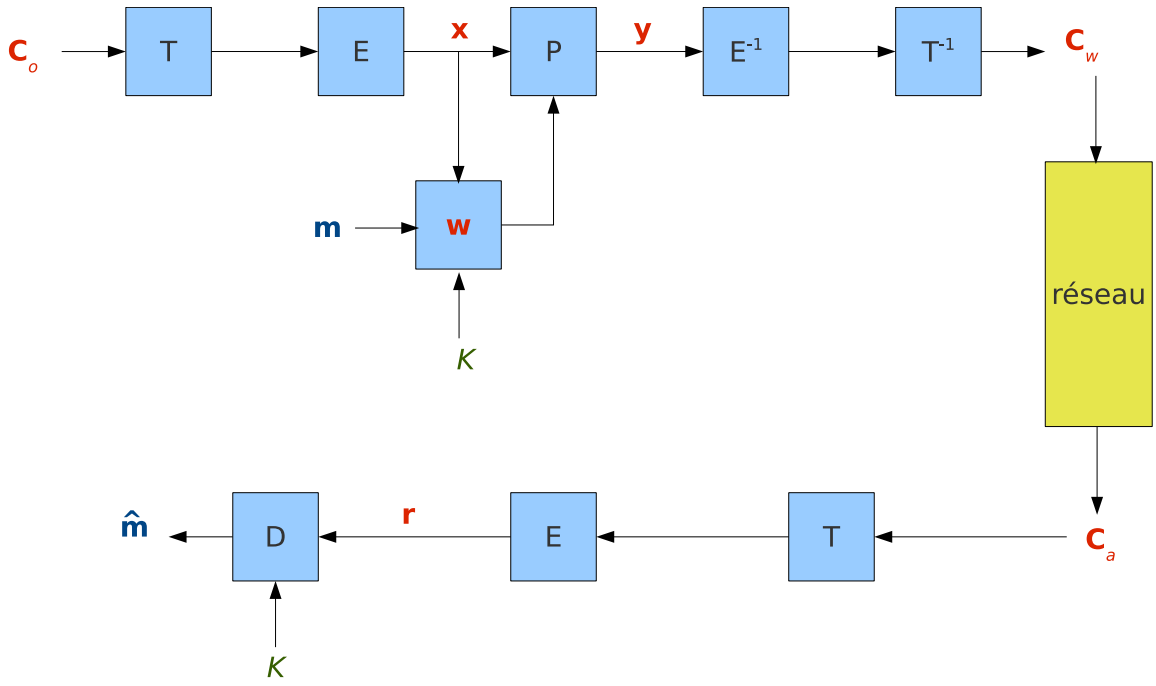


FIGURE 1.2 – Modélisation d'une chaine de tatouage avec information sous-jacente (la connaissance du contenu hôte $\mathbf{x}$ permet d'améliorer la robustesse ainsi que l'imperceptibilité du tatouage [26]).

est préalablement transformé à l'aide d'une fonction T afin d'en extraire un signal caractéristique $\mathbf{x}$. Ce signal représente les informations les plus pertinentes du contenu que l'on souhaite tatouer, il peut s'agir par exemple :

- des coefficients de la Transformée en Cosinus Discrète par Bloc (Block-DCT) d'une image,
- des coefficients de la Transformée de Fourier Rapide après fenêtrage (Window-FFT) d'un son,
- des coefficients de la Transformée de Laplace de la géométrie d'un maillage tridimensionnel.

Il est préférable, si possible, d'utiliser des coefficients pouvant être modélisés mathématiquement afin de mieux contrôler les contraintes définies par le cadre applicatif.

Le choix de la fonction d'extraction E est tout aussi lié à ces contraintes. Pour le tatouage d'images fixes, il peut s'agit des coefficients correspondants aux hautes et moyennes fréquences après transformée. Les hautes fréquences seront beaucoup plus sensibles aux attaques de robustesse par compression mais offriront un meilleur rendu d'imperceptibilité après tatouage.

Le message ou marque $\mathbf{m}$ à insérer est modulé en un signal de tatouage $\mathbf{w}$ par une fonction S paramétrée par la clé secrète K (une modulation S peut aussi dépendre du vecteur hôte $\mathbf{x}$, on parle dans ce cas d'insertion avec information sous-jacente [26]). Le signal tatoué $\mathbf{y}$ est obtenu par $\mathrm{P}(\mathbf{x}, \mathbf{w}, \alpha, \lambda)$ où α permet de fixer la puissance d'insertion du signal $\mathbf{w}$, et λ permet de contrôler l'information sous-jacente. Nous obtenons ensuite le contenu tatoué $\mathbf{C}_t = \mathrm{T}^{-1}(\mathrm{E}^{-1}(\mathbf{y}))$. Ce contenu est alors envoyé sur le réseau, par un canal non sécurisé. Il peut subir des attaques de robustesse et de sécurité et devient alors $\mathbf{C}_a$, contenu attaqué.

Pour le décodage (resp. la détection), nous extrayons un signal $\mathbf{r}$ du contenu attaqué $\mathbf{C}_a$ après transformée. A l'aide de la clé secrète, nous décodons (resp. détecte) le message (resp. marque) $\hat{\mathbf{m}}$.

1.3 Définition d'une clé secrète

1.3.1 Introduction

Dans la plupart des schémas actuels de tatouage, l'insertion et le décodage (détection) d'un message (d'une marque) dans un contenu respectent le principe de Kerckhoffs et dépendent d'une clé secrète symétrique [2]. La clé est l'unique paramètre du schéma de tatouage qui est privé des utilisateurs. Dans *La Cryptographie Militaire* [47], Kerckhoffs définit les conditions que doit respecter un système cryptographique afin de pouvoir être utilisé pour un temps illimité par deux correspondants. Ces conditions peuvent être appliquées à un système de tatouage. Voici les trois premiers points importants (repris directement de [47]) ainsi que leur transcription au tatouage numérique :

1. *Le système doit être matériellement, sinon mathématiquement indéchiffrable ;*
   - L'estimation de la clé secrète par un adversaire est impossible de par la complexité du calcul nécessaire pour cette tâche ou des propriétés mathématiques du système.
2. *Il faut qu'il n'exige pas le secret, et qu'il puisse sans inconvénient tomber entre les mains de l'ennemi ;*
   - Le code source de l'algorithme de tatouage ainsi que les grandeurs utilisées doivent être publiques. La seule donnée privée (inconnue d'un adversaire) est une clé secrète.
3. *La clef doit pouvoir en être communiquée et retenue sans le secours de notes écrites, et être changée ou modifiée au gré des correspondants ;*
   - La clé secrète doit être numériquement légère et pouvoir être stockée de façon sûre. De plus le tatouage d'un contenu avec une clé différente doit pouvoir être effectué facilement.

Nous nous intéressons dans ce manuscrit de thèse au tatouage multi-bits : la marque à insérer contient un message, la clé secrète associée est définie comme la localisation dans un contenu numérique de régions d'encodage et de décodage.

La clé secrète utilisée par un distributeur de contenus est choisie de manière aléatoire pour minimiser l'information mutuelle entre cette même clé et le contenu à tatouer (une dépendance entre clé et contenu faciliterait l'estimation du secret par un adversaire). De plus, le distributeur n'a pas forcément accès au contenu hôte original lors du décodage du message.

Nous prenons comme hypothèse de travail le choix d'une même clé secrète pour le tatouage de plusieurs contenus. Si cette dernière hypothèse paraît très restrictive, elle permet surtout d'éviter les problèmes liés à la resynchronisation entre un contenu tatoué et la clé secrète qui lui est associé. Prenons le cas d'un film ou chaque image est tatouée avec une clé différente, le décodage devient très complexe si, après récupération du contenu sur un canal non-sûr, celui-ci a été tronqué de quelques images, compressé ou déformé géométriquement. Le système de tatouage devrait alors être couplé d'un système de resynchronisation efficace. De même, cette resynchronisation ne peut être assurée que si celle-ci est sécurisée ; utiliser plusieurs clés secrètes devient alors plus coûteux (implantation, stockage).

Afin de bien différencier les différents acteurs intervenant dans la chaine de tatouage, nous posons une définition :

Définition 1 *:*

1. *Le **distributeur** est la personne chargée de tatouer un ou plusieurs contenus (œuvres) numériques, il est le seul à connaître la clé secrète utilisée pour l'ajout de la marque. Le distributeur peut être l'auteur de l'œuvre ou, par délégation, un tiers de confiance ;*

2. Il existe des méthodes de tatouage asymétriques mais celles-ci sont plus difficiles à manier [33].

2. *L'**utilisateur** est une personne jouissant de certains privilèges liés au contenu : lecture, diffusion dans le cadre familial, etc ;*
3. *L'**adversaire** est une personne tentant d'estimer la clé secrète. Selon le degré d'estimation, l'adversaire peut apporter des modifications non autorisées par le distributeur au tatouage. Selon le cadre applicatif, il peut s'agir de la destruction de la marque (protection de la copie), sa modification (estampillage) ou même la copie de cette marque dans un autre contenu numérique (droits d'auteur : usurpation d'identité). L'adversaire peut être un utilisateur ou une personne ayant corrompu un (ou plusieurs) utilisateur(s).*

Une clé secrète de tatouage peut être la localisation de coefficients à modifier dans un contenu, un bruit rajouté dans le signal hôte, il existe autant de rôles pour une clé secrète que de schémas d'insertion. Néanmoins, celle-ci peut être formellement définie par un ensemble de N_h mots de code dans une variété de dimension N_d propre au domaine d'insertion du tatouage dans un signal. Le nombre de mots de code utilisé doit être supérieur ou égal aux nombre de messages possibles à insérer dans le contenu, nous avons alors :

$$N_h \geq 2^{N_c}. \tag{1.1}$$

où N_c est le nombre de bits d'un message $\mathbf{m}$ à insérer.

Tatouer un signal hôte $\mathbf{x}$ revient alors à "pousser" ce signal vers le mot de code associé au message que l'on veut insérer afin d'obtenir un signal tatoué $\mathbf{y}$, l'emplacement de ces mots de codes (ou régions d'insertion et de décodage) dans la variété de dimension N_d étant en bijection avec une clé secrète K_s. La figure 1.3 donne un exemple de tatouage d'un signal $\mathbf{x}$ avec $N_c = 2$ bits. Les cellules colorées désignent les mots de code secrets dans une variété de dimension N_d, un message étant codé par un ou deux mots de codes. Le signal tatoué $\mathbf{y}$ est choisi dans la zone de décodage correspondant au message que l'on veut insérer. Cette figure permet d'illustrer les contraintes de robustesse et de distorsion (liée à l'imperceptibilité) :

– *robustesse* : le choix d'un $\mathbf{y}$ au centre d'une cellule (éloigné des frontières de décodage) permettra une meilleure robustesse du tatouage,

– *distorsion* : la distance entre $\mathbf{x}$ et $\mathbf{y}$ est proportionnelle à la distorsion provoquée par l'ajout de tatouage.

Les cellules de Voronoi illustrées dans la figure du sommaire de ce premier chapitre peuvent jouer le rôle d'une clé secrète : une robustesse maximale sera assurée pour un signal tatoué situé au centroïde d'une cellule. Au contraire, le choix d'un signal tatoué situé à la frontière de la cellule la plus proche (codant le message voulu) de celle du signal hôte permettra un meilleur contrôle de la distorsion.

Les méthodes par quantification utilisent des cellules de Voronoi pour le tatouage. Nous verrons comment prendre en compte la contrainte de la sécurité dans la suite de ce manuscrit de thèse.

1.3.2 Estimation de la clé secrète par les adversaires

Une attaque de sécurité est liée à l'estimation de la clé secrète (mots de code) par un ou plusieurs adversaires. Deux scénarios d'attaque sont alors possibles :

1. **Attaque par oracle** : cette attaque est surtout utilisée dans le cadre du tatouage zéro-bit. L'adversaire a accès au détecteur. Il peut alors générer ses propres contenus avec plusieurs clés, et, en fonction de la réponse du détecteur, obtenir des informations sur la (ou les) frontière(s) de détection ;
2. **Attaque par fuite d'information** : ces attaques seront celles qui seront approfondies dans ce manuscrit de thèse. L'adversaire a accès à plusieurs contenus tatoués et tente d'estimer les régions de décodage. Il peut aussi avoir d'autres informations (accès aux contenus hôtes, à la nature des messages) mais n'a a priori aucun accès à la clé secrète ni au décodeur. La figure 1.4 illustre cette attaque, elle représente une distribution de contenus tatoués (messages et mots de code différents) avec la clé secrète utilisée

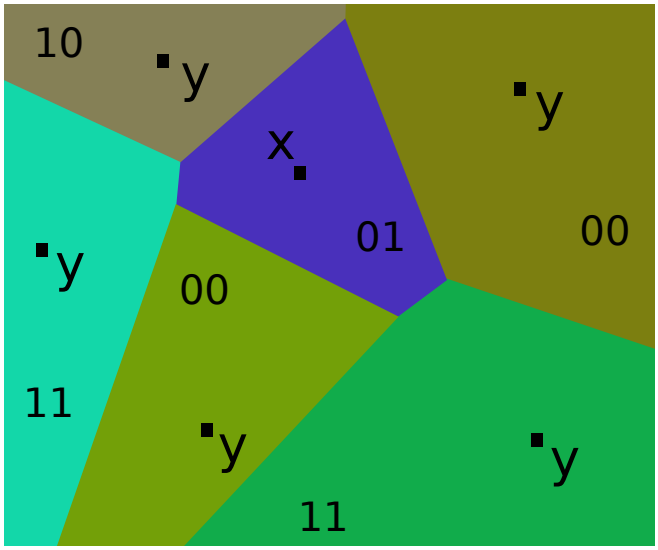


FIGURE 1.3 – Clé secrète d'un schéma de tatouage dans une variété à N_v dimensions : sous-variété de dimension N_d. Les cellules colorées représentent les mots de codes secrets, un message de $N_c = 2$ bits est codé par un ou deux mots de codes dans cet exemple. Le signal tatoué $\mathbf{y}$ est choisi dans la zone de décodage correspondant au message à insérer.

dans la figure 1.3. Nous remarquons que plus le nombre de contenus tatoués possédés par l'adversaire est important, plus il est facile de distinguer les différents mots de code. Dans la suite, nous notons N_o le nombre de contenus auquel un adversaire a accès.

Cette dernière figure 1.4 montre aussi le compromis qui doit être fait entre les contraintes imposées par le système de tatouage. Plus les contenus s'éloignent des frontières des zones de décodage (tatouage robuste), plus il est alors facile d'estimer les mots de code lors d'une attaque par fuite d'information. Une meilleure robustesse implique un défaut de sécurité. Dans le cas où les signaux se confondent avec les frontières de décodage, il est alors impossible d'estimer correctement les mots de code mais l'estimation de la variété privée de dimension N_d est peut-être possible.

Selon le degré d'estimation de la variété privée et des zones de décodage, l'adversaire pourra alors :

1. **altérer le message** d'un signal tatoué en l'amenant au niveau de la plus proche frontière entre le mot de code du message actuel et un mot de code d'un message différent permettant ainsi de minimiser la distorsion avec une probabilité de 1 (attaque au pire cas). La figure 1.5 illustre cette attaque : l'adversaire a estimé les frontières de décodage de la clé secrète grâce à la méthode utilisée figure 1.4, il construit un signal attaqué $\mathbf{r}$ cachant le message $1\ 1$ à partir d'un signal $\mathbf{y}$ (dont le message initial est $0\ 0$) en minimisant la distorsion de manière déterministe [3],
2. **copier** le message d'un signal tatoué sur un nouveau contenu hôte : l'adversaire a accès à la variété privée dans laquelle a lieu l'ajout de tatouage et peut alors copier les projections du contenu tatoué dans cette variété sur un autre contenu sans l'altérer considérablement.

La figure 1.6 illustre le processus de tatouage et d'attaque dans le cas d'attaque par fuite d'information : un adversaire a accès à plusieurs contenus tatoués avec la même clé secrète. Selon le nombre de contenus en sa

3. Notons que l'adversaire n'a a priori aucune information sur la valeur des messages répartis dans les mots de code, en utilisant plusieurs mots de codes pour un même message, l'adversaire ne peut pas déterminer si le message original a été modifié.

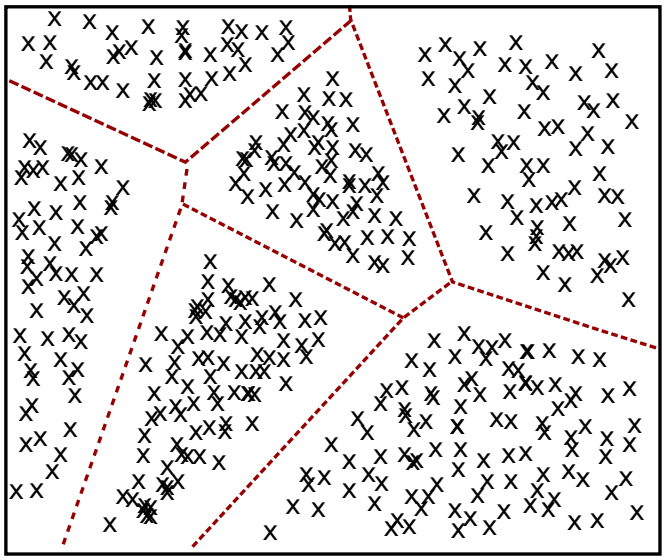


FIGURE 1.4 – Distribution de plusieurs contenus tatoués avec la clé secrète représentée dans la figure 1.3. Plus le nombre de contenus auquel l'adversaire a accès est important, plus la distinction entre les régions de décodage est apparente.

possession et la sécurité du schéma de tatouage, ce dernier aura une estimation de la clé secrète partielle ou totale et pourra alors altérer les messages cachés (effacement ou modification des messages) ou copier les messages sur d'autres contenus vierges.

1.4 Contextes d'attaque en fuite d'information et classes de sécurité

L'idée des attaques par fuite d'information est d'obtenir assez de contenus tatoués N_o afin d'estimer correctement la clé secrète K_s. Notons que, selon le principe de Kerckhoffs, l'adversaire a accès à toutes les données publiques du tatouage (domaine de transformation dans lequel a lieu l'algorithme de tatouage : transformée T et fonction d'extraction E, nombre de bits du message N_c, etc.). Il n'a au préalable aucune information sur la clé secrète K_s. Nous notons :

- $\mathcal{X}$: ensemble de contenus orignaux,
- $\mathcal{Y}(K_i)$: ensemble de contenus tatoués avec une clé K_i,
- $\mathcal{S}$: ensemble des clés secrètes possibles,
- $\mathcal{Y}_{K_s}$: l'ensemble des contenus tatoués avec la clé secrète K_s.

Les données connues de l'adversaire sont alors :

- $\mathrm{p}(\mathcal{X})$: la distribution de signaux hôtes. Celle-ci peut-être connue explicitement (formule analytique) ou alors modélisée : l'adversaire peut générer ses propres signaux hôtes de dimension N_d à l'aide du système de tatouage en partant de contenus vierges (transformation T suivie de la fonction d'extraction E,
- $\mathrm{p}(\mathcal{Y}(K_i))$: la distribution (connue explicitement ou modélisée) de signaux tatoués à l'aide d'une clé K_i, K_i étant généré par l'adversaire,
- $\mathrm{p}(\mathcal{Y}_{K_s}|K_i)$: la distribution des N_o signaux tatoués à l'aide de la clé secrète K_s conditionnellement à une clé K_i choisie par l'adversaire.

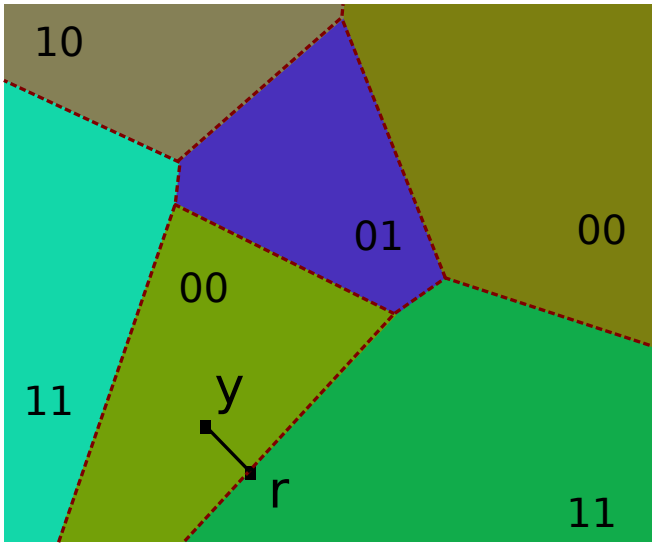


FIGURE 1.5 – Modification d'un message inséré dans un contenu en minimisant la distorsion avec une probabilité de 1. Les frontières des zones de décodage sont estimées à l'aide de l'attaque par fuite d'information réalisée dans la figure 1.4. L'adversaire construit un signal attaqué $\mathbf{r}$ cachant le message 11 à partir d'un signal $\mathbf{y}$ (dont le message initial est 00) en minimisant la distorsion de manière déterministe.

L'utilisation d'une clé secrète privée symétrique ne garantit EN AUCUN CAS à elle seule la sécurité du schéma de tatouage. Selon le cadre applicatif dans lequel le système de tatouage est utilisé, un ou plusieurs adversaires peuvent estimer la clé secrète. Dans [20], les auteurs définissent des contextes d'attaque par fuite d'information en fonction des données que possèdent les adversaires : KOA (*Known Original contents Attack*), KMA (*Known Message Attack*) et WOA (*Watermarked contents Only Attack*). Dans [70] les auteurs ajoutent le cas CMA (*Constant Message Attack*). Ces contextes sont présentés et illustrés par des exemples dans la table 1.4.

Dans ce manuscrit de thèse nous travaillons principalement sous l'hypothèse WOA qui est la plus restrictive en terme d'attaque par fuite d'information. Dans ces quatre contextes d'attaque, les contenus sont tous tatoués avec la même clé secrète K_s. Nous mesurons la sécurité de l'algorithme de tatouage par l'IM (Information Mutuelle) I en nats entre les N_o signaux $\mathcal{Y}_{K_s}$ et la clé secrète K_s :

$$I(\mathcal{Y}_{K_s}; K_s) = H(\mathcal{Y}_{K_s}) - H(\mathcal{Y}_{K_s}|K_s), \tag{1.2}$$

où $H(V)$ désigne l'entropie d'une variable aléatoire V :

$$\mathrm{H}(V) = \int \mathrm{p}(V) \log \mathrm{p}(V) d(V). \tag{1.3}$$

L'entropie mesure la quantité d'information délivrée par une source d'information : ici notre variable aléatoire V. Il s'agit de la moyenne de l'information associée aux diverses occurrences de V (une occurrence non fréquente contient plus d'information qu'une occurrence fréquente). L'entropie peut aussi être interprétée comme la dispersion dans une distribution de probabilité (voir l'article fondateur de Shannon [75]).

Une information mutuelle nulle implique l'indépendance entre les deux variables, c'est-à-dire que l'information donnée par une de ces variables n'a aucune incidence sur la connaissance de l'autre.

La sécurité pour les contextes d'attaque définis précédemment est donnée par :

– **CMA** : $I(\mathcal{Y}_{K_s}; K_s|M_0)$, où $\forall i \in [N_o]$, où $\forall i \in [N_o]$ le message M_0 est inséré dans chaque signal Y_i,

Contexte d'attaque	Définition	Exemple pratique
KOA : *Known Original contents Attack*.	L'adversaire a accès à plusieurs contenus tatoués et possède aussi les originaux.	Images récupérées d'un film tatoué par le distributeur ainsi qu'une bande annonce (*trailer*) préalablement diffusée avant la commercialisation de ce film et non tatouée.
KMA : *Known Message Attack*.	L'adversaire a accès à plusieurs contenus tatoués et connaît les messages insérés.	Bibliothèque musicale dans laquelle chaque fichier sonore est tatoué avec le nom de l'artiste ou compositeur (connu de l'attaquant).
CMA : *Constant Message Attack*.	L'adversaire a accès à plusieurs contenus tatoués avec le même message (l'adversaire ne connait pas le message mais sait que celui-ci est le même pour chaque contenu).	Galerie de photographies d'un artiste publiées sur un site internet, chaque photographie est tatouée avec l'identifiant de l'artiste (l'identifiant est inconnu des visiteurs du site).
WOA : *Watermarked contents Only Attack*.	L'adversaire a uniquement accès à plusieurs contenus tatoués avec des messages différents.	Film dont chaque image est tatouée avec un tronçon d'un identifiant (par exemple, celui reliant la copie du film à son utilisateur, contexte de tatouage en estampillage). Cet identifiant est bien entendu inconnu de l'utilisateur.

TABLE 1.1 – Contextes d'attaque par fuite d'information.

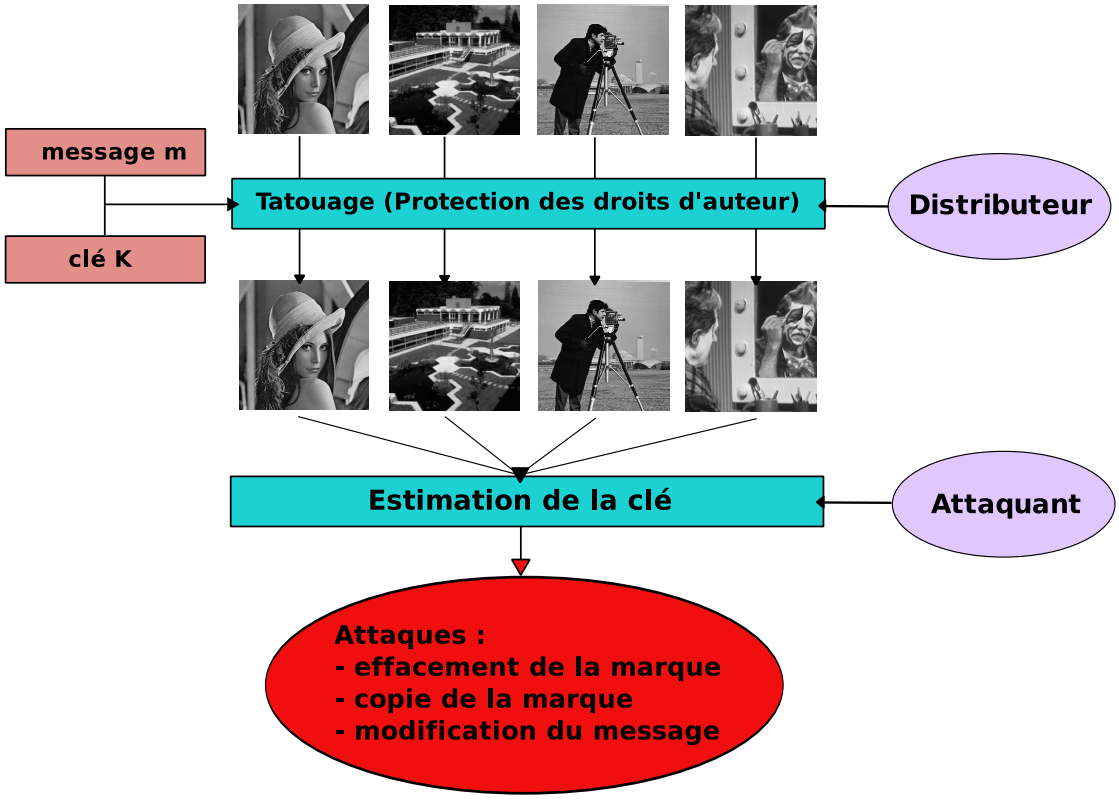


FIGURE 1.6 – Attaque par fuite d'information : un adversaire a accès à plusieurs contenus tatoués avec la même clé secrète. Selon le nombre de contenus en sa possession et la sécurité du schéma de tatouage, ce dernier aura une estimation de la clé secrète partielle ou totale et pourra alors altérer les messages cachés (effacement ou modification des messages) ou copier les messages sur d'autres contenus vierges.

- **KMA** : $I(\mathcal{Y}_{K_s}; K_s | M_0, \ldots, M_{N_o-1})$, où $\forall i \in [N_o]$, le message M_i est inséré dans le signal Y_i,
- **KOA** : $I(\mathcal{Y}_{K_s}; K_s | X_0, \ldots, X_{N_o-1})$, où $\forall i \in [N_o]$, X_i désigne le signal hôte original.
- **WOA** : $I(\mathcal{Y}_{K_s}; K_s)$, où $\forall i \in [N_o]$, X_i désigne le signal hôte original.

Lorsque l'IM est nulle, il n'y a pas de fuite d'information, l'adversaire ne peut pas estimer la clé secrète. Dans le cas contraire, l'information mutuelle peut augmenter en fonction de N_o, de la distorsion d'insertion, en fonction de la sécurité du schéma utilisé.

Dans [6], les auteurs définissent des classes de sécurité pour les systèmes de tatouage en WOA, relatives aux degrés d'estimation de la clé secrète possible par les adversaires :

- **non-sécurité** (*insécurité*) : pour une clé K_i la distribution $\mathrm{p}(\mathcal{Y}_{K_s}; K_i)$ est unique. Les distributions de signaux tatoués avec des clés différentes sont alors différentes. Formellement :

$$\forall K_i \neq K_j \in \mathcal{S},\ \mathrm{p}(\mathcal{Y}_{K_s} | K_i) \neq \mathrm{p}(\mathcal{Y}_{K_s} | K_j).$$

  Un adversaire peut alors, par une recherche exhaustive (ou technique moins complexe, voir section 1.6.1.2), déterminer la clé secrète K_s : il estime la variété privée de dimension N_d ainsi que les régions de décodages (mots de code) ;
- **clé-sécurité** (*sécurité de la clef*) : il existe un sous-ensemble de clés $\mathcal{S}_c$ contenant la clé K_s tel que, pour toutes clés de ce sous-ensemble, les distributions de signaux tatoués conditionnellement à ces clés sont les

mêmes :

$$\forall K_i \in \mathcal{S}_c,\ \mathrm{p}(\mathcal{Y}_{K_s}|K_i) = \mathrm{p}(\mathcal{Y}_{K_s}|K_s).$$

Un adversaire peut alors déterminer un sous-ensemble de clés contenant la clé secrète, ce sous-ensemble est déterminé par la variété privée de dimension N_d.

– **sous-espace-sécurité** (*sécurité du sous-espace*) : la distribution des signaux tatoués est la même, quelque soit la clé utilisée. Nous avons l'égalité :

$$\forall K_i \neq K_j \in \mathcal{S},\ \mathrm{p}(\mathcal{Y}_{K_s}|K_i) = \mathrm{p}(\mathcal{Y}_{K_s}|K_j).$$

L'adversaire n'a alors aucune information sur la clé secrète ;

– **stégo-sécurité** : cette dernière classe est relative à la stéganographie, pour toute clé, la distribution de contenus tatoués conditionnellement à cette clé est égale à la distribution des contenus hôtes. L'adversaire ne peut alors déterminer si un contenu est tatoué ou non. Nous avons :

$$\forall K_i \in \mathcal{S},\ \mathrm{p}(\mathcal{Y}|K_i) = \mathrm{p}(\mathcal{X}).$$

Cette condition est équivalente au critère de Cachin en stéganographie [18] :

$$\mathrm{D_{KL}}\left(\mathrm{p}(\mathcal{X})||\mathrm{p}(\mathcal{Y}(K_i))\right) = 0, \tag{1.4}$$

où $\mathrm{D_{KL}}$ désigne la divergence de Kullback-Leibler.

Le diagramme en figure 1.7 montre les relations existantes entre les différentes classes de sécurité en WOA. Nous avons *stégo-sécurité* $\subset$ *sous-espace-sécurité* $\subset$ *clé-sécurité* et *clé-sécurité* $\cap$ *non-sécurité* = $\emptyset$.

1.5 Le tatouage par étalement de spectre

Les travaux effectuées décrits dans ce manuscrit de thèse concernent principalement l'étude et l'application de méthodes par étalement de spectre pour le tatouage multi-bits [25]. Les techniques d'étalement de spectre sont des méthodes pour lesquelles la puissance générée sur une ou plusieurs porteuses (caractérisant chaque bit du message) est délibérément étalée ou distribuée dans le domaine fréquentiel ou le domaine spatial selon les conditions initiales imposées par le schéma de tatouage. Cette étalement permet un contrôle efficace de la robustesse.

1.5.1 Construction des signaux tatoués

Nous considérons un message $\mathbf{m}$ de N_c bits ($\mathbf{m} \in \mathbb{F}_2^{N_c}$) que l'on veut insérer dans un signal hôte $\mathbf{x}$ de N_v coefficients ($\mathbf{x} \in \mathbb{R}^{N_v}$). La clé secrète utilisée pour l'insertion et le décodage est un ensemble de N_c porteuses $\{\mathbf{u}_i\}_{i\in[N_c]}$. Ces porteuses sont générées à l'aide d'un GNPA (Générateur de Nombres Pseudo-Aléatoires) initialisé à partir d'une graine $K \in \mathbb{N}$. Elles proviennent de vecteurs gaussiens centrés puis orthogonalisés (par la procédure de Gram-Schmidt) et enfin réduits. Nous obtenons : $\forall i \neq j;\ i,j \in [N_c];\ \langle \mathbf{u}_i|\mathbf{u}_j\rangle = 0$ afin d'éviter l'interférence entre porteuses. Pour chaque bit message $\mathbf{m}$ à cacher, nous utilisons une modulation $\mathrm{s} : \mathbb{F}_2 \times \mathbb{R}^{N_v} \to \mathbb{R}$,

$$\forall i \in [N_c],\ \forall \mathbf{x} \in \mathbb{R}^{N_v},\ \mathrm{s}(\mathbf{m}(i), \mathbf{x}) = \alpha(i,\mathbf{x})(-1)^{\mathbf{m}(i)} - \lambda(\mathbf{x})\langle \mathbf{x}, \mathbf{u}_i\rangle, \tag{1.5}$$

où :

– $\alpha(i,\mathbf{x})$ permet d'ajuster la distorsion sur chaque porteuse,
– $\lambda(\mathbf{x})$ quantifie l'insertion informée (information sous-jacente).

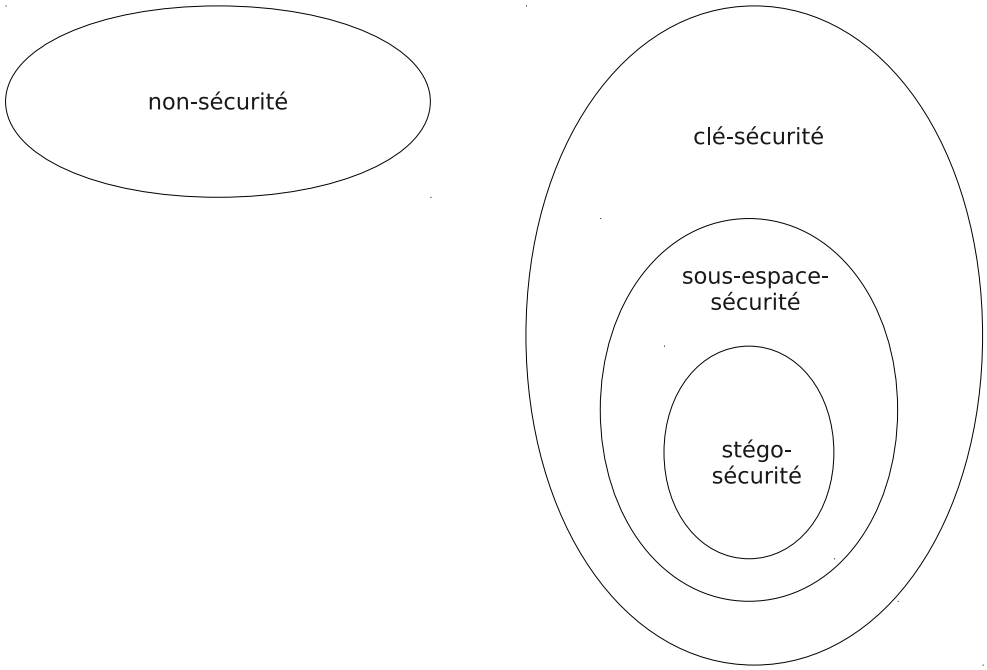


FIGURE 1.7 – Classes de sécurité en WOA. Nous avons les relations *stégo-sécurité* $\subset$ *sous-espace-sécurité* $\subset$ *clé-sécurité* et *clé-sécurité* $\cap$ *non-sécurité* = $\emptyset$.

Nous construisons ensuite le signal tatoué $\mathbf{y}$ comme l'addition du signal hôte et des porteuses pondérés par la modulation choisie :

$$\mathbf{y} = \mathbf{x} + \mathbf{w} = \mathbf{x} + \sum_{i=0}^{N_c-1} \mathrm{s}\left(\mathbf{m}(i), \mathbf{x}\right) \mathbf{u}_i, \tag{1.6}$$

où $\mathbf{w}$ désigne le signal de tatouage. Nous mesurons la distorsion provoquée par l'ajout du signal $\mathbf{w}$ à l'aide du WCR (*Watermark-to-Content power Ratio*) exprimé en décibels :

$$\mathrm{WCR}_{[dB]} = 10 \log_{10} \left(\frac{\sigma^2_{\mathbf{w}}}{\sigma^2_{\mathbf{x}}} \right). \tag{1.7}$$

Comme nous l'avons vu précédemment, le signal tatoué peut être sujet à des attaques de robustesse, intentionnelles ou non. Théoriquement, nous modélisons ces attaques par l'ajout d'un bruit blanc gaussien $\mathbf{n}$ au signal tatoué $\mathbf{y}$, nous considérons alors le signal attaqué $\mathbf{r}$:

$$\mathbf{r} = \mathbf{y} + \mathbf{n}, \tag{1.8}$$

La puissance de l'attaque est exprimée en décibels par le WNR (*Watermark-to-Noise power Ratio*) :

$$\mathrm{WNR}_{[dB]} = 10 \log_{10} \left(\frac{\sigma^2_{\mathbf{w}}}{\sigma^2_{\mathbf{n}}} \right). \tag{1.9}$$

Cependant, dans des applications pratiques du tatouage, pour mesurer l'influence des attaques de robustesse, nous utilisons subsidiairement le WCNR (*Watermarked Content-to-Noise power Ratio*) :

$$\mathrm{WCNR}_{[dB]} = 10\log_{10}\left(\frac{\sigma^2_{\mathbf{y}}}{\sigma^2_{\mathbf{n}}}\right). \tag{1.10}$$

Le décodage du message est assuré par le calcul des corrélations z entre le signal (probablement attaqué) $\mathbf{r}$ et les porteuses secrètes $\{\mathbf{u}_i\}_{i\in[N_c]}$. Nous avons :

$$\forall i \in [N_c],\ \mathrm{z}_{\mathbf{r},\mathbf{u}_i} = \frac{1}{N_v}\langle \mathbf{r}, \mathbf{u}_i\rangle. \tag{1.11}$$

Dans ce manuscrit de thèse, nous utiliserons parfois la notations vectorielle pour le calcul des corrélations : le vecteur de corrélations $\mathbf{z}_{\mathbf{r}}$ entre le signal $\mathbf{r}$ et les porteuses secrètes $\{\mathbf{u}_i\}_{i\in[N_c]}$:

$$\forall i \in [N_c],\ \mathbf{z}_{\mathbf{r}}(i) = \mathrm{z}_{\mathbf{r},\mathbf{u}_i}. \tag{1.12}$$

Lors du calcul de ces corrélations, nous obtenons explicitement :

$$\forall i \in [N_c],\ \mathrm{z}_{\mathbf{r},\mathbf{u}_i} = \mathrm{z}_{\mathbf{x},\mathbf{u}_i} + \mathrm{s}(\mathbf{m}(i),\mathbf{x}) + \mathrm{z}_{\mathbf{n},\mathbf{u}_i}. \tag{1.13}$$

Le premier terme représente l'interférence due au signal hôte, qui peut être utilisée pour l'insertion informée mais qui est statistiquement négligeable comparé au second terme (modulation). Le dernier terme ne peut pas être prédit au moment de l'insertion car il dépend de la puissance de l'attaque de robustesse que subit le contenu tatoué.

Une corrélation forte en valeur absolue permettra une meilleure robustesse. En effet, plus la corrélation est éloignée du seuil de décodage (origine 0), plus l'attaque de robustesse devra être puissante (afin de basculer la corrélation de l'autre côté de l'origine). Cependant, celle-ci se fera au détriment de l'imperceptibilité.

Si $\mathbf{m}'$ représente le message estimé après décodage, nous avons :

$$\forall i \in [N_c],\ \mathbf{m}'(i) = \begin{cases} 0 \text{ si } \mathrm{z}_{\mathbf{r},\mathbf{u}_i} \leq 0, \\ 1 \text{ si } \mathrm{z}_{\mathbf{r},\mathbf{u}_i} > 0. \end{cases} \tag{1.14}$$

Ce décodage est issu des règles classiques de décodage pour les communications par étalement de spectre. Nous mesurons ses performances à l'aide du taux d'erreur binaire BER (*Bit Error Rate*) :

$$\mathrm{BER} = \frac{1}{N_c}\mathrm{d_h}(\mathbf{m},\mathbf{m}') = \frac{1}{N_c}\sum_{i=0}^{N_c-1}\mathbf{m}(i)\oplus\mathbf{m}'(i), \tag{1.15}$$

où $\mathrm{d_h}$ est la distance de Hamming. Les sous-sections suivantes décrivent deux modulations utilisées en tatouage par étalement de spectre : SS (*Spread Spectrum* ou étalement de spectre classique) et ISS (*Improved Spread Spectrum* ou étalement de spectre amélioré). Ces modulations sont non-sûres.

1.5.2 Étalement de spectre classique (SS)

L'étalement de spectre classique SS est analogue à la modulation BPSK (*Binary Phase-Shift Keying* ou modulation binaire par déplacement de phase) pour les communications numériques. En reprenant les notations de l'équation (1.5), nous avons :

$$\alpha_{SS}(i,\mathbf{x}) = \sqrt{\frac{\sigma^2_{\mathbf{x}}10^{\mathrm{WCR}/10}}{N_c}},\ \lambda_{SS}(\mathbf{x}) = 0. \tag{1.16}$$

La valeur de α_{SS} est proportionnelle à la puissance de l'insertion. Plus celle-ci est élevé, plus l'insertion est robuste (au détriment de l'imperceptibilité du tatouage). Dans ce schéma, l'insertion informée n'est pas activée (on n'utilise pas l'information donnée par le signal hôte $\mathbf{x}$ pour l'insertion).

1.5.3 Étalement de spectre amélioré (ISS)

L'étalement de spectre amélioré ISS proposé par Malvar et Florêncio [51] permet de diminuer ou d'annuler l'interférence due au signal hôte pour l'insertion. Le calcul des valeurs de α_{ISS} et λ_{ISS} de l'équation (1.5) est optimisé afin de rejeter l'interférence provoquée par le signal hôte (insertion informée) et de minimiser la probabilité d'erreur en fonction du WCR et du NCR (*Noise-to-Content power Ratio*) donné par :

$$\mathrm{NCR}_{[dB]} = 10\log_{10}\left(\frac{\sigma^2_{\mathbf{n}}}{\sigma^2_{\mathbf{x}}}\right), \tag{1.17}$$

où $\mathbf{n}$ est le bruit blanc gaussien ajouté au signal tatoué représentant une attaque de robustesse. Les expressions analytiques de α_{ISS} et de λ_{ISS} sont données par :

$$\alpha_{ISS}(i,\mathbf{x}) = \sqrt{1 - \frac{\lambda^2 N_c}{N_v 10^{\mathrm{WCR}/10}}}, \tag{1.18}$$

et :

$$\lambda_{ISS}(\mathbf{x}) = \frac{1}{2}\Bigg(1 + 10^{\mathrm{NCR}/10} + \frac{N_v 10^{\mathrm{WCR}/10}}{N_c} - \sqrt{\left(1 + 10^{\mathrm{NCR}/10} + \frac{N_v 10^{\mathrm{WCR}/10}}{N_c}\right)^2 - 4\frac{N_v 10^{\mathrm{WCR}/10}}{N_c}}\Bigg). \tag{1.19}$$

Si $\lambda_{ISS} = 0$, ce schéma est réduit au schéma classique SS. Si $\lambda_{ISS} = 1$, l'interférence du signal hôte est totalement annulée.

1.6 Attaques de sécurité en étalement de spectre

Les attaques de sécurité en étalement de spectre sont liées à l'estimation de la clé secrète par un ou plusieurs adversaires, la clé secrète utilisée est un ensemble de N_c porteuses $\{\mathbf{u}_i\}_{i\in[N_c]}$. Pratiquement, un adversaire tentera d'accéder

- à vect $\left(\{\mathbf{u}_i\}_{i\in[N_c]}\right)$, l'espace vectoriel engendré par les porteuses et sous-variété privée de dimension $N_d = N_c$,
- aux mots de codes ou zones de décodage déterminées par l'équation (1.14) au nombre de $N_h = 2^{N_c}$.

1.6.1 Séparation aveugle de sources

Nous avons vu précédemment que le problème de la sécurité dans un contexte WOA relève de la connaissance par un adversaire de plusieurs signaux tatoués avec la même clé secrète. En reprenant l'équation (1.6) et considérant les N_o contenus tatoués sous formes de vecteurs colonnes, nous obtenons la relation matricielle suivante :

$$\mathbf{Y} = \mathbf{X} + \mathbf{W} = \mathbf{X} + \mathbf{US}, \tag{1.20}$$

avec :

- $\mathbf{Y} \in \mathcal{M}_{N_v,N_o}(\mathbb{R})$ les signaux tatoués,
- $\mathbf{X} \in \mathcal{M}_{N_v,N_o}(\mathbb{R})$ les signaux hôtes,
- $\mathbf{W} \in \mathcal{M}_{N_v,N_o}(\mathbb{R})$ les signaux de tatouage,
- $\mathbf{U} \in \mathcal{M}_{N_v,N_c}(\mathbb{R})$ les porteuses secrètes,

– $\mathbf{S} \in \mathcal{M}_{N_c,N_o}(\mathbb{R})$ les modulations des messages insérés.

Le problème de l'estimation des matrices $\mathbf{U}$ et $\mathbf{S}$ à partir de $\mathbf{Y}$ est connu sous le nom de SAS (Séparation Aveugle de Sources ou *blind source separation*) avec $\mathbf{Y}$ les *observations*, $\mathbf{X}$ le *bruit*, $\mathbf{U}$ la *matrice de mélange* et enfin $\mathbf{S}$ les *sources*. Les attaques de sécurité en étalement de spectre consistent à obtenir le plus d'information possible sur $\mathbf{U}$: les porteuses $\{\mathbf{u}_i\}_{i\in[N_c]}$. L'ACP (Analyse en Composantes Principales) permet d'estimer $\mathrm{vect}(\mathbf{U})$, le sous-espace engendré par les porteuses secrètes. L'ACI (Analyse en Composantes Indépendantes) [40, 41] permet de résoudre le problème SAS quand les sources $\mathbf{S}$ sont statistiquement indépendantes. Dans le contexte WOA, nous posons l'hypothèse suivante : l'adversaire possède plusieurs signaux tatoués avec des messages tirés de façon aléatoire.

1.6.1.1 Analyse en Composantes Principales

L'ACP traite avec la *sous-espace-sécurité* et permet, pour un adversaire, d'estimer le sous-espace privé engendré par les porteuses si l'insertion altère la matrice de covariance des contenus tatoués $\mathbf{Y}$. L'ACP implique le calcul des valeurs propres de cette matrice de covariance et vise à trouver la transformation linéaire optimale permettant de garder le sous-espace possédant la plus grande variance (l'insertion des messages augmente la variance des signaux dans les directions des porteuses). Cette technique permet d'estimer $\mathrm{vect}\left(\{\mathbf{u}_i\}_{i\in[N_c]}\right)$. Une estimation réussie permet pour un adversaire la modification du message tout en minimisant la distorsion comme pour l'attaque explicitée figure 1.5. Nous utilisons la distance chordale [24,69] pour mesurer la précision de l'estimation du sous-espace secret. Si $\hat{\mathbf{U}}$ désigne les porteuses estimées (après orthogonalisation), la distance chordale est définie par :

$$\mathrm{d}_c = \frac{1}{\sqrt{N_c}} \left(\sum_{i=0}^{N_c-1} \sin^2(\theta_i) \right)^{1/2}, \tag{1.21}$$

où $\theta_0 \dots \theta_{N_c-1}$ sont les angles principaux entre $\mathbf{U}$ et $\hat{\mathbf{U}}$ [69]. La figure 1.8 illustre les angles principaux entre deux espace vectoriels.

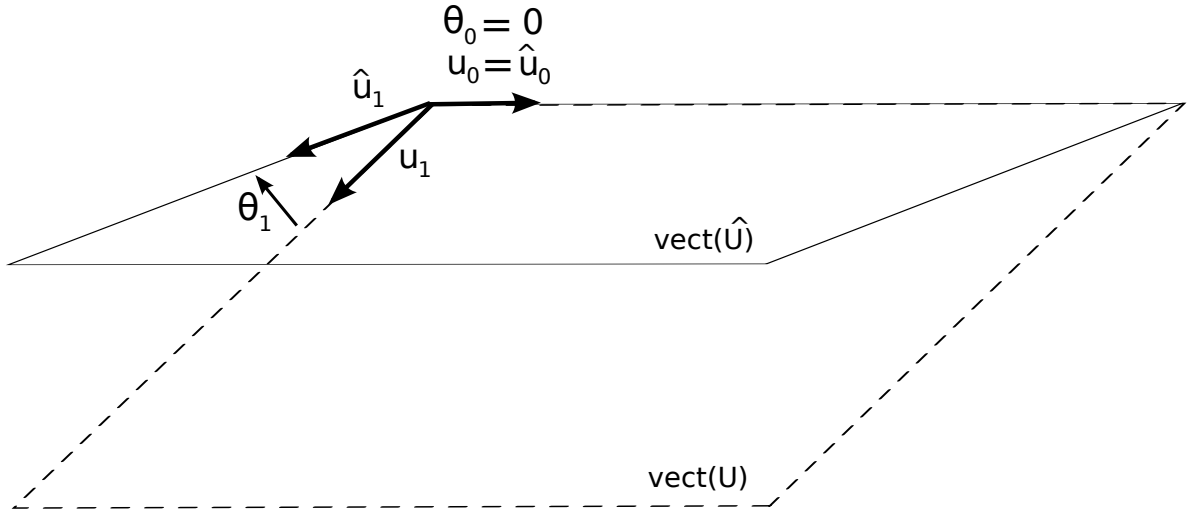


FIGURE 1.8 – Angles principaux $\theta_0 \dots \theta_{N_c-1}$ entre $\mathbf{U}$ et $\hat{\mathbf{U}}$, $N_c = 2$.

La distance chordale peut aussi être calculée en utilisant la norme de Frobenius pour les matrices :

$$\mathrm{d}_c\left(\mathbf{U}, \hat{\mathbf{U}}\right) = \frac{1}{\sqrt{2N_c}} \|\mathbf{U}^t\mathbf{U} - \hat{\mathbf{U}}^t\hat{\mathbf{U}}\|_F. \tag{1.22}$$

La distance chordale d_c vaut alors 0 quand les sous-espaces engendrés par les deux matrices sont égaux, et 1 (son maximum) les sous-espaces sont orthogonaux.

1.6.1.2 Analyse en Composantes Indépendantes

L'ACI [40,41] permet de décider si le schéma de tatouage appartient à la *clé-sécurité* ou bien à la *non-sécurité*. Elle permet, après une ACP, l'estimation des porteuses $\mathbf{U}$ quand les modulations sont indépendantes. L'ACI permet d'estimer les porteuses secrètes en maximisant l'indépendance statistique des composantes estimées pas à pas. Cependant, l'estimation par ACI est sujette à trois contraintes :

1. les porteuses sont estimées au signe près,
2. l'ordre des porteuses n'est pas connu. Dans le contexte WOA, la connaissance de quelques messages permet de lever cette indétermination (de l'ordre de $\log_2(N_c)$),
3. les porteuses ne peuvent être estimées si les sources ($\mathbf{S}$) sont distribuées selon une loi gaussienne ou sont dépendantes (ce qui n'est pas le cas avec les modulations SS et ISS).

De par ces contraintes, nous utilisons alors le score $\mathrm{S}_{\mathrm{est}}$ suivant, obtenu par un adversaire, afin de quantifier la précision de l'estimation des porteuses :

$$\mathrm{S}_{\mathrm{est}} = \frac{1}{N_c} \sum_i \left(\max\nolimits_j^1 |\mathrm{z}(\mathbf{u}_j, \hat{\mathbf{u}}_i)| - \max\nolimits_j^2 |\mathrm{z}(\mathbf{u}_j, \hat{\mathbf{u}}_i)| \right), \tag{1.23}$$

où $\{\hat{\mathbf{u}}_i\}_{i\in[N_c]}$ est l'ensemble des porteuses estimées par ACI, et $\max^{1,\text{ resp. }2}$ est le premier (resp. second) maximum de la valeur absolue entre la corrélation normalisée z entre chaque porteuse exacte $\mathbf{u}_i$ et chaque porteuse estimée $\hat{\mathbf{u}}_j$. Ce procédé a déjà été utilisé avec succès [8]. Comme $\{\mathbf{u}_i\}_{i\in[N_c]}$ représente une base du sous-espace secret de dimension N_c, nous pouvons conclure qu'une estimation correcte des porteuses entrainera un score S proche de 1 quand le nombre de signaux tatoués qu'un adversaire possède (N_o) est assez satisfaisant pour une ACI. Au contraire, une mauvaise estimation des porteuses produire un score S proche de 0, ou, de façon équivalente, un score qui ne converge pas lorsque N_o augmente.

1.6.1.3 À propos de la génération des porteuses

Lors de nos expériences, nous générons un ensemble de porteuses secrètes à l'aide d'un GNPA initialisé à partir d'une graine $K \in \mathbb{N}$ par la méthode suivante :

– le générateur de nombre pseudo-aléatoire MT19937 [62] permet de générer des nombres de façon uniforme,
– l'algorithme Ziggurat [52] permet une distribution gaussienne des nombres générés précédemment.

Le MT19937, variante du générateur Mersenne Twister, possède une période de $2^{19937-1}$ (nombre premier de Mersenne). Ce générateur est insuffisant pour des applications en cryptographie car des algorithmes tel Berlekamp-Massey [12, 53] permettent d'en prédire le comportement. La question qui se pose dans notre utilisation de ce générateur est de savoir s'il est possible, pour un pirate, de pouvoir regénerer les porteuses avec comme seule hypothèse la connaissance de la première d'entre elles (c'est-à-dire celle qui n'a pas été modifiée par l'algorithme de Gram-Schmidt).

En *clé-sécurité*, il n'est pas possible d'estimer les porteuses (on peut seulement estimer le sous-espace engendré par celles-ci). Le pirate ne peut donc connaître la première porteuse. Les possibles failles de sécurité liées au générateur aléatoire ne sont donc d'aucune utilité pour le pirate.

En *non-sécurité*, une estimation des porteuses est possible par ACI. Avant de pouvoir exploiter les failles du MT19937 le pirate se heurte à plusieurs barrières :

– la position de la première porteuse n'est pas connue et celle-ci est estimée au signe près, il doit alors effectuer $2N_c$ tests,
– la précision de l'estimation doit être sans failles,
– l'inversion de l'algorithme Ziggurat n'est pas évidente.

1.6.2 Attaque des schémas classiques par étalement de spectre

Nous avons vu à la section 1.5 deux modulations pour l'étalement de spectre, l'étalement de spectre classique (SS) et l'étalement de spectre amélioré (ISS). La figure 1.9 montre les distributions de N_o signaux hôtes et tatoués avec des messages tirés aléatoirement avec les modulations SS et ISS dans le sous-espace secret de dimension $N_c = 2$ (la distribution des corrélations entre les signaux hôtes $\mathbf{x}$ et $\mathbf{y}$ et deux porteuses secrètes). Les signaux hôtes ont été générés selon la loi normale $\mathcal{N}(0, 1)$. Les paramètres utilisés sont les suivants : $N_o = 2000$, $N_v = 512$, $N_c = 2$, $\mathrm{WCR} = -10\ dB$, $\mathrm{NCR} = -10\ dB$. Nous remarquons, pour les distributions SS et ISS, la présence de quatre "clusters" (c'est-à-dire une constellation en communications numériques). Ces clusters sont situés dans les mots de code : les régions de décodage définies en fonction du signe des corrélations par l'équation (1.14). Ils correspondent au $N_h = 2^{N_c}$ messages possibles : $(0, 0),\ (0, 1),\ (1, 0)$ et $(1, 1)$. Nous remarquons que, comme prévu par la construction de la modulation ISS, la variance des corrélations est diminuée pour améliorer la robustesse.

Ces modulations sont non-sûres (*non-sécurité*) [19]. A l'aide d'une ACP, un adversaire est capable de trouver le sous-espace engendré par les porteuses ; par une ACI, il a accès au zones de décodage. Il peut alors altérer les messages des contenus tout en minimisant la distorsion avec une probabilité de 1 (attaque déterministe) dans le sous-espace secret comme le montre la figure 1.10. Des expériences illustrant la propriété de *non-sécurité* des modulations SS et ISS sont proposées dans la section 3.2.3 lors d'une implantation des techniques par étalement de spectre sur des images naturelles. Ces expériences utilisent le score $\mathrm{S}_{\mathrm{est}}$ (équation (1.23)) pour l'ACI.

1.7 Méthodes sûres par étalement de spectre

Cette partie présente des schémas de tatouage multi-bits permettant d'atteindre la *clé-sécurité*, *sous-espace-sécurité* ou *stégo-sécurité* dans le cadre du tatouage par étalement de spectre.

1.7.1 Tatouage circulaire (CW)

Le tatouage circulaire CW (*Circular Watermarking*) est une modulation basée sur l'ISS. En reprenant les notations de la section sur l'étalement de spectre, la formule 1.5 s'utilise avec :

$$\alpha_{CW}(i, \mathbf{x}) = \mathbf{d}(i)\alpha_{ISS}(i, \mathbf{x}), \tag{1.24}$$

et :

$$\lambda_{CW}(\mathbf{x}) = \lambda_{ISS}(\mathbf{x}). \tag{1.25}$$

Le paramètre $\mathbf{d} \in \mathbb{R}^{N_c}$ permet de générer des points sur une hyper-sphère unitaire de dimension $N_c - 1$. Ses composantes sont issues d'un vecteur aléatoire gaussien $\mathbf{g}$ centré réduit normalisé :

$$\forall i \in [N_c],\ \mathbf{d}(i) = \frac{|\mathbf{g}(i)|}{\|\mathbf{g}\|}, \tag{1.26}$$

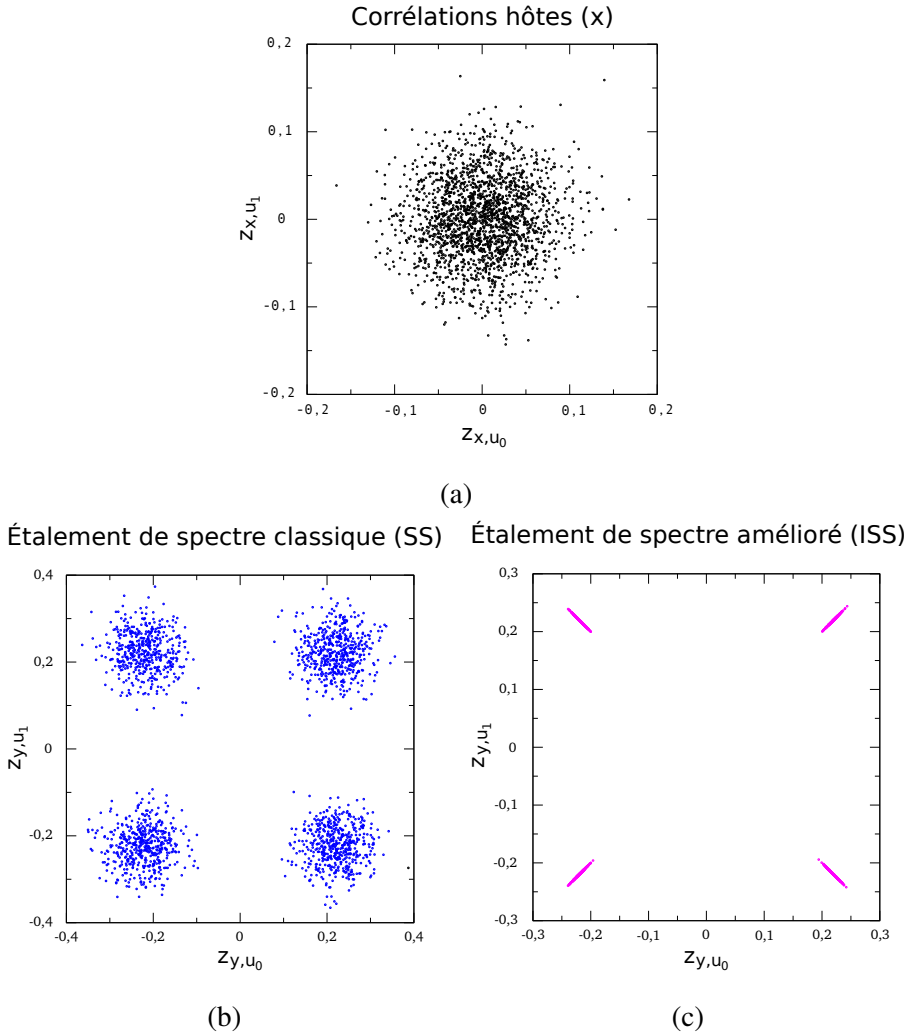


FIGURE 1.9 – Corrélations normalisées entre les signaux hôtes (a), SS (b) et ISS (c) et deux porteuses secrètes $\{\mathbf{u}_0, \mathbf{u}_1\}$: sous-espace secret de dimension N_c. Paramètres $N_c = 2,\ N_v = 512,\ \mathrm{WCR} = -10\ dB,\ \mathrm{NCR} = -10\ dB$. Nous remarquons la présence de quatre clusters situés dans les $N_h = 2^{N_c} = 4$ mots de codes. De par la construction de la modulation ISS, la variance des corrélations est diminuée pour améliorer la robustesse.

permettant ainsi d'étaler de façon circulaire les corrélations des signaux tatoués dans la région du mot de code désiré pour le tatouage. La modulation CW respecte la propriété suivante, dite de *circularité* [6] :

$$\mathrm{p}\left(\mathrm{z}_{\mathbf{y},\mathbf{u}_0}, \ldots, \mathrm{z}_{\mathbf{y},\mathbf{u}_{N_c-1}}\right) = \mathrm{p}\left(\sqrt{\sum_{i=0}^{N_c-1} \mathrm{z}^2_{\mathbf{y},\mathbf{u}_i}}\right). \tag{1.27}$$

La propriété de circularité signifie que la distribution des corrélations peut être réduite à une distribution qui dépend uniquement de la norme euclidienne de ces corrélations. Cette propriété est une condition nécessaire de la *clé-sécurité*. C'est-à-dire que pour un sous-ensemble de clés (bases de $\mathrm{vect}\,(\{\mathbf{u}_i\})$), la distribution des signaux tatoués (conditionnellement à la clé secrète) sera identique. Un adversaire peut avoir accès à ce sous-espace mais n'aura aucun moyen de déterminer les mots de code.

Toutefois, dans cette implantation du CW, les corrélations peuvent être étalées sur les bords de mauvaises

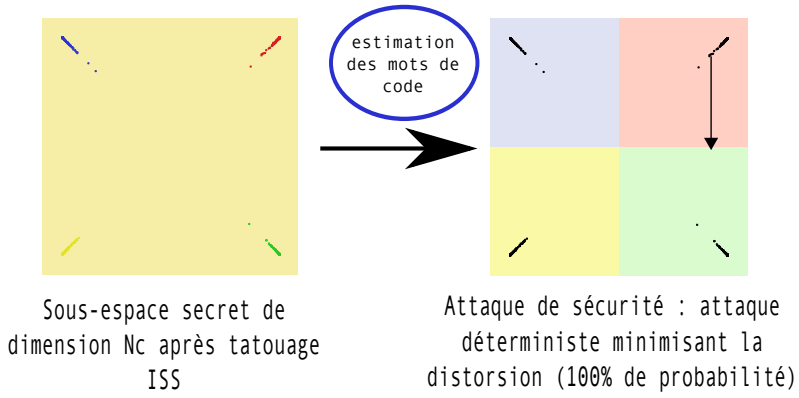


FIGURE 1.10 – Altération du message d'un contenu tatoué (ISS) par le déplacement de ses corrélations avec les porteuses secrètes vers une région de décodage différente après estimation des mots de codes ($N_c = 2$). SS et ISS sont des modulations *non-sûres*.

régions de mot de code (dépassement dû à l'interférence du signal hôte). Ce problème peut occasionnellement entrainer des erreurs lors du décodage (même lorsqu'il n'y a pas d'attaques). Nous proposons, dans ce manuscrit de thèse, de remplacer l'implantation du CW par une nouvelle version stochastique, appelée "zero-error-bit" CW. Le procédé consiste à générer aléatoirement une nouvelle perturbation $\mathbf{d}$ jusqu'à ce que la corrélation d'un contenu tatoué $\mathbf{y}$ se situe à l'intérieur de la région de mot de code relative au message après insertion (contrôlé par l'équation (1.14)). Notons que la complexité de ce procédé dépend du rejet de l'interférence du signal hôte. Si $\lambda(\mathbf{x}) = 1$, le contenu tatoué sera toujours dans la bonne zone de décodage. Nous obtenons alors une implantation du tatouage circulaire sans erreur au décodage dans un contexte sans attaque, sans changer la fonction de densité de la distribution des $\mathbf{y}$.

La figure 1.11 montre la distributions de N_o signaux hôtes et tatoués avec des messages tirés aléatoirement avec la modulation CW dans le sous-espace secret de dimension $N_c = 2$. Les signaux hôtes ont été générés selon la loi $\mathcal{N}(0, 1)$.

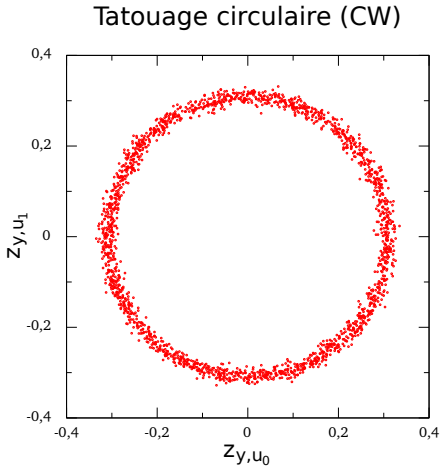


FIGURE 1.11 – Corrélations normalisées entre les signaux tatoués par CW et deux porteuses secrètes $\{\mathbf{u}_0, \mathbf{u}_1\}$: sous-espace secret de dimension N_c. Paramètres : $N_c = 2,\ N_v = 512,\ \mathrm{WCR} = -10\ dB,\ \mathrm{NCR} = -10\ dB$.

La modulation CW est **clé-sûre** [6]. Grâce à la circularité de la distribution, pour toute base de vect $\left(\{\mathbf{u}_i\}_{i\in[N_c]}\right)$ (rotations du sous-espace privé), la distribution est identique. Un adversaire peut estimer le sous-espace secret de dimension N_c, par exemple par ACP, mais ne peut localiser les régions de décodage. Il peut alors modifier les corrélations par translations dans l'espace privé (ajout de bruit gaussien, symétries centrales dans le sous-espace secret) mais ne peut être sûr de modifier le message présent comme le montre la figure 1.12. De plus ces modifications entraineront une grande distorsion (contrairement aux modulations SS ou ISS où la distorsion est minimisée pour l'attaque).

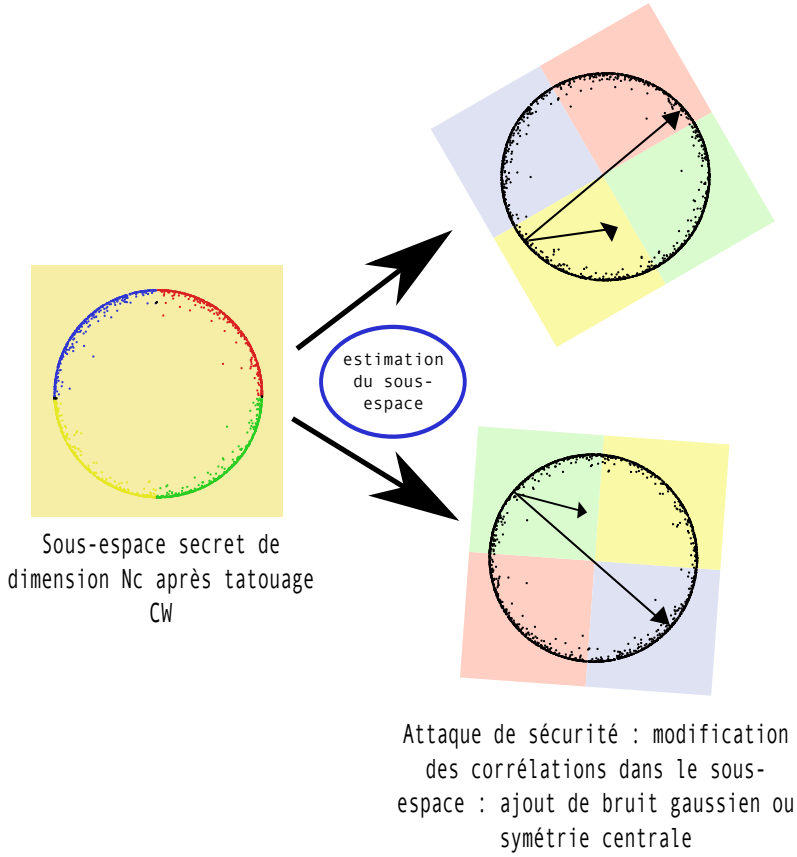


FIGURE 1.12 – Estimation du sous-espace secret ($N_c = 2$) après tatouage CW de plusieurs contenus. L'adversaire ne peut estimer les régions de mots de code de par la circularité de la modulation. L'adversaire peut effectuer des attaques à l'intérieur du sous-espace mais au détriment de la distorsion et sans savoir (a priori) que l'attaque est un succès. CW est une modulation *clé-sûre*.

1.7.2 Tatouage naturel (NW & RNW)

Dans [7], deux nouvelles modulations, le tatouage naturel NW (*Natural Watermarking*) et le tatouage naturel robuste RNW (*Robust Natural Watermarking*) sont proposées. Leur but est de préserver la distribution des corrélations $\mathrm{z}_{x,\mathbf{u}_i}$ supposée circulaire (vérifiant l'équation (1.27)) après tatouage. Cette distribution peut être mise à l'échelle selon un facteur η afin de fixer le budget de distorsion conditionné par l'application pratique visée. Notons que nous ne pouvons définir les modulations NW et RNW uniquement lorsque la répartition des contenus hôtes est circulaire, il peut s'agit d'une distribution gaussienne ou encore laplacienne (ondelettes d'une image).

En reprenant l'équation (1.5), nous avons :

$$\alpha_{NW}(i, \mathbf{x}) = \eta |\mathrm{z}_{\mathbf{x},\mathbf{u}_i}|, \tag{1.28}$$

et :

$$\lambda_{NW}(\mathbf{x}) = 1, \tag{1.29}$$

avec :

$$\eta = \sqrt{\frac{N_v - 1}{N_c} 10^{\mathrm{WCR}/10} - 1}. \tag{1.30}$$

Pour chaque porteuse $\mathbf{u}_i$, nous obtenons :

$$|\mathrm{z}_{\mathbf{y},\mathbf{u}_i}| = |\eta| |\mathrm{z}_{\mathbf{x},\mathbf{u}_i}|. \tag{1.31}$$

Lorsque $\eta = 1$, nous parlons de tatouage naturel (NW), la distribution des signaux hôtes et tatoués dans le sous-espace engendré par les porteuses est identique. Nous effectuons une symétrie centrale si la corrélation concernée n'est pas du même signe que celle correspondant au bit à cacher (figure 1.13). Cette symétrie permet de conserver la distribution des corrélations du contenu hôte après insertion, nous avons l'égalité :

$$\forall i \in [N_c],\ \mathrm{p}(\langle \mathrm{z}_{\mathbf{x},\mathbf{u}_i} \rangle) = \mathrm{p}(\langle \mathrm{z}_{\mathbf{y},\mathbf{u}_i} \rangle). \tag{1.32}$$

où p désigne la distribution de probabilité.

Cette propriété permet une méthode de tatouage *stégo-sûre* : théoriquement, un adversaire ne peut certifier que les signaux sont tatoués. La variance des signaux étant inchangée, une ACP est alors impossible [4]. Cependant, il n'est pas possible de fixer au préalable la distorsion engendrée par l'ajout du signal de tatouage. Si le contenu hôte se situe déjà dans la région de décodage correspondant au message à insérer, aucun signal de tatouage n'est ajouté, nous obtenons alors $\mathrm{WCR} = -\infty$.

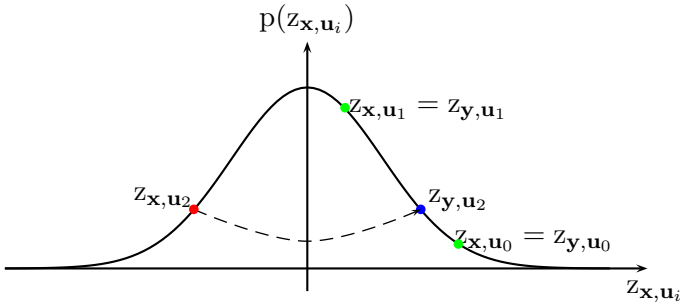


FIGURE 1.13 – Tatouage Naturel (NW) avec $\mathbf{m} = \{1, 1, 1\}$ ($N_c = 3$). Seul le troisième bit du message nécessite une symétrie centrale.

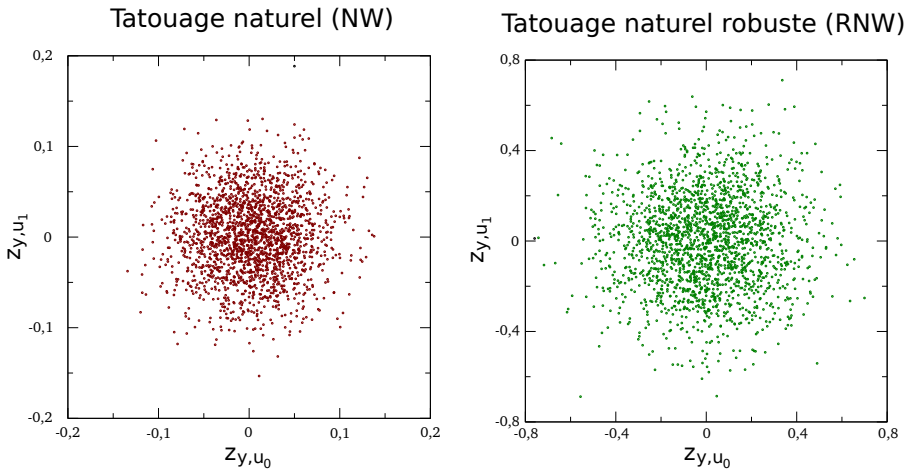
Lorsque $\eta > 1$, nous parlons de tatouage naturel robuste (RNW), la robustesse de ce schéma augmentant avec la valeur de η. Sous l'hypothèse d'une distribution circulaire des corrélations entre les signaux hôtes et les porteuses, les classes de sécurité de ces schémas dans le cadre WOA sont définies par :

- si $N_c = N_v$:
  - si $\eta = 1$, alors NW est stégo-sûr,
  - si $\eta > 1$, RNW est sous-espace-sûr,

4. Nous supposons une distribution des contenus hôtes circulaire, pratiquement cette hypothèse n'est pas forcément vérifiée, voir l'implantation sur des images naturelles dans le chapitre 3.

- si $N_c < N_v$:
  - si $\eta = 1$, alors NW est sous-espace-sûr,
  - si $\eta > 1$, RNW est clé-sûr.

La figure 1.14 montre la distribution de N_o signaux tatoués avec des messages tirés aléatoirement avec les modulations NW et RNW dans le sous-espace secret de dimension $N_c = 2$. Les signaux hôtes ont été générés selon la loi normale $\mathcal{N}(0, 1)$ qui est circulaire. Les paramètres utilisés sont les suivants : $N_o = 2000$, $N_v = 512$, $N_c = 2$. Pour RNW, la distorsion est fixée à $\mathrm{WCR} = -10\ dB$.



(a) (b)

FIGURE 1.14 – Corrélations normalisées entre les signaux tatoués par NW (a) et RNW (b) et deux porteuses secrètes $\{\mathbf{u}_0, \mathbf{u}_1\}$: sous-espace secret de dimension N_c. Paramètres $N_c = 2,\ N_v = 512$.

1.8 Conclusion

Dans ce chapitre nous avons abordé la notion de sécurité en tatouage numérique en prenant comme exemple le tatouage par étalement de spectre. Plusieurs points ont été abordés :

i) La notion de sécurité est ici liée au principe de Kerckhoffs : la clé secrète est l'unique paramètre inconnu des adversaires. Cependant, l'utilisation d'une clé secrète n'implique pas qu'une méthode de tatouage est sûre.

ii) Dans le cadre WOA (l'adversaire a accès a plusieurs contenus tatoués avec des messages différents et la même clé), nous distinguons quatre classes de sécurité. Les méthodes d'étalement de spectre classiques (SS et ISS) sont *non-sûres*, un attaquant peut distinguer les régions de décodage de chaque message et ainsi estimer la clé secrète. Les modulations CW, RNW et NW atteignent la *clé-sécurité* de par leur circularité (équation (1.27)). Nous voyons alors que la sécurité des schémas dépend de la distribution des contenus, la contrainte de distorsion n'étant pas prise en compte. Il est alors possible de construire des méthodes sûres par étalement de spectre, non plus en utilisant les formules d'insertion classiques, mais plutôt en générant directement des distributions (données par le niveau de sécurité souhaité) dans le domaine tatoué. Cette nouvelle approche du tatouage s'appelle le **tatouage sûr basé sur modèle** et est similaire à celle proposée par Sallee en stéganographie [74].

Chapitre 2

Insertions sûres construites à partir de modèles statistiques

Gaspard Monge (1746 - 1818), mathématicien français dont les contributions concernent la géométrie descriptive, l'analyse infinitésimale et la géométrie analytique.

Sommaire

DANS ce chapitre sont développées nos contributions en tatouage sûr de signaux réalisées pendant l'ensemble du doctorat. Nous présentons deux nouvelles techniques de tatouage : le **tatouage ρ circulaire**, modulation par étalement de spectre, permettant une flexibilité entre la clé-sécurité et la non-sécurité, le compromis entre sécurité et robustesse étant fixé par le cadre applicatif et le **tatouage par la loi du χ^2**, permettant la stégo-sécurité. La contrepartie de cette dernière méthode est sa faible résistance face aux attaques de robustesse. En effet, celle-ci est limitée en comparaison avec les techniques sûres par étalement de spectre. Les performances des schémas de tatouage étant liées aux contraintes de robustesse, d'imperceptibilité et de sécurité, nous présentons ensuite deux améliorations des schémas sûrs par étalement de spectre. En gardant les mêmes performances en terme de distorsion et de robustesse, nous sommes en mesure de minimiser la distorsion provoquée par l'ajout du tatouage grâce à deux techniques :

i) l'algorithme des Hongrois, méthode discrète permettant de minimiser la distorsion globale entre nuages de points,

ii) la théorie du transport, méthode continue minimisant le poids de déplacement entre deux distributions, la fonction de densité étant connue.

2.1 Nouvelles techniques de tatouage sûr

Nous nous intéressons dans cette première section à la construction de nouvelles méthodes de tatouage dont la sécurité s'inscrit dans le cadre WOA, en effet les modulations CW, NW et RNW vues dans le chapitre précédent ont leurs limites :

- les techniques NW et RNW (pouvant atteindre la *stégo-sécurité*, la *sous-espace-sécurité* ou la *clé-sécurité*) sont malheureusement faibles du point de vue de la robustesse et peuvent alors êtres difficiles à utiliser dans un cadre pratique (exemple de tatouage d'images naturelles, section 3.3.2),
- la modulation CW ne peut être appliquée lorsque $N_c = 1$ bit, cette restriction peut poser problème dans le cadre pratique de l'estampillage de contenus (cadre présenté dans le chapitre 4).

Nous présentons ici deux nouvelles techniques de tatouage : le tatouage ρ circulaire par étalement de spectre et le tatouage par la loi du χ^2 agissant sur la norme euclidienne au carré des signaux.

2.1.1 Tatouage ρ circulaire (ρ-CW)

Le tatouage ρ circulaire ρ-CW (ρ *Circular Watermarking*) est une nouvelle modulation par étalement de spectre basée sur la modulation CW. Les corrélations sont étalées sur la région de mot de code correspondant au message à insérer. Cependant ces corrélations sont translatées suivant un vecteur ρ sur chaque dimension et ensuite normalisées. Formellement, suivant l'équation (1.5), nous obtenons :

$$\alpha_{\rho\text{-CW}}(i, \mathbf{x}) = \mathbf{d}_\rho(i)\alpha_{\text{ISS}}(i, \mathbf{x}), \tag{2.1}$$

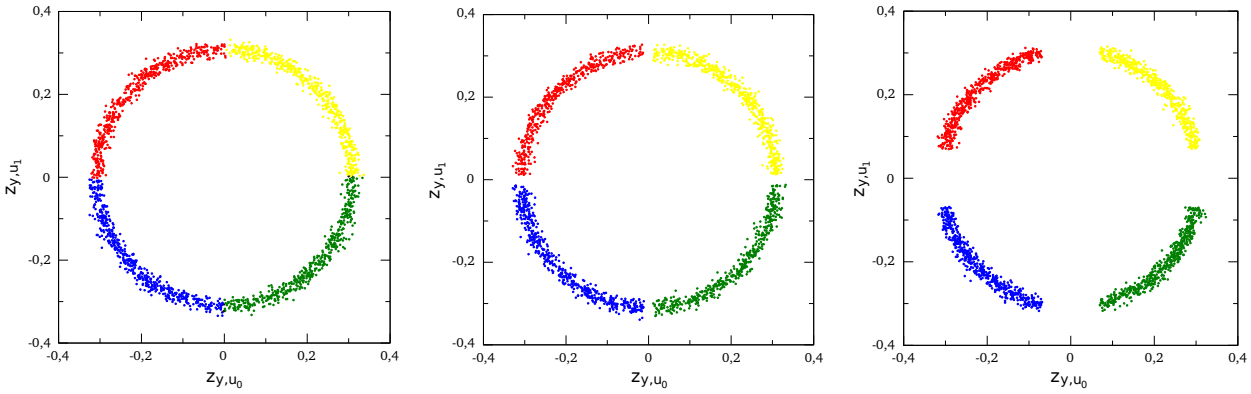
avec :

$$\mathbf{d}_\rho \in \mathbb{R}^{N_c},\ \mathbf{d}_\rho = (\mathbf{d} + \rho.\mathbf{1})/\|\mathbf{d} + \rho.\mathbf{1}\|, \tag{2.2}$$

où $\mathbf{1}$ désigne le vecteur constant dont chaque composante vaut 1.

Quand $\rho >> 0$, ρ-CW est une modulation non-sûre ; quand $\rho = 0$, ρ-CW désigne la modulation CW. Cette modulation permet la flexibilité entre robustesse et sécurité. La qualité de l'estimation des porteuses par un attaquant est liée à la distorsion de l'insertion (WCR) et au nombre d'observations (N_o). Cette modulation peut être vue comme un trait d'union entre la *clé-sécurité* et la *non-sécurité*.

La figure 2.1 montre la distribution de N_o signaux tatoués avec des messages tirés aléatoirement avec la modulation ρ-CW dans le sous-espace secret de dimension $N_c = 2$. Les signaux hôtes ont été générés selon la loi normale $\mathcal{N}(0, 1)$. Les paramètres utilisés sont les suivants : $N_o = 2000$, $N_v = 512$, $N_c = 2$, $\mathrm{WCR} = -10\ dB$, $\mathrm{NCR} = -10\ dB$.



$\rho = 0.$ $\rho = 0.04$ $\rho = 0.3$

FIGURE 2.1 – Corrélations normalisées entre les signaux tatoués par ρ-CW et deux porteuses secrètes $\{\mathbf{u}_0, \mathbf{u}_1\}$: sous-espace secret de dimension N_c pour $\rho = 0, 0.04, 0.3$. Paramètres $N_c = 2$, $N_v = 512$, $\mathrm{WCR} = -10\ dB$, $\mathrm{NCR} = -10\ dB$.

Nous remarquons l'éloignement entre les différentes régions de décodage qui augmente avec la valeur du scalaire ρ. Cette représentation des corrélations illustre une fois de plus le compromis sécurité/robustesse des schémas de tatouage. Plus les régions de décodage sont éloignées, plus le tatouage sera résistant aux attaques de robustesse. En contrepartie, il sera plus facile pour un attaquant d'estimer les axes des porteuses secrètes (sous l'hypothèse de posséder une collection de contenus tatoués).

Pour quantifier pratiquement la sécurité du ρ-CW en fonction de ρ, nous simulons ici le tatouage de N_o signaux gaussiens. Dans le cadre WOA, un adversaire applique ensuite une ACI sur ces signaux pour estimer les porteuses secrètes $\{\mathbf{u}_i\}_{i\in[N_c]}$. Grâce aux porteuses obtenues, il est alors en mesure d'estimer les bits insérés. Nous appelons ici ϵ l'erreur binaire d'estimation.

La figure 2.2 montre la valeur de ϵ en fonction du nombre de contenus tatoués possédés par l'adversaire pour $\rho = 0$ (CW classique), 0.04, 0.1, 0.4 ainsi que pour la modulation ISS. Nous avons vu que l'ACI était sujette à deux limites (section 1.6.1.2) : les porteuses sont estimées au signe près et l'ordre des porteuses n'est pas connu par l'adversaire. Pour cette expérience, nous supposons que l'adversaire est capable de soulever ces indéterminations. De plus les valeurs de ϵ sont calculées en moyenne sur 10 clés secrètes différentes. Les paramètres sont ici : $N_c = 16$ $N_v = 512$, $\mathrm{WCR} = -10\ dB$, $\mathrm{NCR} = -10\ dB$.

Comme nous pouvons le constater, lorsque $\rho = 0$, nous utilisons le CW classique et il n'est alors pas possible pour l'adversaire d'estimer les bits insérés même avec un grand nombre de contenus : ϵ reste stationnaire à 0.34. Cependant, lorsque ρ augmente, l'estimation des messages est de plus en plus précise avec un nombre de contenus nécessaires qui diminue, se rapprochant de la *non-sécurité* de l'ISS. L'estimation ϵ dépend alors de la valeur de ρ et du nombre de contenus N_o.

Nous testons à présent la robustesse du ρ-CW pour plusieurs valeurs de ρ ainsi que pour la modulation ISS. La figure 2.3 montre le taux d'erreur binaire (BER) en fonction du WNR pour ces modulations. Comme nous pouvons le constater, la robustesse augmente lorsque le paramètre ρ augmente pour se rapprocher de l'ISS.

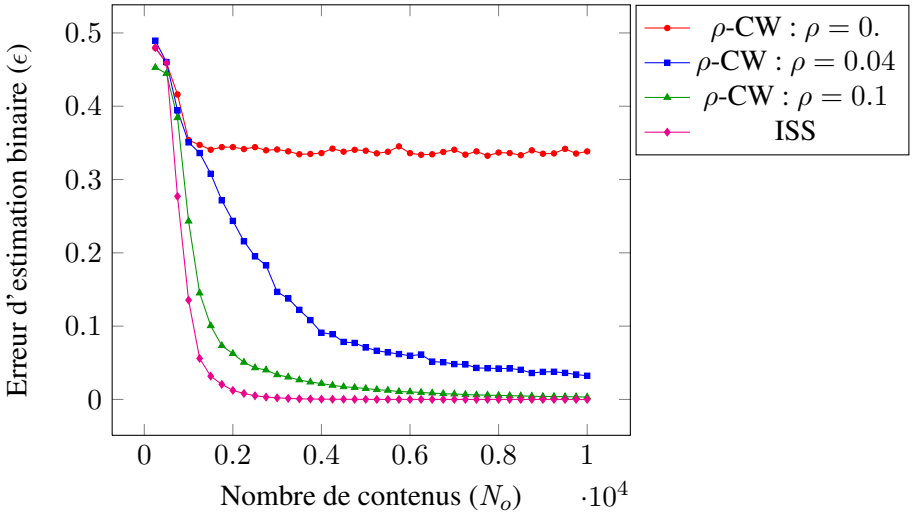


FIGURE 2.2 – Erreur d'estimation binaire en fonction du nombre de contenus tatoués possédés par un adversaire utilisés dans le cadre d'une ACI. Paramètres : $N_c = 16$ $N_v = 512,$ $\text{WCR} = -10\ dB,$ $\text{NCR} = -10\ dB$.

Les expériences réalisées sur la modulation ρ-CW reflètent le compromis entre sécurité et robustesse que subissent les schémas de tatouage. Ils valident aussi la flexibilité de cette nouvelle modulation ρ-CW, hybride entre la modulation ISS (très robuste et non-sûre) et la modulation CW (peu robuste et clé-sûre).

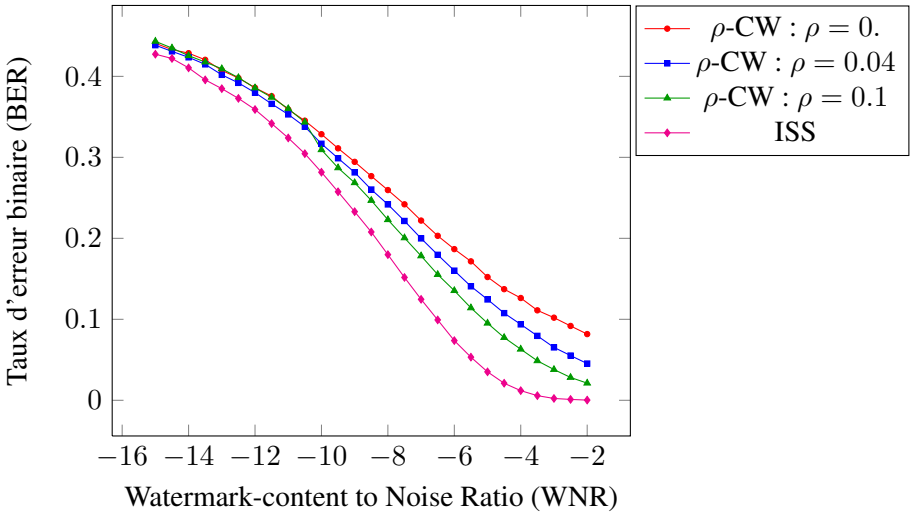


FIGURE 2.3 – Taux d'erreur binaire (BER) en fonction du WNR pour ISS et ρ-CW. Paramètres : $N_c = 16$ $N_v = 512,$ $\text{WCR} = -10\ dB,$ $\text{NCR} = -10\ dB$. Comme nous pouvons le constater, la robustesse augmente avec ρ et reflète (avec la figure 2.2) le compromis entre sécurité et robustesse des schémas de tatouage.

2.1.2 Tatouage par la loi du χ^2 (χ^2W)

2.1.2.1 Construction

En se basant sur les définitions des classes de sécurité (section 1.4), nous proposons un schéma de tatouage *stégo-sûr* appelé tatouage par la loi du CHI^2, χ^2W (*CHI^2 Watermarking*). Cette technique modifie la norme

euclidienne des signaux hôtes (supposés gaussiens) après insertion, tout en gardant la même distribution entre corrélations hôtes et corrélations tatouées. La distribution de ces normes peut être modélisée par une loi du χ^2. Les mots de code de la clé secrète sont alors définis par des intervalles réels ($\subset \mathbb{R}$) dans l'ensemble des normes de vecteurs gaussiens.

Pour insérer un message $\mathbf{m} \in \mathbb{F}_2^{N_c}$, nous choisissons une norme dans un intervalle réel (mot de code) codant le message, nous multiplions ensuite le signal hôte $\mathbf{x}$ par un scalaire α afin d'avoir la norme désirée pour le signal $\mathbf{y}$. Contrairement aux schémas par étalement de spectre, le sous-espace de tatouage (espace des normes euclidiennes) n'est pas privé, le secret est entièrement défini par la partition de l'axe positif réel représentant les normes des signaux, ce qui ne cause aucune faille de sécurité dans le contexte d'attaque WOA.

Cette insertion est facile à implanter, elle permet la *stégo-sécurité* car la distribution des signaux hôtes et tatoués est identique. Nous obtenons le signal tatoué $\mathbf{y} = \alpha\mathbf{x}$, $\alpha \in \mathbb{R}^+$. Cette méthode est basée sur une distribution de normes de signaux hôtes gaussiens. Si $\mathbf{x} \sim \mathcal{N}(0,1)$, alors $\|\mathbf{x}\|^2 \sim \chi^2(N_v)$. Les mots de code sont représentés par une partition de l'axe réel positif $\mathbb{R}^+$. Pour insérer un message $\mathbf{m}$ dans un signal hôte $\mathbf{x}$, nous choisissons aléatoirement une norme $\|\mathbf{y}\|^2$ dans l'intervalle correspondant à un mot de code du message souhaité et nous construisons :

$$\mathbf{y} = \sqrt{\frac{\|\mathbf{y}\|^2}{\|\mathbf{x}\|^2}}\mathbf{x}. \tag{2.3}$$

Le signal de tatouage $\mathbf{w}$ est alors donné par $\mathbf{w} = \mathbf{y} - \mathbf{x}$. Ce procédé peut être considéré comme une variante de l'insertion stochastique de Moulin et Briassouli [66] qui s'utilise sur différentes distributions hôtes. La figure 2.4 illustre les régions de décodages lors d'insertions de messages de longueur $N_c = 2$ bits, d'une part par étalement de spectre, d'autre part par tatouage χ^2W. Nous remarquons alors que, contrairement aux techniques par étalement de spectre, l'insertion de messages par la méthode χ^2W agit sur la norme euclidienne des signaux hôtes, les frontières des zones de décodages sont définies comme des seuils dans l'intervalle réel $\mathbb{R}^+$.

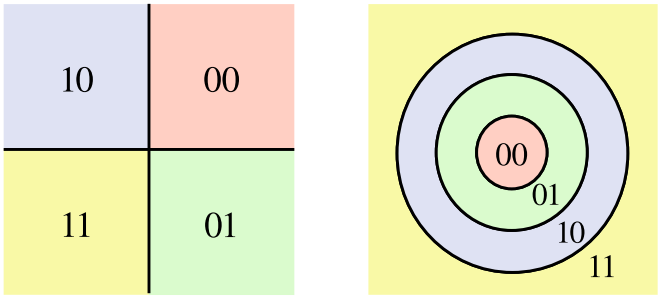


FIGURE 2.4 – Représentations en 2 dimensions des régions de décodage pour l'étalement de spectre (gauche) et pour le χ^2W (droite) pour $N_c = 2$ bits. Les zones de décodage pour le χ^2W sont définies par des intervalles $\subset \mathbb{R}^+$.

Les conséquences d'une telle insertion sont les suivantes :

i) La manière de choisir une norme dans le mot de code désiré n'est pas optimale du point de vue de la distorsion : nous générons des nombres réels de manière aléatoire dans l'intervalle du mot de code désiré ;

ii) Pour définir la partition secrète de $\mathbb{R}^+$, nous utilisons un estimateur de la fonction quantile de la loi du χ^2 ;

iii) Chaque zone de décodage du message doit avoir la même probabilité : en tirant une norme de façon aléatoire dans $\mathbb{R}^+$, nous devons pouvoir générer un signal tatoué dont la nature du message inséré doit apparaître avec une probabilité de $\frac{1}{2^{N_c}}$ (exemple figure 2.5). Cependant, sous cette condition, il faut avoir plusieurs

mots de code pour un message. Sans cette condition, un adversaire peut alors estimer les frontières de décodage en utilisant une fonction quantile : il sépare alors la fonction de répartition des normes en 2^{N_c} parties équiprobables de probabilité $\frac{1}{2^{N_c}}$. Nous avons alors la condition :

$$N_h > 2^{N_c}, \tag{2.4}$$

afin d'éviter une faille de sécurité.

La figure 2.6 montre un exemple de partition secrète dans $\mathbb{R}^+$ pour l'application du χ^2W avec $N_c = 2$ et $N_h = 8$ (deux mots de code pour un message). De par l'équation 2.4, un adversaire ne peut estimer les zones de décodage.

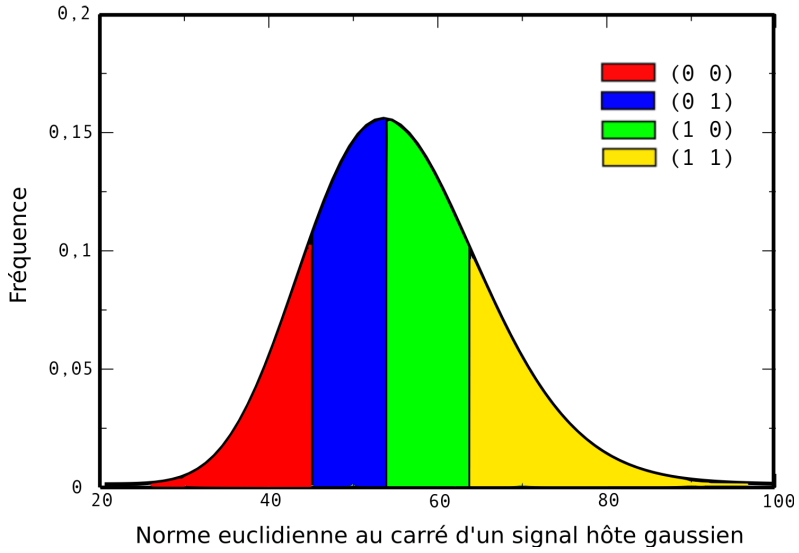


FIGURE 2.5 – Mauvais exemple de partition secrète pour χ^2W : distribution des normes euclidiennes au carré de signaux gaussiens. Ces normes suivent une loi $\chi^2(N_v)$, $N_v = 55$. De plus nous avons $N_h = 2^{N_c} = 2^2 = 4$ (un mot de code pour chaque message). Les zones de décodage apparaissent avec la même probabilité : les aires de couleurs sont égales. Sous cette condition d'équiprobabilité, une **faille de sécurité apparaît** : un adversaire peut alors délimiter les réels correspondants aux frontières de décodage (qui ne dépendent pas de la clé secrète et qui peuvent être déterminés à l'aide d'une fonction quantile.)

2.1.2.2 Implantation sur signaux gaussiens

Dans cette sous-section, nous avons modélisé le χ^2W avec $N_c = 2,\ N_v = 55,\ \mathrm{WCR} = -18\ dB,\ N_h = 8$ et $N_o = 2000$ observations (tatouage de N_o signaux gaussiens). Afin d'obtenir la condition d'équiprobabilité, nous utilisons un estimateur de la fonction quantile de la loi du χ^2 donné par [38], lui-même utilisant un estimateur de la fonction de répartition de la loi normale donné par [42]. Nous utilisons comme clé secrète pour l'insertion, la partition définie dans la table 2.1 et représentée figure 2.7.

La figure 2.8 montre les projections sur les deux premières composantes de signaux gaussiens hôtes et tatoués par χ^2W avec les paramètres $N_c = 2,\ N_v = 55,\ \mathrm{WCR} = -18\ dB,\ N_h = 8$ et $N_o = 2000$. Nous remarquons que, comme pour la modulation NW (figure 1.14), la distribution des signaux ne change pas après tatouage, illustrant la *stégo-sécurité* du χ^2W.

Nous testons à présent la robustesse du χ^2W face à l'ajout de bruit gaussien (quantifié ici par le WCNR). La figure 2.9 montre le taux d'erreur binaire (BER) en espérance sur $N_o = 2000$ signaux tatoués en fonction d'un

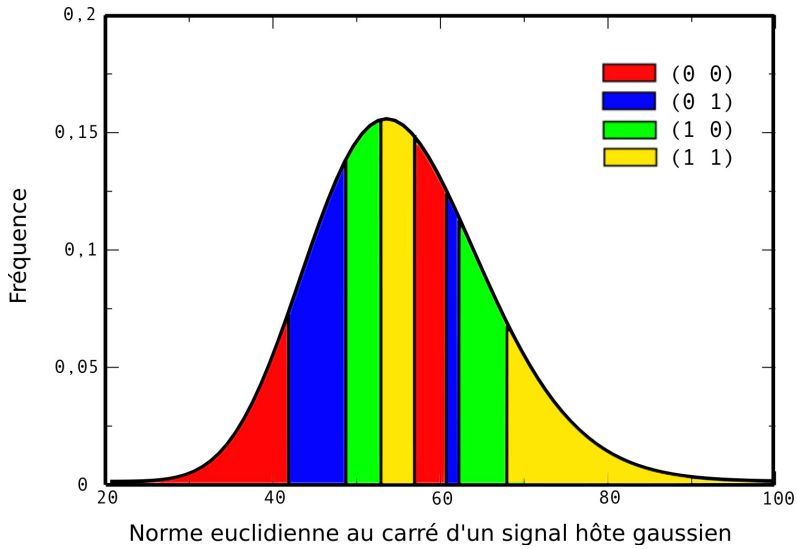


FIGURE 2.6 – Bon exemple de partition secrète pour χ^2W. Nous observons une distribution des normes euclidiennes au carré de signaux gaussiens. Ces normes suivent une loi $\chi^2(N_v)$, $N_v = 55$. De plus nous avons $N_h = 8 \geq 2^{N_c} = 2^2 = 4$ (deux mots de code pour un message). Chaque message apparaît avec la même probabilité mais grâce à la condition donnée par 2.4, les frontières de décodage dépendent d'une clé secrète. L'adversaire ne peut alors déterminer les mots de code, cette construction est *stégo-sûre*.

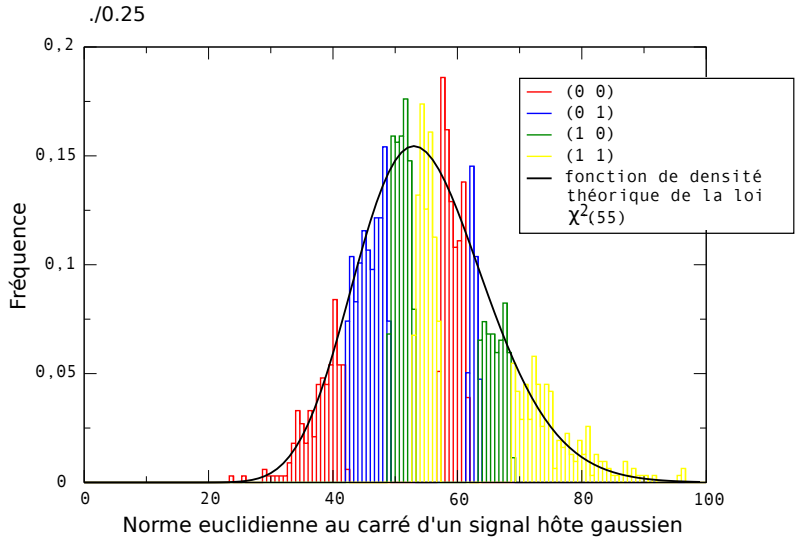
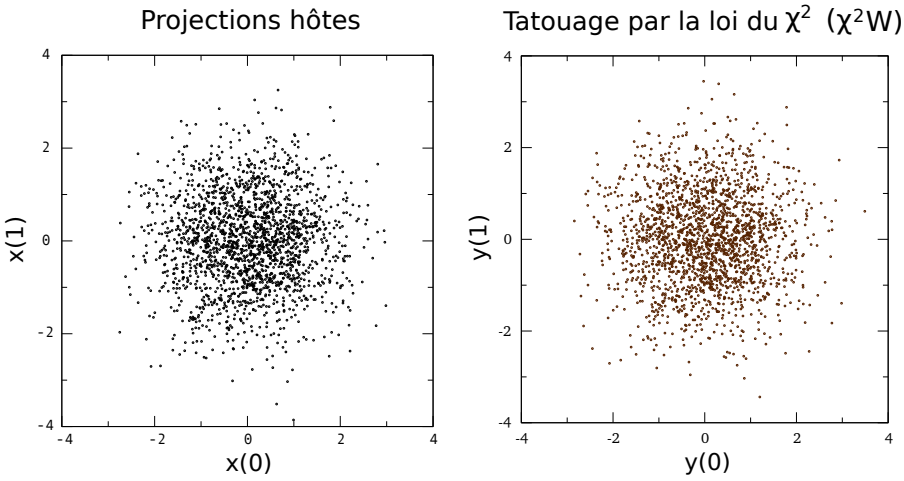


FIGURE 2.7 – Partition secrète pour χ^2W et distribution associé, histogramme généré avec les paramètres $N_c = 2, N_w = 8, N_v = 55$. Cette partition est la même que celle définie par la table 2.1, chaque message est codé par deux mots de code. Nous avons généré ici $N_o = 2000$ vecteurs gaussiens pour chaque message et avons calculé leur norme respective.

WCNR cible pour les deux schémas par étalement de spectre NW et CW ainsi que pour le χ^2W. Nous cachons ici $N_c = 2$ bits. Pour les modulations NW et CW, nous fixons $N_v = 256$ et pour le χ^2W, nous utilisons $N_v = 55$ afin d'obtenir la même distorsion pour les trois schémas, à savoir $\mathrm{WCR} = -18\ dB$ en moyenne (le CW est

message	probabilité	intervalle de $\mathbb{R}^+$
(0 0)	0.1	[0 ;42.06]
(0 1)	0.2	[42.06 ;49.055]
(1 0)	0.15	[49.055 ;53.037]
(1 1)	0.15	[53.037 ;57.016]
(0 0)	0.15	[57.016 ;61.665]
(0 1)	0.05	[61.665 ;63.577]
(1 0)	0.1	[63.577 ;68.796]
(1 1)	0.1	[68.796 ;+∞[

TABLE 2.1 – χ^2W : exemple de clé secrète générée pour χ^2W avec $N_c = 2$, $N_w = 8$, $N_v = 55$. Ce tableau montre les intervalles réels (mots de code) en fonction de sa probabilité. Chaque message apparaît avec la même probabilité. Cette partition est la même que celle illustrée dans l'histogramme des normes figure 2.7.



(a) (b)

FIGURE 2.8 – Projections sur les deux premières composantes de signaux gaussiens hôtes (a) et tatoués χ^2W (b) de dimension $N_v = 55$. Paramètres : $N_c = 2$, $N_v = 55$, WCR $= -18\ dB$, $N_h = 8$, $N_o = 2000$. La distribution des signaux est identique avant et après tatouage, condition suffisante de la *stégo-sécurité*.

la seule technique parmi les trois testées ici où il est possible de fixer la distorsion). Comme nous pouvons le remarquer, le CW est plus robuste que le NW (conformément aux résultats théoriques). La faible robustesse du χ^2W s'explique par le fait que, pour ce schéma d'insertion, les régions de décodage sont très proches les unes des autres (voir figure 2.4). Par conséquent, un signal tatoué corrompu par l'ajout de bruit aura une plus grande probabilité de basculer dans une autre région de mot de code pour le χ^2W que pour les méthodes NW et CW.

Nous avons vu précédemment que le niveau de sécurité des méthodes de tatouage dans le contexte WOA est entièrement défini par la distribution des signaux après insertion : distribution des corrélations entre les signaux tatoués et les porteuses secrètes pour les méthodes par étalement de spectre (BPSK), distribution des normes

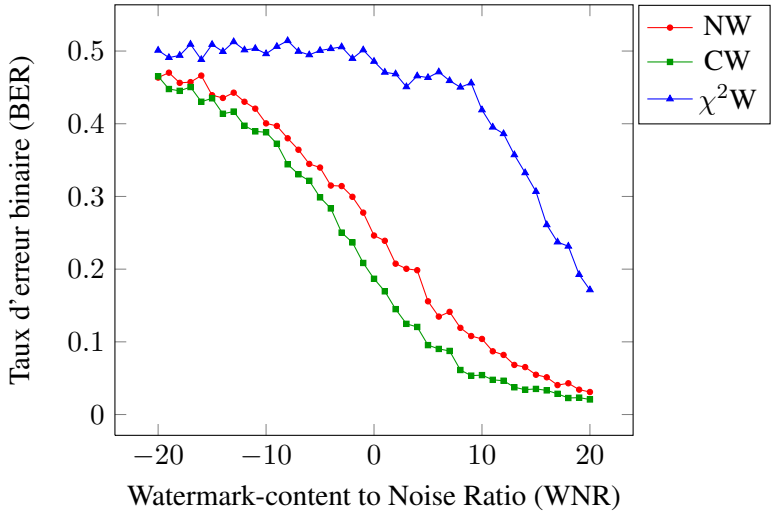


FIGURE 2.9 – BER en fonction du WCNR pour NW, CW et χ^2W. Pour les trois techniques, nous obtenons : WCR $\cong -18\ dB$. La faible robustesse du χ^2W s'explique par le fait que, pour ce schéma d'insertion, les régions de décodage sont très proches les unes des autres (voir figure 2.4). Par conséquent, un signal tatoué corrompu par l'ajout de bruit aura une plus grande probabilité de basculer dans une autre région de mot de code pour le χ^2W que pour les méthodes NW et CW.

euclidiennes au carré pour le tatouage par la loi du χ^2. Le tatouage numérique dans le cadre WOA est alors présenté sous un nouvel angle, nous générons d'abord la distribution des signaux tatoués dans le sous-espace privé induit par la clé secrète, cette distribution étant en bijection avec le niveau de sécurité souhaité (par exemple par le distributeur de contenus). Les signaux tatoués sont alors générés par injection (ou rétro-projection) dans le domaine de dimension N_v. Cette nouvelle technique de tatouage s'appelle le **tatouage basé sur modèle**.

Cependant, du point de vue de la distorsion, les méthodes de tatouage présentées précédemment ne sont pas optimales : dans le contexte WOA, nous souhaitons tatouer N_o contenus hôtes. Chaque point de la distribution des contenus dans le sous-espace de tatouage (espace des corrélations de dimension N_c pour l'étalement de spectre, espace des normes euclidiennes au carré pour le χ^2W) est alors associé à un point correspondant au contenu tatoué dans une distribution induite par la sécurité désirée. La question qui se pose alors est : comment associer chaque point d'une distribution hôte avec chaque point d'une distribution tatouée dans une région de mot de code désirée tout en minimisant la moyenne des distances euclidiennes entre ces points (la distance euclidienne étant proportionnelle à la distorsion provoquée par le tatouage) ? La figure 2.10 illustre cette problématique.

Les formules de modulations (CW, NW) par étalement de spectre ou du χ^2W ne nous serviraient alors plus qu'à une chose : **générer un modèle**. L'affectation au modèle s'effectuerait ensuite en minimisant la distorsion.

Nous proposons dans ce chapitre deux moyens pour résoudre ce problème :

– la minimisation par l'algorithme des Hongrois,

– la minimisation par les éléments de la théorie du transport.

2.2 Minimisation de la distorsion par la méthode dite des Hongrois

2.2.1 Graphe biparti et couplage parfait minimal

Avant de présenter la méthode dite des Hongrois, quelques rappels en théorie des graphes s'avèrent indispensables. Nous commençons par poser trois définitions :

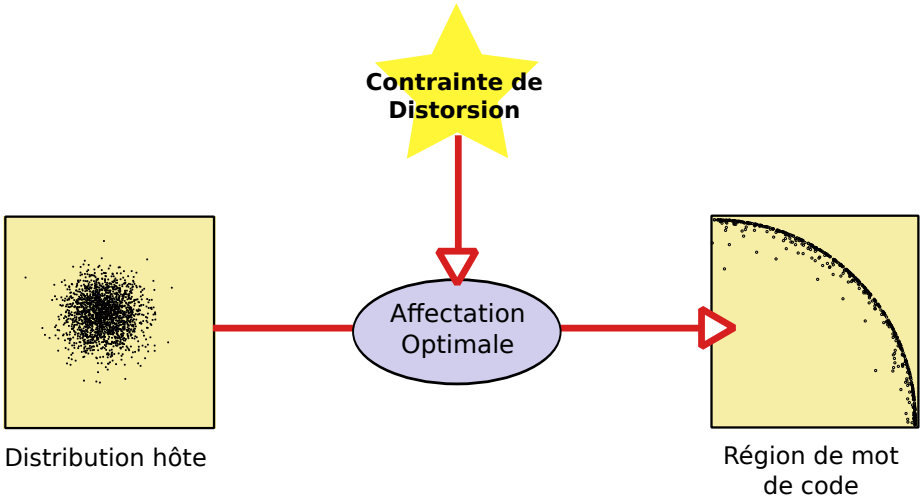


FIGURE 2.10 – Affectation d'une distribution de signaux hôtes (ici gaussienne en 2 dimensions) à une distribution tatouée dans une région de mot de code désirée par le tatoueur (ici le message (00) dans le cas du CW avec $N_c = 2$) soumise à la contrainte de la minimisation de la distorsion.

Définition 2 *:*

- *Un graphe biparti est un graphe $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ avec $\mathcal{V}$ l'ensemble des sommets et $\mathcal{E}$ l'ensemble des arrêtes respectant la propriété suivante : $\mathcal{V}$ est représenté comme une union de deux ensembles disjoints ($\mathcal{V} = \mathcal{A} \sqcup \mathcal{B}$), chaque arête de $\mathcal{E}$ est de la forme $[a, b]$ avec $a \in \mathcal{A}$ et $b \in \mathcal{B}$. De plus $\mathcal{G}$ satisfait $\#\mathcal{A} = \#\mathcal{B} = N_m$;*
- *Un graphe biparti pondéré $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathrm{P})$ est un graphe biparti dans lequel chaque arête de $\mathcal{E}$ est pondérée par une fonction $\mathrm{P} : \mathcal{E} \to \mathbb{R}^+$;*
- *Un couplage parfait (ou couplage complet) $\mathcal{M}$ est défini comme un sous-ensemble de $\mathcal{E}$ de N_m arrêtes où chaque sommet de $\mathcal{V}$ est incident à exactement une arête de $\mathcal{M}$.*

Dans cette section, nous nous intéressons au *Problème d'Affectation (PA)*, nous cherchons le couplage parfait minimal $\mathcal{M}^*$, c'est-à-dire le couplage parfait dont la somme des poids des arêtes est minimal. Formellement, nous cherchons :

$$\mathcal{M}^* = \arg\min_{\mathcal{M}} \sum_{t \in \mathcal{M}} \mathrm{P}(t). \tag{2.5}$$

La figure 2.11 montre un exemple de graphe biparti pondéré de taille $N_m = 3$ et le couplage parfait minimal qui lui est associé.

2.2.2 L'algorithme des Hongrois

La méthode des Hongrois (aussi appelé algorithme des Hongrois) [50] est un algorithme efficace pour résoudre le PA dans un graphe biparti pondéré en temps polynomial $O({N_m}^3)$. Nous considérons $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathrm{P})$ un graphe biparti pondéré avec :

- $\mathcal{V} = \mathcal{A} \sqcup \mathcal{B}$,
- $\mathcal{A} = \{a_0, \ldots, a_{N_m - 1}\}$ (sommets de départ),
- $\mathcal{B} = \{b_0, \ldots, b_{N_m - 1}\}$ (sommets d'arrivée).

Nous considérons alors la matrice $\mathbf{D} \in \mathcal{M}_{N_m, N_m}(\mathbb{R})$ de pondération telle que $\mathbf{D}(i, j) = \mathrm{P}([a_i, b_j]),\ i, j \in [N_m]$. Il s'agit de trouver N_m éléments de cette matrice de façon à ce que chaque ligne et chaque colonne

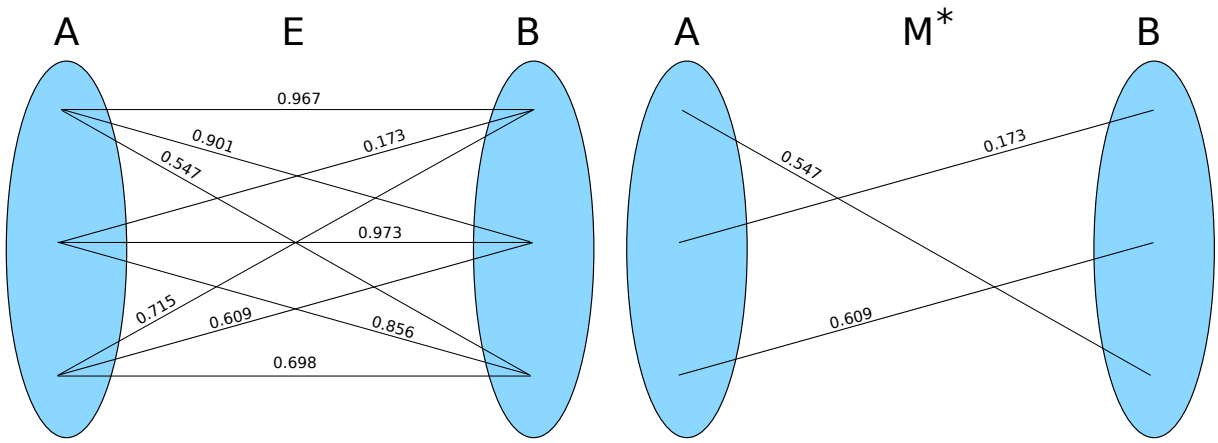


FIGURE 2.11 – Exemple d'un graphe biparti pondéré avec 2 partitions $\mathcal{A}$ et $\mathcal{B}$ de $N_m = 3$ sommets ainsi que son couplage parfait minimal $\mathcal{M}^*$ (les poids entre deux sommets sont notés sur chaque arête correspondante).

contienne un élément choisi, et que la somme des éléments choisis soit minimale. La méthode des Hongrois est présentée dans l'algorithme 1.

Algorithme 1 Algorithme (ou méthode) des Hongrois.

ENTRÉE : $\mathbf{D} \in \mathcal{M}_{N_m,N_m}$.

SORTIE : $\mathcal{M}^*$, l'ensemble des arêtes $[a_i, b_j]$ telles que $\mathbf{D}(i,j) = 0$.

Soustraire ligne par ligne et colonne par colonne le plus petit élément.

Tant Que On ne peut affecter un 0 par ligne et par colonne. **Faire**

Choisir au hasard une affectation de 0. Ces 0 choisis sont dits 0 encadrés, les autres sont dits 0 barrés.

Répéter

Marquer toute ligne n'ayant pas de 0 encadré.

Marquer toute colonne ayant un 0 barré dans une ligne marquée.

Marquer toute ligne ayant un 0 encadré dans une colonne marquée.

Jusqu'à Marquage impossible

Tracer un trait sur les lignes non marquées et les colonnes marquées.

Prendre le plus petit chiffre du tableau restant et le retrancher de tous les éléments non rayés et l'ajouter aux éléments rayés deux fois.

Fin Tant Que

Le couplage parfait minimal recherché est alors :

$$\mathcal{M}^* = \{[a_i, b_j] \ : \ \mathbf{D}(i,j) = 0\}. \tag{2.6}$$

2.2.3 Application aux méthodes sûres par étalement de spectre

Cette section décrit le fonctionnement d'un tatouage par étalement de spectre basé sur modèle par la méthode des Hongrois. Nous utilisons ici les distributions sûres données par les méthodes par étalement de spectre CW et NW.

2.2.3.1 Construction des graphes bipartis

Nous voulons générer 2^{N_c} (nombre de messages possibles) graphes bipartis $\mathcal{G} = (\mathcal{X} \sqcup \mathcal{Y}, \mathcal{E})$ avec $\mathcal{X}$ les corrélations des signaux hôtes avec les porteuses secrètes pour l'ensemble de départ, $\mathcal{Y}$ les corrélations de signaux tatoués avec la méthode par étalement de spectre choisie et un message $\mathbf{m}$. Le but est alors de trouver le couplage parfait minimal entre les deux partitions $\mathcal{X}$ et $\mathcal{Y}$.

Formellement, nous utilisons N_m signaux hôtes $\{\mathbf{x}_i\}_{i\in[N_m]}$. Pour chaque signal $\mathbf{x}_i$, nous construisons le vecteur $\mathbf{z}_{\mathbf{x}_i}$ des corrélations avec les porteuses secrètes :

$$\mathbf{z}_{\mathbf{x}_i} = \left(z_{\mathbf{x}_i,\mathbf{u}_0}, \ldots, z_{\mathbf{x}_i,\mathbf{u}_{N_c-1}}\right). \tag{2.7}$$

Nous voulons alors construire 2^{N_c} distributions d'arrivée correspondant aux régions de décodage dans le sous-espace privé de dimension N_c. Comme nous l'avons vu précédemment, les modulations CW et NW permettent de construire les distributions pouvant atteindre une classe de sécurité désirée (stégo-sécurité, clé-sécurité, etc.). Nous tatouons alors, pour chaque message $\mathbf{m}$, les signaux hôtes $\mathbf{x}_i$ avec la modulation choisie afin d'obtenir les signaux tatoués $\mathbf{y}_i,\ i \in [N_m]$. Nous construisons le vecteur $\mathbf{z}_{\mathbf{y}_i}$ des corrélations avec nos porteuses :

$$\mathbf{z}_{\mathbf{y}_i} = \left(z_{\mathbf{y}_i,\mathbf{u}_0}, \ldots, z_{\mathbf{y}_i,\mathbf{u}_{N_c-1}}\right). \tag{2.8}$$

Nous obtenons alors, pour chaque message, le graphe biparti pondéré $\mathcal{G} = (\mathcal{X} \sqcup \mathcal{Y}, \mathcal{E}, \mathrm{P})$ avec :

- $\mathcal{X} = \{\mathbf{z}_{\mathbf{x}_i}\}_{i\in[N_m]}$,
- $\mathcal{Y} = \{\mathbf{z}_{\mathbf{y}_j}\}_{j\in[N_m]}$ (dépend du message $\mathbf{m}$),
- $\mathcal{E}$ l'ensemble des arêtes $[\mathbf{z}_{\mathbf{x}_i}, \mathbf{z}_{\mathbf{y}_j}]$ du graphe $\mathcal{G}$,
- P la fonction de poids des arêtes de $\mathcal{G}$, nous utilisons la norme euclidienne :

$$\mathrm{P}([\mathbf{z}_{\mathbf{x}_i}, \mathbf{z}_{\mathbf{y}_j}]) = \|\mathbf{z}_{\mathbf{x}_i} - \mathbf{z}_{\mathbf{y}_j}\|.$$

Après application de la méthode des Hongrois, nous nous retrouvons avec 2^{N_c} couplages parfaits minimaux $\mathcal{M}_k^*,\ k \in [2^{N_c}]$ entre des corrélations hôtes et corrélations marquées en utilisant l'algorithme des Hongrois.

2.2.3.2 Réduction de la complexité des affectations

Précédemment, nous construisions un graphe biparti pondéré pour chaque message à insérer. Afin de réduire la complexité des calculs de bijections (affectations), nous pouvons utiliser la propriété de symétrie de nos distributions propre à l'étalement de spectre (les axes de symétrie étant les porteuses secrètes). Les points de la distribution d'arrivée sont alors calculés afin d'insérer un message constant. Pour la suite de ce manuscrit, nous appellerons N_m-carte un triplet $(\mathcal{X}, \mathcal{Y}, \mathcal{M}^*)$ construit avec N_m signaux hôtes et tatoués avec le message constant $(1, 1, \ldots, 1)$.

La figure 2.12 montre un graphe biparti construit avec la modulation CW ($N_c = 2$) bits et la figure 2.13 montre la 3-carte avec le couplage parfait minimal trouvé par la méthode des Hongrois.

La figure 2.14 présente les affectations optimales trouvées par la méthode des Hongrois d'une distribution hôte gaussienne ($N_c = 2$) vers une distribution CW pour le message $(1, 1, \ldots, 1)$. Pour plus de clarté, la distribution hôte est découpée en neuf zones (seuils choisis en fonction de la norme euclidienne des corrélations). Chaque zone hôte est affectée à une zone tatouée de la même couleur dans la distribution CW grâce à la méthode des Hongrois. Nous remarquons que plus les corrélations hôtes s'éloignent de l'origine, plus elles sont affectées au niveau des frontières de décodage.

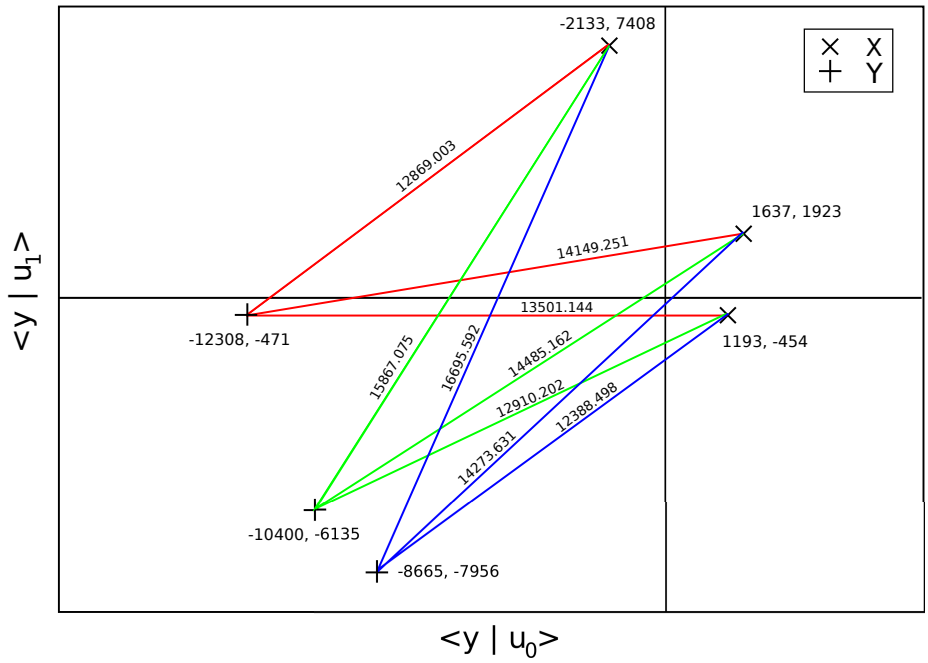


FIGURE 2.12 – Projections de signaux tatoués sur $N_c = 2$ porteuses secrètes : construction d'un graphe biparti pondéré pour la méthode CW avec 2 partitions de $N_m = 3$ sommets, corrélations hôtes et corrélations marquées. Les distances euclidiennes (poids) sont notées sur chaque arête.

2.2.3.3 Insertion basée sur modèle

Nous considérons $(\mathcal{X}, \mathcal{Y}, \mathcal{M}^*)$ la N_m-carte construite dans la section précédente. Nous voulons alors tatouer un signal $\mathbf{x}$ avec un message arbitraire $\mathbf{m}$ en utilisant notre N_m-carte. Nous commençons tout d'abord par créer le vecteur de corrélations $\mathbf{z_x}$ d'après l'équation (2.7). Nous voulons affecter $\mathbf{z_x}$ avec un point de $\mathcal{Y}$ en utilisant $\mathcal{M}^*$.

Nous avons vu précédemment que les éléments de $\mathcal{Y}$ de la N_m-carte sont construits de façon à cacher le message $(1, 1, \ldots, 1)$. Nous relevons alors les indices du message $\mathbf{m}$ dont les bits diffèrent du symbole "1". Notons qu'une affectation d'une corrélation hôte à une région de mot de code donnée permet une affectation à n'importe quelle autre région de décodage par symétrie axiale selon les porteuses secrètes. Par conséquent, des changements de signe doivent intervenir sur les coefficients de $\mathbf{z_x}$ qui ont subi des symétries.

Rappelons qu'une telle méthode d'affectation n'est possible que si la répartition des mots de codes peut se construire par un ensemble de symétries/rotations (comme ici pour les méthodes par étalement de spectre).

Après affectation, les symétries inverses doivent être effectuées pour pouvoir cacher le message correct $\mathbf{m}$. Formellement, nous construisons le vecteur $\mathbf{R} \times \mathbf{z_x}$ avec $\mathbf{R} \in \mathcal{M}_{N_c,N_c}(\mathbb{Z})$:

$$\mathbf{R}(i,j) = \begin{cases} 0 \text{ if } i \neq j, \\ (-1)^{\mathbf{m}(i)+1} \text{ if } i = j, \end{cases} \tag{2.9}$$

afin d'effectuer des changements de signe au niveau des indices des corrélations représentant les bits différents de "1". Ensuite, nous prenons le plus proche voisin (distance euclidienne minimale) de $\mathbf{R} \times \mathbf{z_x}$ dans $\mathcal{X}$, par exemple $\mathbf{z}_{\mathbf{x}_{i_0}}$. Grâce au couplage parfait minimal $\mathcal{M}^*$ nous trouvons $\mathbf{z}_{\mathbf{y}_{j_0}}$, l'affectation de $\mathbf{z}_{\mathbf{x}_{i_0}}$. Nous effectuons alors des

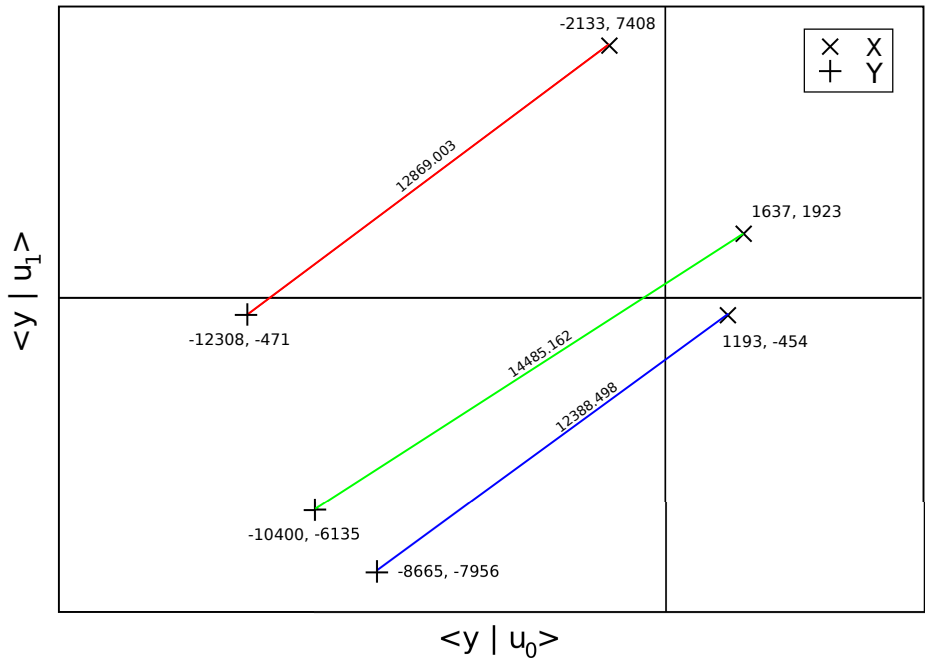


FIGURE 2.13 – Projections de signaux tatoués sur $N_c = 2$ porteuses secrètes : couplage parfait minimal du graphe biparti pondéré construit figure 2.12 pour la méthode CW avec 2 partitions de $N_m = 3$ sommets, corrélations hôtes et corrélations marquées. Les distances euclidiennes (poids) sont notées sur chaque arête. Le couplage parfait minimal associe les éléments de chaque partition (ensembles de départ $\mathcal{X}$ et d'arrivée $\mathcal{Y}$) en minimisant la somme des distances euclidiennes (proportionnelles à la distorsion provoquée par le tatouage).

symétries inverses afin de calculer le vecteur marqué $\mathbf{z_y}$:

$$\mathbf{z_y} = \mathbf{R}^{-1}\mathbf{z}_{\mathbf{y}_{j_0}} = \mathbf{R}\mathbf{z}_{\mathbf{y}_{j_0}}, \tag{2.10}$$

le vecteur de corrélation de notre signal marqué par la méthode de tatouage basé sur modèle par notre N_m-carte. Pour construire notre signal tatoué $\mathbf{y} \in \mathbb{R}^{N_v}$, nous effectuons une différence entre $\mathbf{z_y}$ et $\mathbf{z_x}$ pour obtenir $\mathbf{z_w}$, le vecteur de corrélation du signal de tatouage. La rétro-projection dans l'espace de dimension N_v est assurée par :

$$\mathbf{w} = \sum_{i=0}^{N_c-1} \mathbf{z_w}(i)\mathbf{u}_i. \tag{2.11}$$

Finalement, nous obtenons le signal tatoué $\mathbf{y} = \mathbf{x} + \mathbf{w}$. La figure 2.15 illustre la procédure de tatouage en utilisant la 3-carte construite figure 2.13.

Pour la suite de ce manuscrit, nous appellerons **tatouage circulaire Hongrois HCW** (*Hungarian Circular Watermarking*), **tatouage naturel Hongrois HNW** (*Hungarian Natural Watermarking*) et **tatouage naturel Hongrois robuste HRNW** (*Hungarian Robust Natural Watermarking*) les schémas par étalement de spectre basés sur les modèles des schémas classiques sûrs par étalement de spectre.

La figure 2.16 montre les distributions des corrélations entre les signaux hôtes (gaussiens), NW et HNW et deux porteuses secrètes. Paramètres : $N_c = 2,\ N_v = 256,\ N_o = 2000$. Nous constatons que les corrélations NW et HNW sont identiques, notre tatouage basé sur modèle ne modifie pas la sécurité initiale du schéma. La distribution des corrélations CW et HCW sur deux porteuses est présentée figure 2.17 avec les mêmes paramètres que précédemment et WCR $= -18\ dB$. Nous remarquons que le HCW ne modifie pas la sécurité du schéma CW clé-sûr classique.

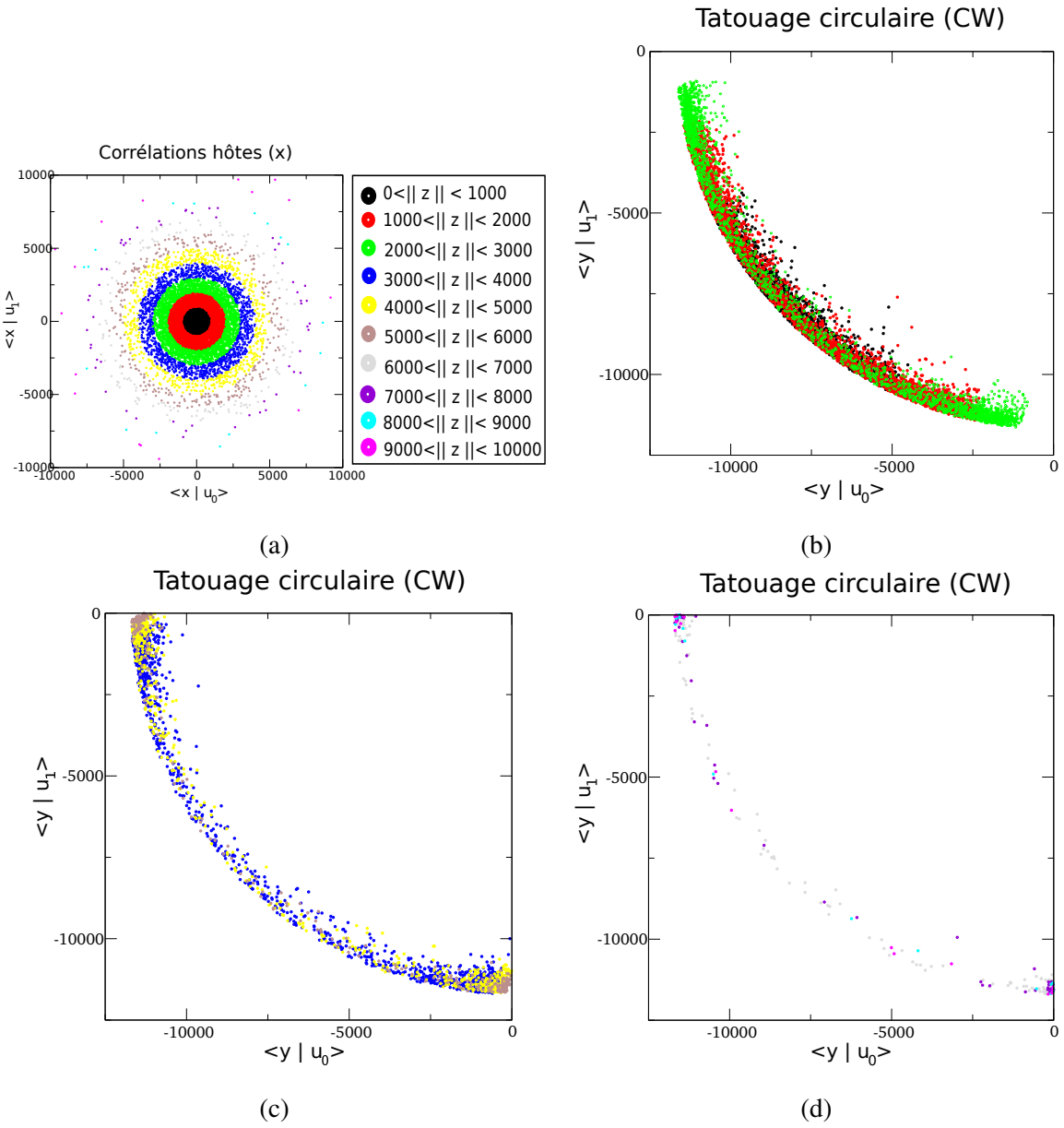


FIGURE 2.14 – Affectations optimales trouvées par l'algorithme des Hongrois d'une distribution hôte gaussienne (a) ($N_o = 10000$) vers une distribution CW (b) avec $N_o = 7554$, (c) avec $N_o = 2205$ et (d) avec $N_o = 241$ pour le message $(1, 1)$. La distribution hôte est découpée en 10 zones (seuils choisis en fonction de la norme euclidienne des corrélations). Chaque zone hôte est affectée à une zone tatouée de la même couleur dans la distribution CW grâce à la méthode des Hongrois. Nous remarquons que plus les corrélations hôtes s'éloignent de l'origine, plus elles sont affectées au niveau des frontières de décodage.

2.2.4 Application au tatouage par la loi du χ^2

La méthode de tatouage basé sur modèle par l'algorithme des Hongrois pour le χ^2W utilise le même principe que pour les techniques par étalement de spectre. Nous voulons créer 2^{N_c} graphes bipartis qui contiennent, pour l'ensemble de départ, les normes de N_m signaux hôtes gaussiens $\mathbf{x}_j$. Nous construisons les points de l'ensemble d'arrivée en sélectionnant uniquement les réels positifs dans l'intervalle de mot de code du message désiré. Nous

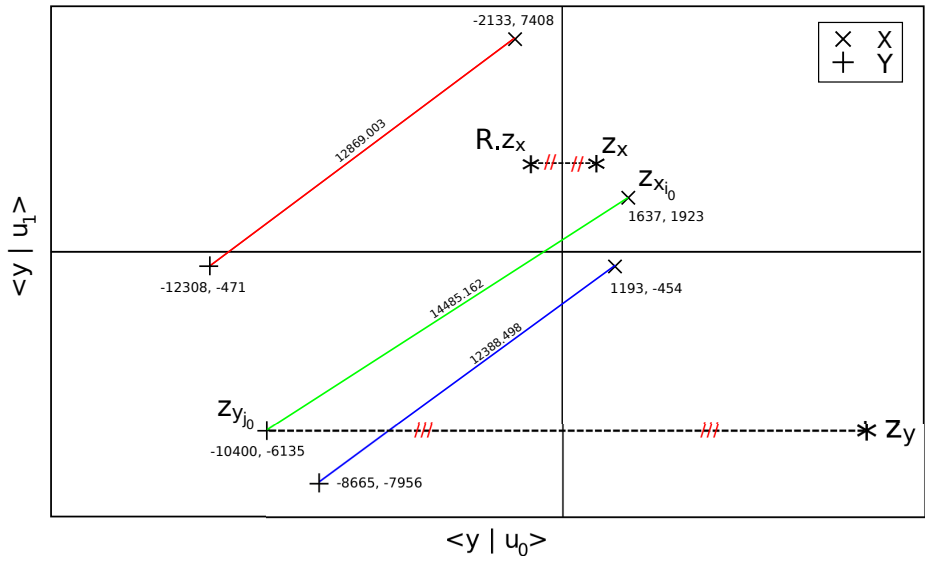


FIGURE 2.15 – Illustration du tatouage basé sur modèle. Insertion du message $\mathbf{m} = (0, 1)$, $N_c = 2$ bits. Après le calcul du vecteur de corrélations $\mathbf{z_x}$ d'un signal hôte, nous construisons le vecteur des corrélations tatouées $\mathbf{z_y}$ en utilisant la 3-carte construite figure 2.13.

générons alors, pour chaque message, N_m signaux gaussiens $\mathbf{y}_j$, avec $\|\mathbf{y}_j\|^2$ dans la bonne région de mot de code. Nous pouvons construire, pour tout $k \in [2^{N_c}]$, un graphe biparti pondéré $\mathcal{G} = (\mathcal{X} \sqcup \mathcal{Y}, \mathcal{E}, \mathrm{P})$ avec :

- $\mathcal{X} = \{\|\mathbf{x}_i\|^2\}_{i\in[N_m]}$,
- $\mathcal{Y} = \{\|\mathbf{y}_j\|^2\}_{j\in[N_m]}$,
- $\mathcal{E}$ l'ensemble des arêtes $[\|\mathbf{x}_i\|^2, \|\mathbf{y}_j\|^2]$ du graphe $\mathcal{G}$,
- P la fonction de poids des arêtes de $\mathcal{G}$, nous utilisons la valeur absolue de la différence entre chaque sommet de l'arête.

En utilisant la méthode des Hongrois, nous trouvons 2^{N_c} couplages parfaits minimaux $\mathcal{M}_k^*$ entre les normes hôtes et les normes tatouées (en fonction du message à cacher). Pour la suite de ce manuscrit, nous appellerons N_m-atlas l'ensemble $\{(\mathcal{X}, \mathcal{Y}_k, \mathcal{M}_k^*)\}_{k\in[2^{N_c}]}$ construit avec N_m signaux.

L'insertion est similaire à la méthode de tatouage basé sur modèle pour les modulations BPSK. Pour insérer un message $\mathbf{m}$ dans un signal $\mathbf{x}$ à l'aide d'un N_m-atlas, nous calculons tout d'abord $\|\mathbf{x}\|^2$. Nous cherchons ensuite le plus proche voisin de $\|\mathbf{x}\|^2$ dans $\mathcal{X}$. Grâce à $\mathcal{M}_k^*$ (k dépend de $\mathbf{m}$), nous trouvons $\|\mathbf{y}\|^2$. Finalement, nous obtenons $\mathbf{y} = \sqrt{\frac{\|\mathbf{y}\|^2}{\|\mathbf{x}\|^2}}\mathbf{x}$. Nous appelons cette technique de tatouage **tatouage hongrois par la loi du** χ^2 : **H**χ^2**W** (*Hungarian CHI*^{2} *Watermarking*).

La figure 2.18 montre la projection des signaux hôtes, χ^2W et Hχ^2W sur les deux premières composantes des vecteurs. Paramètres : $N_c = 2,\ N_v = 55,\ N_h = 8,\ N_o = 2000$. La distorsion est donnée par $\mathrm{WCR} = -18\ dB$ pour le χ^2W. Nous remarquons que la distribution sur ces deux composantes est la même pour les deux méthodes.

Nous comparons à présent les méthodes χ^2W et Hχ^2W avec les méthodes NW, HNW, CW et HCW en terme de distorsion et de robustesse. Les paramètres utilisés sont les mêmes que ceux choisis dans la section 2.1.2.2, à savoir $N_c = 2$ bits, $N_v = 256$ pour les modulations NW et CW, $N_v = 55$ pour le χ^2W afin d'obtenir la même distorsion pour les trois schémas, à savoir $\mathrm{WCR} = -18\ dB$ en moyenne. De nouveau, le CW est la seule technique parmi les trois testées ici où il est possible de fixer la distorsion. La table 2.2 montre l'impact de l'utilisation de méthodes basées sur modèle sur la distorsion. Celle-ci est calculée en moyenne sur $N_o = 2000$

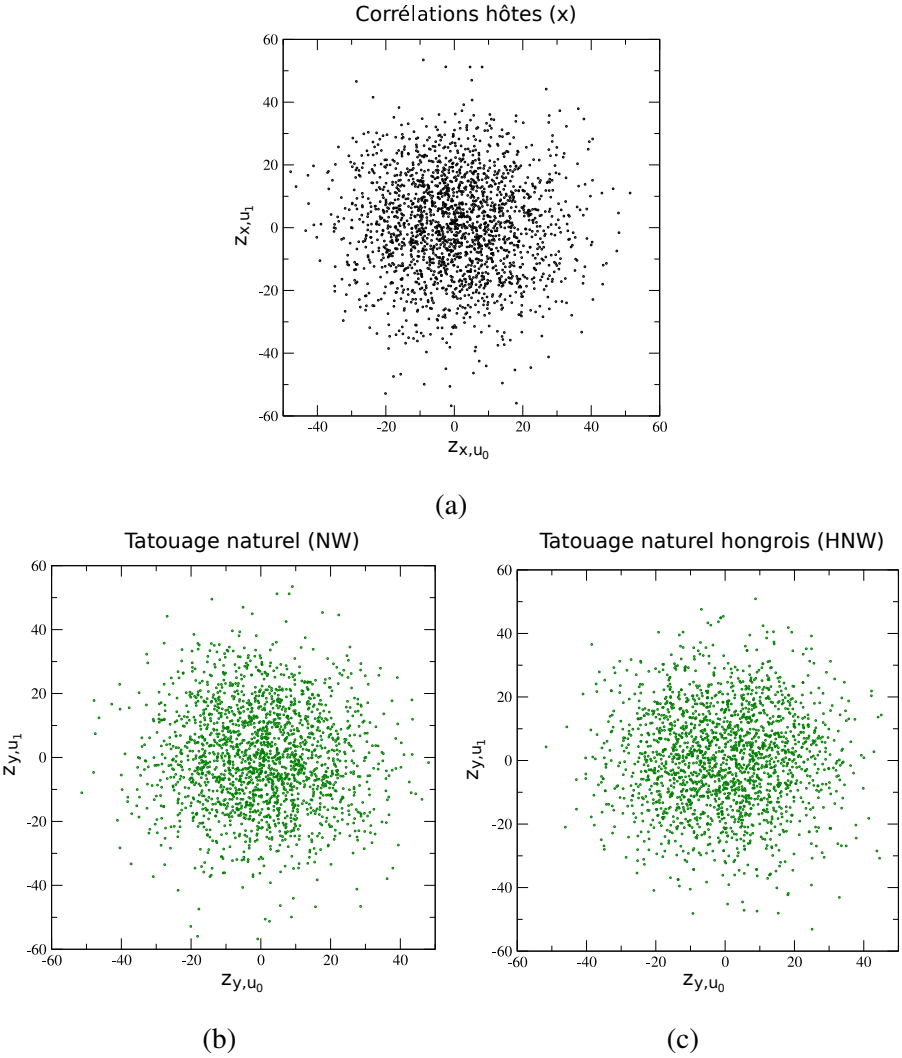
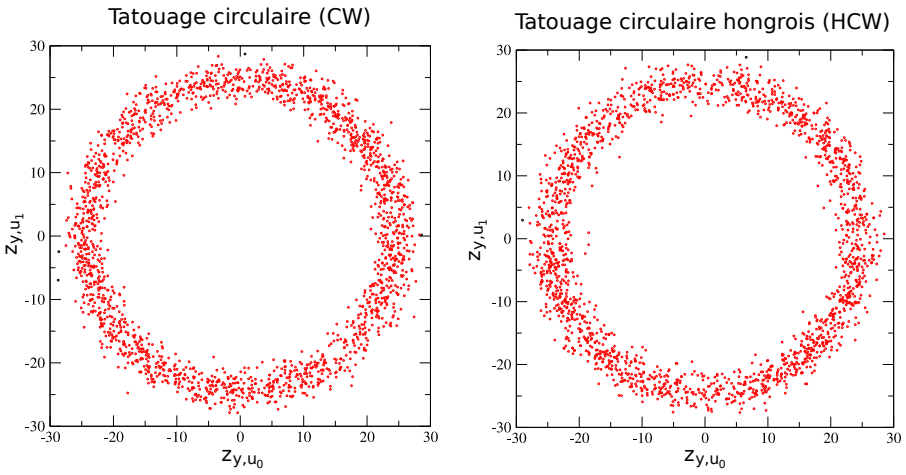


FIGURE 2.16 – (a) : distribution des projections des signaux hôtes sur deux porteuses. (b) and (c) : distributions des projections des signaux tatoués sur deux porteuses secrètes pour NW et HNW. Paramètres : $N_c = 2,\ N_v = 256,\ N_o = 2000$. Les corrélations NW et HNW sont identiques, notre tatouage basé sur modèle ne modifie pas la sécurité initiale du schéma.

signaux pour les six méthodes. Nous remarquons alors des gains de 2.7 dB pour NW, 1.1 dB pour CW et 3.6 dB pour χ^2W. Ce dernier résultat est dû au fait que, pour ce schéma d'insertion, il y a plus d'un mot de code pour un message et ces mots de code sont sur l'axe positif réel $\mathbb{R}^+$.

La figure 2.19 montre la robustesse des six schémas face à l'ajout de bruit gaussien (quantifié par l'utilisation d'un WCNR). Cette figure provient de celle présentée figure 2.9 à laquelle nous avons rajouté les optimisations par la méthode des Hongrois. Nous remarquons que l'utilisation de cette dernière méthode ne modifie pas la robustesse des schémas initiaux.



(a) (b)

FIGURE 2.17 – Distribution des projections des signaux tatoués sur deux porteuses pour CW (a) et HCW (b). Paramètres : $N_c = 2,\ N_v = 256,\ N_o = 2000,\ \mathrm{WCR} = -18\ dB,\ \mathrm{NCR} = -10\ dB$. Nous remarquons que le HCW ne modifie pas la sécurité du schéma CW clé-sûr classique.

Technique d'insertion	WCR (méthode classique)	WCR (méthode des Hongrois)
NW	-18.07	-20.76
CW	-17.97	-19.11
χ^2W	-18.02	-21.65

TABLE 2.2 – Distorsion en moyenne sur $N_o = 2000$ signaux pour NW, CW et χ^2W ainsi que leurs optimisations par la méthode des Hongrois. Paramètres : $N_c = 2$ bits, $N_v = 256$ pour les modulations NW et CW et $N_v = 55$ pour le χ^2W.

2.3 La théorie du transport appliquée au tatouage sûr

La méthode du tatouage basé sur modèle par la méthode des Hongrois présentée dans la section précédente permet de minimiser la distorsion globale causée par le tatouage dans le cadre WOA. Cependant, cette dernière méthode présente les deux inconvénients suivants :

1. Les N_m-cartes et N_m-atlas doivent être générés et stockés "en dur" avant le processus de tatouage, cette condition est très restrictive car ces données dépendent du nombre de bits N_c que le distributeur de contenus veut cacher ;
2. La complexité cubique de l'algorithme des Hongrois peut devenir un réel problème quand N_m augmente.

Nous proposons dans cette section une nouvelle méthode de tatouage basé sur modèle par étalement de spectre permettant d'affecter une distribution de corrélations hôtes (corrélations avec les porteuses secrètes) à une distribution de corrélations tatouées par un moyen optimal (du point de vue de la distorsion) basé sur la **théorie du transport**. Contrairement à la méthode basée sur modèle par l'algorithme des Hongrois qui est une méthode discrète, nous considérons les mesures de probabilité continues des corrélations hôtes et tatouées pour une affectation optimale des distributions.

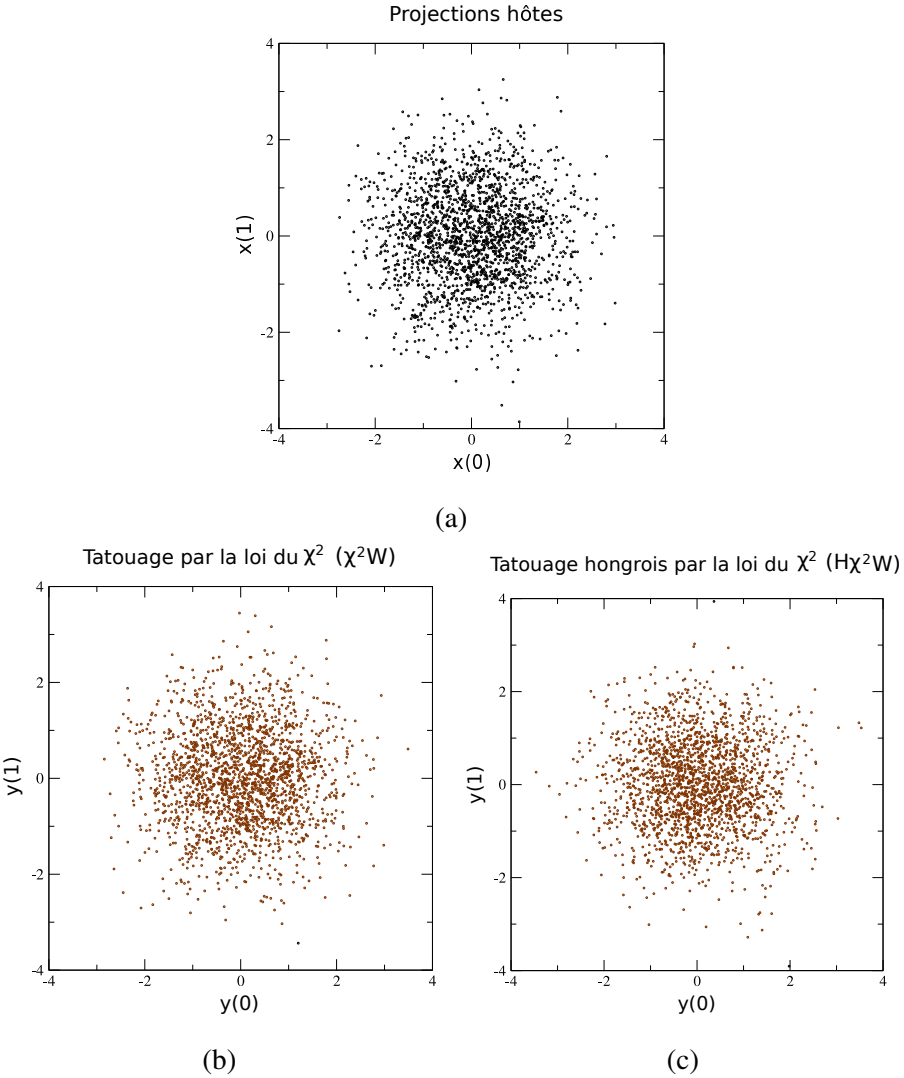


FIGURE 2.18 – (a) : projection des signaux hôtes sur les deux premières composantes des vecteurs. (b) and (c) : projection des signaux tatoués sur les deux premières composantes des vecteurs pour χ^2W et Hχ^2W.

2.3.1 Le problème du transport par Monge et Kantorovitch

Le problème du transport peut être vu comme un problème d'ingénierie, défini par Monge [64] en 1781 : nous considérons un tas de sable $\mathcal{X} \subset \mathbb{R}^{N_c}$ de distribution μ que l'on veut déplacer dans des creux $\mathcal{Y} \subset \mathbb{R}^{N_c}$ de distribution ν (μ et ν sont des mesures de probabilité).

Une fonction de coût c rentre alors en compte :

$$\begin{array}{rccc} \mathrm{c}: & \mathbb{R}^{N_c} \times \mathbb{R}^{N_c} & \rightarrow & [0, +\infty[\\ & (\mathbf{z_x}, \mathbf{z_y}) & \mapsto & \mathrm{c}(\mathbf{z_x}, \mathbf{z_y}), \end{array} \tag{2.12}$$

où $\mathrm{c}(\mathbf{z_x}, \mathbf{z_y}) = \mathrm{h}(\mathbf{z_y} - \mathbf{z_x})$ représente le coût nécessaire pour déplacer une unité de masse de sable $\mathbf{z_x}$ à $\mathbf{z_y}$. L'objectif est ici de trouver une bijection $\mathrm{T} : \mathcal{X} \rightarrow \mathcal{Y}$ avec $\nu = \mathrm{T}_\sharp \mu = \mu \circ \mathrm{T}^{-1}$ (ν est la mesure image de μ par T) qui minimise le coût total de déplacement du sable vers les creux.

La figure 2.20 illustre le problème du transport.

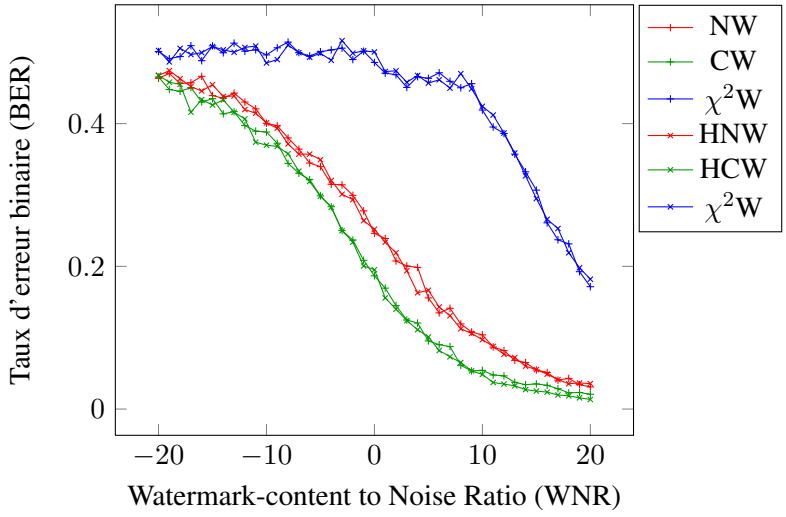


FIGURE 2.19 – BER en fonction du WCNR pour NW, CW et χ^2W ainsi que pour leur optimisations par la méthode des Hongrois HNW, HCW et $H\chi^2$W. Pour les trois premières techniques, nous obtenons : WCR $\cong -18\ dB$.

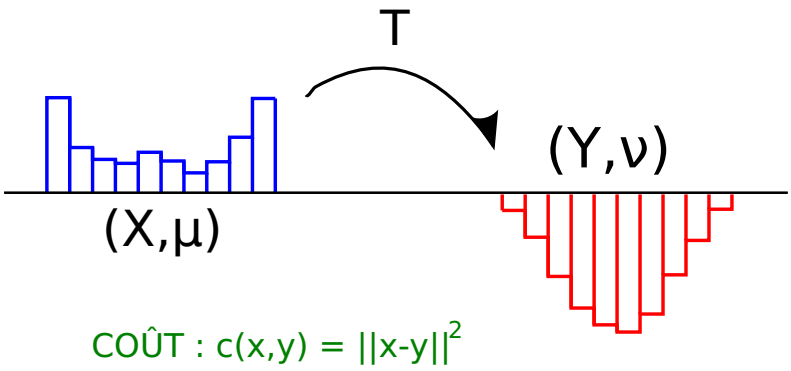


FIGURE 2.20 – Problème du transport : on cherche une bijection T qui minimise le coût total c de déplacement du sable $\mathcal{X}$ de distribution μ vers les creux $\mathcal{Y}$ de distribution ν.

Formellement, nous recherchons un plan de transport T qui minimise :

$$\inf_{\mathrm{T}} \left\{ \int_{\mathbb{R}^{N_c}} \mathrm{c}\left(\mathbf{z_x}, \mathrm{T}\left(\mathbf{z_x}\right)\right) \mu(\mathbf{z_x}) \mathrm{d}\mathbf{z_x} \mid \nu = \mathrm{T}_\sharp \mu \right\}. \tag{2.13}$$

Une solution T du problème est appelé plan de transport optimal. Cette recherche est nommée comme étant le problème Monge-Kantorovitch car il a tout d'abord été formalisé par Monge et les principales contributions ont été apportées par Kantorovitch [45, 46].

Pour $N_c = 1$, un plan de transport optimal T pour une fonction de coût convexe h est donné par [72, 81] :

$$\mathrm{T} = \mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu, \tag{2.14}$$

où P_δ désigne la fonction de répartition d'une mesure de probabilité δ.

La figure 2.21 illustre le plan de transport optimal d'une distribution $\mu \sim \mathcal{N}(0,1)$ vers une distribution $\nu \sim \mathcal{N}(0.5, 0.5)$. L'exemple pris ici est le déplacement du scalaire $x = -1.5$ vers $y = \mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu(x) = -0.25$.

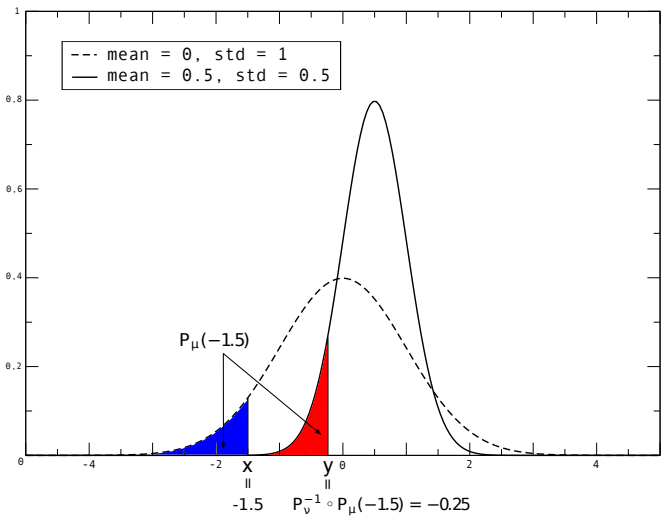


FIGURE 2.21 – Plan de transport optimal d'une distribution $\mu \sim \mathcal{N}(0,1)$ vers une distribution $\nu \sim \mathcal{N}(0.5, 0.5)$. L'exemple pris ici est le déplacement du scalaire $x = -1.5$ vers $y = \mathrm{P}_{\nu}^{-1} \circ \mathrm{P}_{\mu}(x) = -0.25$. Les aires en rouge et bleu sont égales.

Plus généralement, pour tout N_c, nous avons le théorème 1 [48] :

Théorème 1 ***Critère d'optimalité de Knott-Smith*** *: si μ (respectivement ν) représente la distribution de $\mathcal{X}$ (resp. $\mathcal{Y}$), les conditions suffisantes pour que le plan de transport T minimise le problème Monge-Kantorovitch avec $\mathrm{c}(\mathbf{z_x}, \mathbf{z_y}) = \|\mathbf{z_x} - \mathbf{z_y}\|^2$ sont :*

i) $\mathrm{T}(\mathcal{X})$ *a pour distribution ν,*

ii) la matrice jacobienne $\mathbf{J}_{\mathrm{T}}(\mathbf{z_x})$ de T est symétrique et semi-définie positive.

Nous avons vu que la sécurité des schémas par étalement de spectre s'appuie sur la distribution des corrélations z entre les signaux tatoués et les porteuses secrètes. De plus, nous connaissons l'expression analytique de cette distribution pour le tatouage naturel (NW) et le tatouage naturel robuste (RNW), nous sommes alors capables d'utiliser le théorème 1 afin d'affecter les corrélations hôtes et tatouées de façon à minimiser la distorsion.

2.3.2 Application au tatouage naturel NW et au tatouage robuste RNW

Une des propriétés du NW et du RNW est la préservation de la distribution des corrélations (mis à l'échelle par un facteur η pour le RNW) par rapport aux porteuses secrètes avant et après tatouage des signaux supposés gaussiens. De [7], nous avons :

$$\mathrm{z}_{\mathbf{x},\mathbf{u}_i} \sim \mathcal{N}\left(0, \frac{\sigma_{\mathbf{x}}^2}{N_v}\right) = \mu. \tag{2.15}$$

Nous considérons tout d'abord l'insertion du message de N_c bits $(0, 0, \ldots, 0)$ pour chaque signal hôte. Nous obtenons :

$$\mathrm{z}_{\mathbf{y},\mathbf{u}_i} \sim \mathcal{N}^{+}\left(0, \frac{\eta^2\sigma_{\mathbf{x}}^2}{N_v}\right) = \nu, \tag{2.16}$$

où $\mathcal{N}^+$ désigne une distribution gaussienne tronquée dans la région $\mathbb{R}^+$. Si $\delta = \mathcal{N}\left(0, \frac{\eta^2\sigma_{\mathbf{x}}^2}{N_v}\right)$, la distribution $\nu = \mathcal{N}^+\left(0, \frac{\eta^2\sigma_{\mathbf{x}}^2}{N_v}\right)$ est donnée par sa fonction de répartition :

$$\mathrm{P}_\nu(t) = \begin{cases} 0 \text{ si } t < 0, \\ 2\mathrm{P}_\delta(t) - 1 \text{ si } t \geq 0. \end{cases} \tag{2.17}$$

Une application des résultats théoriques précédents est alors pertinente, nous cherchons le plan de transport optimal T_0 permettant une affectation de la distribution des $\mathrm{z}_{\mathbf{x},\mathbf{u}_i}$ à la distribution des $\mathrm{z}_{\mathbf{y},\mathbf{u}_i}$. La stratégie adoptée ici est d'utiliser un plan de transport optimal pour chaque dimension du sous-espace privé de dimension N_c grâce à l'équation (2.14). Le calcul du plan de transport optimal T_0 pour insérer le message $(0, \ldots, 0)$ (affectation dans une région de mot de code précise) est donné par :

$$\begin{array}{rccc} \mathrm{T}_0 : & \mathbb{R}^{N_c} & \rightarrow & \mathbb{R}^{N_c} \\ & \mathbf{z_x} & \mapsto & \mathbf{z_y} = \mathrm{T}_0(\mathbf{z_x}), \end{array} \tag{2.18}$$

où

$$\mathrm{T}_0 \begin{pmatrix} \mathbf{z_x}(0) \\ \vdots \\ \mathbf{z_x}(N_c - 1) \end{pmatrix} = \begin{pmatrix} \mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu(\mathbf{z_x}(0)) \\ \vdots \\ \mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu(\mathbf{z_x}(N_c - 1)) \end{pmatrix}. \tag{2.19}$$

Nous avons :

$$\mathrm{P}_\mu(t) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{t\sqrt{N_v}}{\sigma_{\mathbf{x}}\sqrt{2}}\right)\right), \tag{2.20}$$

et

$$\mathrm{P}_\nu^{-1}(t) = \mathrm{P}_\delta^{-1}\left(\frac{t}{2} + \frac{1}{2}\right) = \frac{\eta\sigma_{\mathbf{x}}\sqrt{2}}{\sqrt{N_v}}\mathrm{erf}^{-1}(t). \tag{2.21}$$

La figure 2.22 montre la courbe convexe $y = \mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu(x)$.

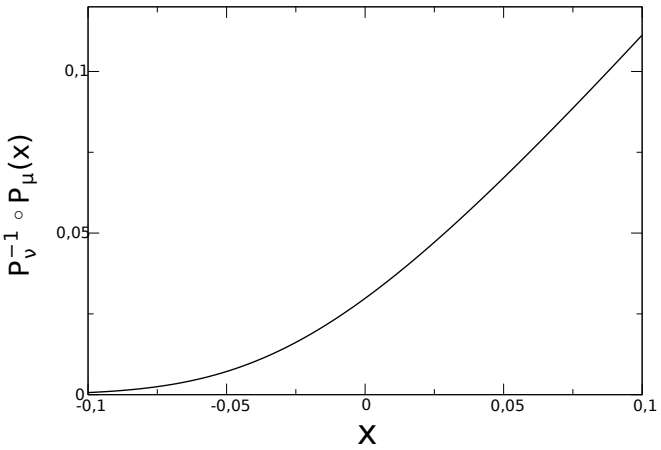


FIGURE 2.22 – Courbe $y = \mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu(x)$ avec $\eta = 1$, $N_v = 512$ et $\sigma_{\mathbf{x}}^2 = 1$.

Proposition 1 *Le plan de transport* T_0 *donné par l'équation (2.19) est optimal.*

Preuve Nous voulons montrer que :

$$\mathrm{T}_0 \begin{pmatrix} \mathbf{x}(0) \\ \vdots \\ \mathbf{x}(n-1) \end{pmatrix} = \begin{pmatrix} \mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu(\mathbf{x}(0)) \\ \vdots \\ \mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu(\mathbf{x}(n-1)) \end{pmatrix}, \tag{2.22}$$

est un plan de transport optimal en appliquant le théorème 1. La condition *i)* est vérifiée car un signal $\mathbf{x}$ est gaussien si et seulement si les composantes $\mathbf{x}(i)$ sont gaussiennes et indépendantes (on utilise la séparabilité de la distribution gaussienne en plusieurs dimensions). Pour la condition *ii)*, nous introduisons la matrice jacobienne de T_0 :

$$\mathbf{J}_{\mathrm{T}_0}\left(\mathbf{x}(0)\dots\mathbf{x}(n-1)\right)$$

$$= \begin{pmatrix} \frac{\partial\left(\mathrm{P}_\nu^{-1}\circ\mathrm{P}_\mu(\mathbf{x}(0))\right)}{\partial\mathbf{x}(0)} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\partial\left(\mathrm{P}_\nu^{-1}\circ\mathrm{P}_\mu(\mathbf{x}(n-1))\right)}{\partial\mathbf{x}(n-1)} \end{pmatrix}. \tag{2.23}$$

Montrons que cette matrice est symétrique et semi-définie positive. La symétrie est triviale par construction du plan de transport T_0. Une matrice $\mathbf{J}$ est semi-definie positive si et seulement si les valeurs propres de $\mathbf{J}$ sont positives ou nulles. Nous devons alors montrer que $\left(\mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu\right)'(t) \geq 0$. Nous avons :

$$\left(\mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu\right)'(t) = \left(\mathrm{P}_\nu^{-1}\right)'\left(\mathrm{P}_\mu(t)\right)\mathrm{f}_\mu(t). \tag{2.24}$$

$\mathrm{f}_\mu(t)$ est positive (densité de probabilité) et $\mathrm{P}_\nu^{-1}(t)$ est une fonction croissante, nous avons alors $\left(\mathrm{P}_\nu^{-1}\right)'(t) \geq 0$. Nous avons $\left(\mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu\right)'(t) \geq 0$ et la matrice jacobienne de T_0 est semi-définie positive. Les conditions *i)* et *ii)* du théorème 1 sont vérifiées : T_0 est un plan de transport optimal.

C.Q.F.D.

Grâce à la propriété de symétrie des corrélations en étalement de spectre, pour insérer des messages qui diffèrent de $(0,\dots,0)$, des changements de signe doivent intervenir sur les coefficients de $\mathbf{z_x}$ au niveau des indices qui ont subi des symétries. Après insertion, des symétries inverses doivent être effectuées pour pouvoir insérer le message correct $\mathbf{m}$.

Formellement, le plan de transport optimal $\mathrm{T}_\mathbf{m}$ pour tout message $\mathbf{m}$ en fonction de η est donné par :

$$\mathrm{T}_{\mathbf{m},\eta}(\mathbf{z_x}) = \mathbf{R}^{-1}(\mathrm{T}_0(\mathbf{R}\mathbf{z_x})), \tag{2.25}$$

avec :

$$\mathbf{R} \in \mathcal{M}_{N_c,N_c}(\mathbb{R}), \mathbf{R}(i,j) = \begin{cases} 0 \text{ si } i \neq j, \\ (-1)^{\mathbf{m}(i)} \text{ sinon.} \end{cases} \tag{2.26}$$

Nous sommes alors en mesure de pouvoir définir deux nouvelles techniques de tatouage par étalement de spectre basées sur la théorie du transport, nommées **tatouage naturel transporté TNW** (*Transportation Natural Watermarking*) et **tatouage naturel transporté robuste TRNW** (*Transportation Robust Natural Watermarking*), les modulations correspondantes sont données par :

$$\mathrm{s}_{TRNW}(\mathbf{m}(i),\mathbf{x}) = \mathrm{T}_{\mathbf{m},\eta}(\mathbf{z_x})(i) - \mathbf{z_x}(i). \tag{2.27}$$

Nous remarquons que nous avons considérablement réduit la complexité du problème d'affectation optimale : au lieu d'utiliser un algorithme de complexité $\sim O(N_m^3)$ (section 2.2.3), l'affectation est linéaire et donnée directement par l'équation (2.25).

La figure 2.23 montre les modulations s_{NW} (équations (1.28) et (1.29)) et s_{TNW} (équation 2.27) en fonction de la corrélation sur la première porteuse secrète $\mathrm{z}_{\mathbf{x},\mathbf{u}_0}$ ($N_c = 1, N_v = 512, \eta = 2, \sigma_{\mathbf{x}}^2 = 1$). Contrairement à la modulation NW, la modulation TNW n'est pas linéaire par morceaux.

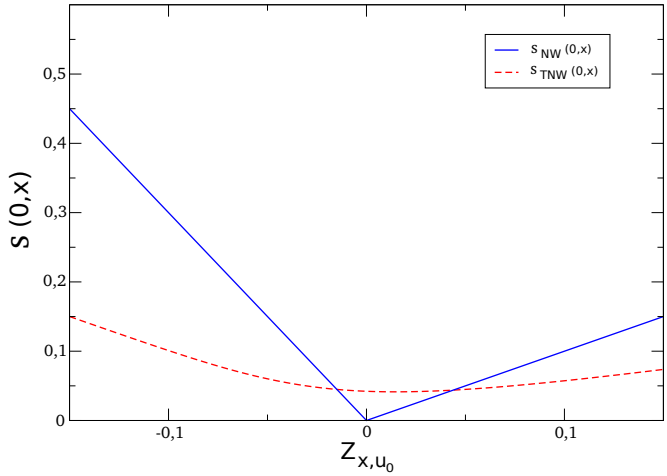


FIGURE 2.23 – Modulations s_{NW} et s_{TNW} en fonction de $\mathrm{z}_{\mathbf{x},\mathbf{u}_0}$ ($N_c = 1, N_v = 512, \eta = 2, \sigma_{\mathbf{x}}^2 = 1$). Contrairement à la modulation NW, la modulation TNW n'est pas linéaire par morceaux.

La figure 2.24 montre la distribution des corrélations entre les signaux hôtes, NW et TNW et deux porteuses secrètes avec les paramètres $N_o = 2000,\ N_v = 512,\ N_c = 10$. Comme prévu, les distributions pour NW et TNW sont identiques à la distribution hôte. L'affectation donnée par la théorie du transport ne modifie pas la sécurité su schéma classique NW.

L'analyse en composantes principales ACP peut être utilisée afin d'estimer le sous-espace engendré par les porteuses. Si un adversaire construit une base de cet espace, il peut alors rendre les messages illisibles (par exemple en annulant les projections dans le sous-espace estimé). Formellement, l'adversaire obtient une base orthogonale $\hat{\mathbf{U}}$ du sous-espace engendré par les porteuses $\mathbf{U}$ grâce à une ACP. Cette estimation dépend de la puissance d'insertion [69]. La figure 2.25 montre la distance chordale (équation (1.22)) entre les porteuses secrètes et les porteuses estimées par ACP (avec $N_o = 2000$ signaux tatoués) en fonction du paramètre η pour les modulations TNW et TRNW avec $N_c = 10$ bits. Nous remarquons tout d'abord que la valeur de la distance chordale est la même pour les deux méthodes pour toute valeur de η. Ce résultat (avec la figure 2.24) permet de valider la distribution obtenue par la méthode basée sur modèle : identique à la distribution du RNW classique. De plus, cette courbe met pratiquement en valeur la notion de *sous-espace-sécurité*. Lorsque $\eta = 1$ (NW et TNW), la distance chordale vaut 1 (son maximum), le sous-espace estimé par ACP est alors différent du sous-espace privé et la sous-espace-sécurité du NW et du TNW est vérifiée. Nous remarquons que pour les méthodes RNW et TRNW, plus la valeur de η (proportionnelle à la variance des corrélations) augmente, plus les modulations tendent vers la *clé-sécurité*. Nous voyons alors le paramètre η comme un paramètre permettant de faire varier une méthode hybride [1] entre sécurité et robustesse.

1. Comme pour la modulation ρ circulaire (section 2.1.1), le TRNW permet aussi le compromis robustesse/sécurité. Cependant, pour le ρ-CW, le paramètre ρ permettait une continuité entre la clé-sécurité et la non-sécurité alors que pour le TRNW, elle est située entre la sous-espace-sécurité et la clé-sécurité.

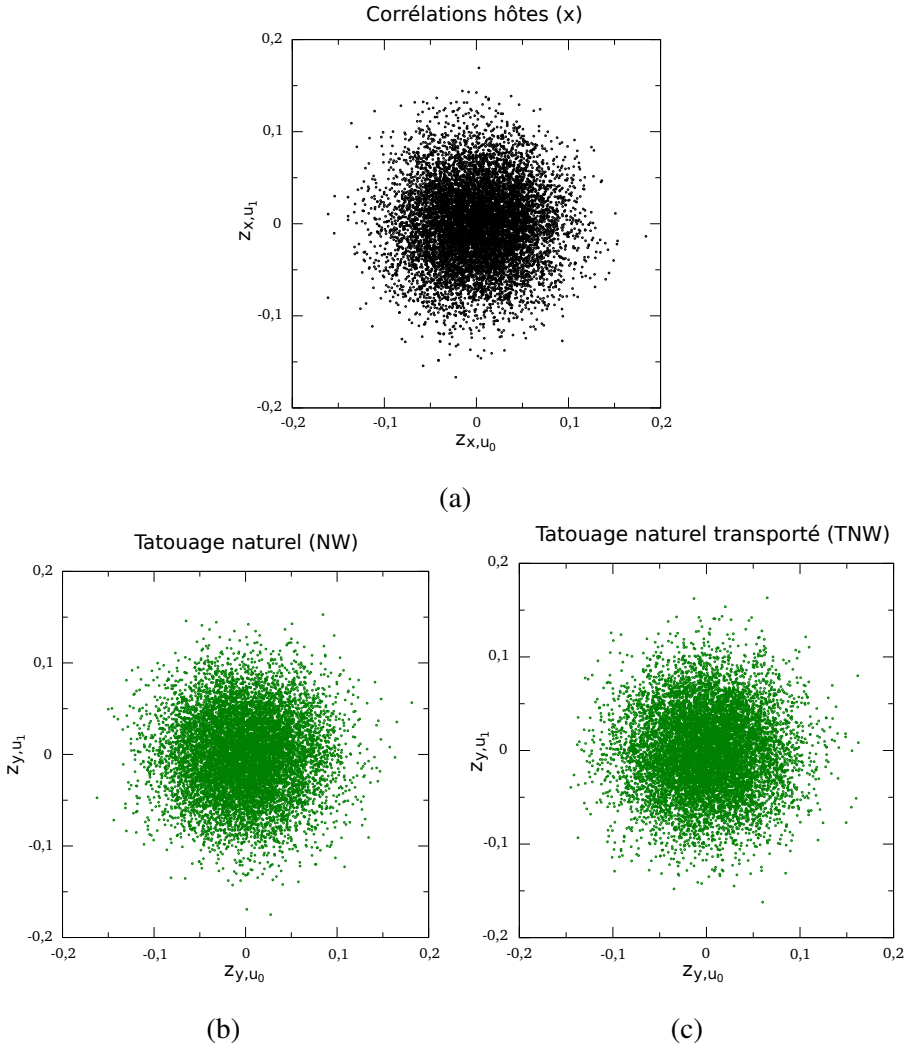


FIGURE 2.24 – Distribution des corrélations entre les signaux hôtes (a), NW (b) et TNW (c) et deux porteuses secrètes avec les paramètres $N_o = 2000,\ N_v = 512,\ N_c = 10$. Comme prévu, les distributions pour NW et TNW sont identiques à la distribution hôte. L'affectation donnée par la théorie du transport ne modifie pas la sécurité su schéma classique NW.

Nous testons à présent la robustesse de ce nouveau schéma basé sur modèle. La figure 2.26 montre le BER (en moyenne sur $N_o = 2000$ signaux) en fonction du WCNR pour les modulations NW et RNW ($\eta = 2$) ainsi que leur variantes transportées TNW et TRNW. Nous remarquons alors que l'utilisation de la théorie du transport ne modifie pas la robustesse des schémas initiaux. De plus nous remarquons que lorsque la variance des corrélations est plus forte ($\eta = 2$), la robustesse face à l'ajout de bruit gaussien en est améliorée.

La table 2.3 quantifie la distorsion (au moyen du WCR) sur $N_o = 2000$ signaux avec $N_c = 10$ bits des méthodes NW, TNW et HNW (respectivement RNW, TRNW et HRNW avec $\eta = 2$) afin de différencier les deux techniques basées sur le modèle du NW (pour cette dernière technique, nous utilisons ici une N_m-map construite avec $N_m = 10000$ signaux). Comme prévu théoriquement, utiliser un plan de transport optimal permet de réduire la distorsion originale du NW de 3.76 dB (contre 2.65 dB en utilisant la méthode des Hongrois) et celle du RNW ($\eta = 2$) de 2.73 dB (contre 2.23 dB en utilisant la méthode des Hongrois). Nous remarquons la supériorité de

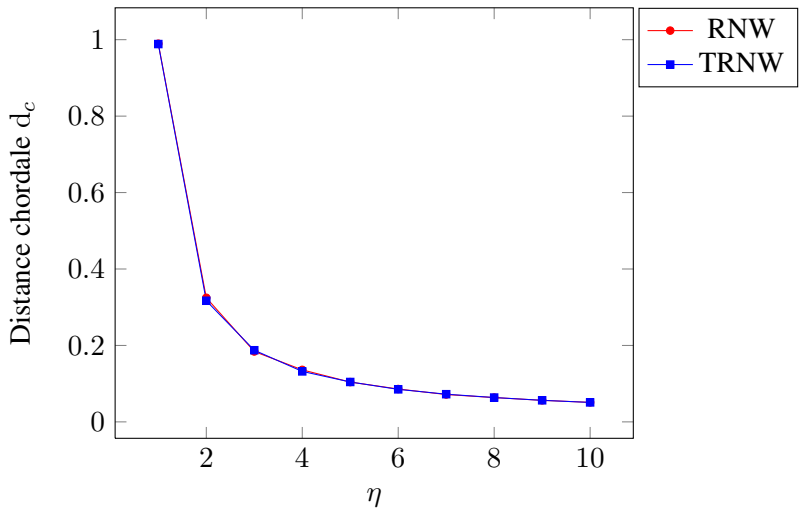


FIGURE 2.25 – Distance chordale entre les porteuses secrètes et les porteuses estimées par un adversaire possédant $N_o = 2000$ signaux tatoués à l'aide d'une ACP en fonction du paramètre η pour les modulations RNW et TRNW (NW et TNW pour $\eta = 1$) avec $N_c = 10$ bits. Nous remarquons que la valeur de la distance chordale est la même pour les deux méthodes pour toute valeur de η. De plus, ce résultat met pratiquement en valeur la notion de sous-espace-sécurité ($\eta = 1$) et de clé-sécurité ($\eta > 1$).

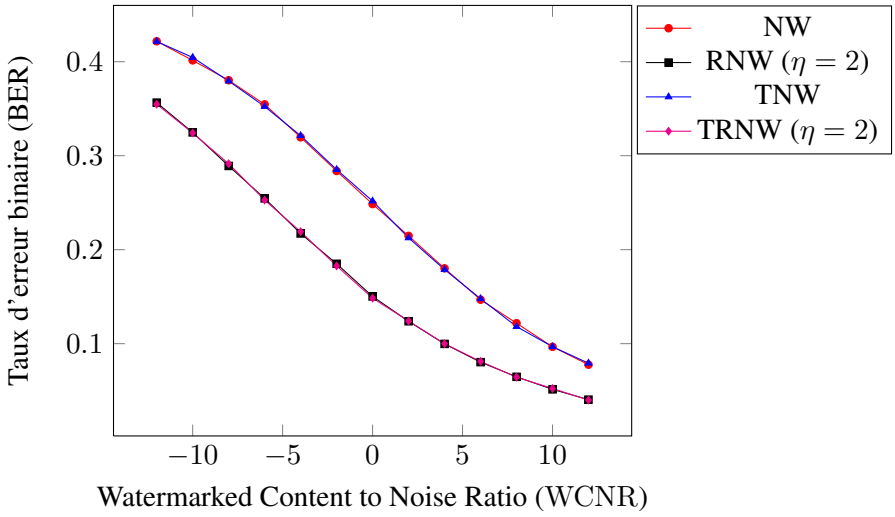


FIGURE 2.26 – BER en fonction du $\mathrm{WCNR}_{[dB]}$ pour les modulations sous-espace-sûres NW et TNW et les modulations RNW et TRNW ($\eta = 2$) avec $N_c = 10$ bits.

la méthode par transport face à la méthode des Hongrois tant du point de vue de la distorsion, que celui de la complexité (l'utilisation d'un plan de transport a une complexité linéaire alors que l'algorithme des Hongrois est en $\mathcal{O}({N_m}^3)$).

La figure 2.27 compare à nouveau la différence en terme de distorsion entre le NW classique et ses améliorations par les techniques basées sur modèle (algorithme des Hongrois et théorie du transport) pour plusieurs valeurs de N_c. Nous pouvons voir de nouveau la supériorité de la méthode basée sur le transport du point de vue de la distorsion dont l'efficacité reste constante selon le nombre de bits insérés contrairement à l'utilisation de

Distribution	Distorsion $\mathrm{WCR}_{[\,dB]}$		
	méthode classique	algorithme Hongrois	plan de transport optimal
NW	-14.09	-16.74	-17.85
RNW ($\eta = 2$)	-10.06	-12.29	-12.79

TABLE 2.3 – Distorsion (WCR) en moyenne sur $N_o = 2000$ signaux pour NW, HNW et TNW (resp. RNW ($\eta = 2$), HRNW et TRNW) avec $N_c = 10$ bits insérés. Nous remarquons la supériorité de la méthode par transport face à la méthode des Hongrois tant du point de vue de la distorsion, que celui de la complexité (l'utilisation d'un plan de transport est linéaire alors que l'algorithme des Hongrois est en $\mathcal{O}({N_m}^3)$.

la méthode des Hongrois dont le gain en distorsion diminue. L'explication que nous donnons et que, même si la méthode des Hongrois est optimale pour trouver le couplage parfait qui minimise le coût total, le tatouage naturel (robuste) Hongrois insère le message en cherchant le plus proche voisin du signal hôte à tatouer dans l'ensemble des sommets de départ pré-calculé $\mathcal{X}$ de la N_m-carte choisie. Cependant, la distance euclidienne entre le vecteur de corrélations hôte par rapport aux porteuses secrètes et celui présent dans la N_m-carte augmente avec la dimension de l'espace des porteuses (de dimension N_c).

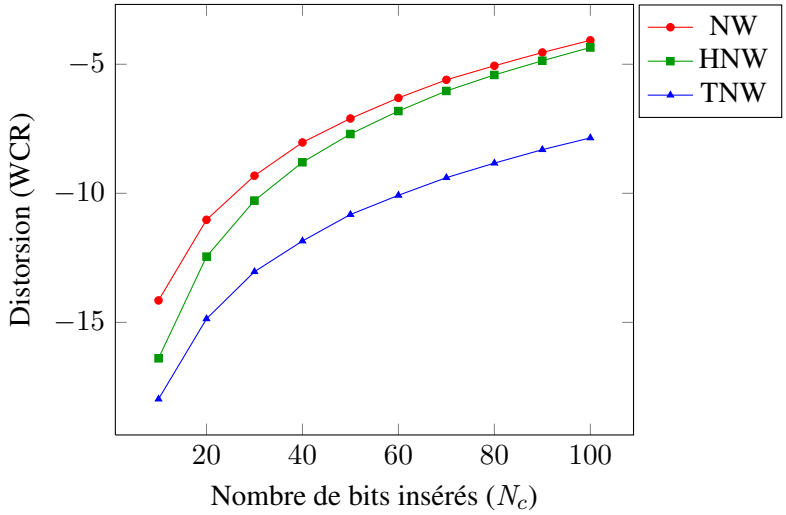


FIGURE 2.27 – WCR en fonction de N_c pour les modulations NW, HNW et TNW. Comme nous pouvons le remarquer, l'amélioration du point de vue de la distorsion est meilleur pour le HNW que pour le TNW lorsque N_c augmente (le gain reste constant pour le TNW alors qu'il tend vers 0 pour le HNW. Pour cette dernière technique, nous perdons l'optimalité de la méthode des Hongrois en choisissant le plus proche voisin de notre vecteur de corrélations hôte dans la N_m-carte pré-calculée.

Dans la table 2.3 nous remarquions de le gain en distorsion apportée par les méthode basées sur modèle est plus important pour une distribution NW que RNW (ici $\eta = 2$). Il est alors intéressant de comparer ce gain pour plusieurs valeurs de η. La figure 2.28 montre le WCR pour RNW, HRNW et TRNW en fonction du paramètre η pour $N_c = 10$ bits insérés. Nous remarquons que le gain en distorsion apportée par les techniques basées sur modèle diminue avec η. De plus la différence entre le HNW et le TNW devient négligeable.

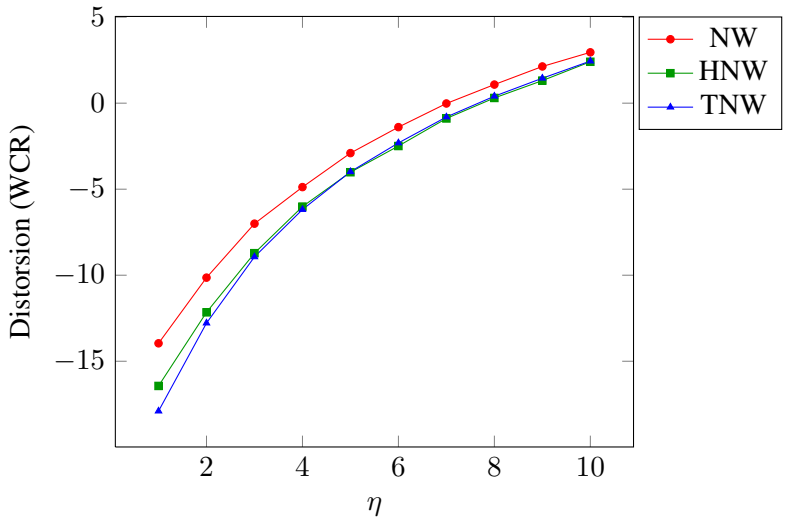


FIGURE 2.28 – Distorsion (WCR) pour les méthodes NW, HNW et TNW en fonction du paramètre η de la distribution RNW pour $N_c = 10$ bits. Nous remarquons que le gain en distorsion est plus élevé lorsque η est proche de 1.

2.4 Conclusion

Dans ce chapitre, nous avons tout d'abord présenté deux nouvelles techniques de tatouage : le tatouage ρ circulaire (ρ-CW), méthode par étalement de spectre hybride entre la clé-sécurité et la non-sécurité ainsi que le tatouage par la loi du χ^2 (χ^2W), méthode permettant d'atteindre la stégo-sécurité. Nos expériences ont une fois de plus mis en valeur le compromis qu'il existe entre les contraintes de robustesse et de sécurité en tatouage : une meilleure sécurité entraine une plus faible résistance aux manipulations que peut subir le support.

Nous avons ensuite présenté une nouvelle manière de tatouer en WOA. Au lieu d'utiliser les formules de modulations pour atteindre le niveau de sécurité souhaité (CW, NW, RNW ou χ^2W), nous générons directement les distributions souhaitées dans la variété privée à laquelle nous affectons ensuite chaque signal hôte. Nous avons rendu cette affectation optimale du point de vue de la distorsion grâce à deux techniques : la méthode des Hongrois, méthode discrète permettant d'affecter deux nuages de points en minimisant la somme totale des déplacements et la théorie du transport, méthode continue réalisant le même objectif mais qui se base sur la distributions des signaux avant et après tatouage. Nos implantations de ces techniques nous a permis d'améliorer les méthodes de tatouage sûres classiques (CW, NW, RNW et χ^2W) en permettant une meilleure robustesse ainsi qu'une diminution de la distorsion sans changer les niveaux de sécurité théoriques des méthodes classiques.

Dans les premier et second chapitres nous avons présenté la sécurité en tatouage de manière assez théorique : les signaux hôtes ont été générés selon une distribution gaussienne. Nous nous intéressons maintenant à l'implantation de nos méthodes dans un cadre pratique : le tatouage d'images fixes où es contraintes liées à l'imperceptibilité et à la robustesse imposent de nouvelles hypothèses de travail.

Chapitre 3

Tatouage d'images naturelles

Goðafoss [1] est une des chutes les plus spectaculaires d'Islande. Elle est localisée dans la région de Mývatn, sur le fleuve Skjálfandafljót et fait 12 m de haut sur 30 m de large. En l'an 1000, le parlement islandais (l'Alþing) décide le passage au christianisme. Pour symboliser l'abandon de ces rites, le chef de clan þorgeir, de retour du þing, décide de jeter dans ces impressionnantes chutes d'eau toutes les idoles païennes. Le nom Goðafoss signifie "la chute des dieux" en souvenir de cet épisode.

Sommaire

1. Auteur : Pierre Kornobis.

CE chapitre présente les résultats obtenus par l'utilisation de méthodes de tatouage par étalement de spectre sur des images naturelles en niveaux de gris dans le contexte d'attaque WOA (l'adversaire possède plusieurs images tatouées avec des messages différents mais avec la même clé secrète). Jusqu'à présent nous travaillions sur des techniques de tatouage dans un cadre théorique (traitements sur signaux gaussiens), dans ce chapitre nous prenons en compte les contraintes propres aux images naturelles, comme le choix du domaine d'insertion de la marque, le système visuel humain pour la contrainte d'imperceptibilité, la compression JPEG (*Joint Photographic Experts Group*) pour la robustesse. Nous avons pris le choix de travailler dans le domaine des ondelettes en utilisant un masquage psychovisuel *a la* Piva [71]. D'une part nous comparons les modulations sûres : tatouage naturel robuste (RNW) et circulaire (CW) avec les méthodes non-sûres : étalement de spectre classique (SS) et amélioré (ISS) du point de vue de la robustesse par compression JPEG et de la clé-sécurité par séparation de sources ACI. D'autre part, nous changeons les paramètres du domaine transformé (changement de niveau lors de la transformée en ondelettes) afin de pouvoir implanter le tatouage naturel (NW). Dans ce nouvel espace de tatouage, nous mettons en valeur la sous-espace-sécurité du tatouage naturel à l'aide d'une attaque par effacement (consistant à annuler les corrélations entre signaux tatoués et porteuses estimées par ACP). Toujours dans ce nouvel espace, nous montrons que les optimisations des modulations sûres basées sur modèle (HCW&TNW) sont applicables en pratique et permettent de diminuer la distorsion sans compromettre la sécurité ni la robustesse des schémas initiaux.

3.1 Schéma de tatouage d'images

La plupart des schémas de tatouage théoriques prennent pour hypothèse des signaux hôtes $\mathbf{x}$ distribués selon une loi gaussienne (distribution circulaire mathématiquement facile à manier). Cependant, ce modèle ne correspond pas aux distributions usuelles des signaux caractéristiques des images (composantes des images satisfaisant les contraintes liées au tatouage comme l'imperceptibilité et la robustesse). Par exemple, les coefficients DCT (transformée en cosinus discrète) sont traditionnellement modélisés par une distribution de Laplace, les coefficients d'ondelettes par une distribution gaussienne généralisée, etc. Pour se rapprocher de l'hypothèse de gaussianité des signaux (particulièrement pour l'application du tatouage naturel NW), nous adoptons l'astuce suivante : nous procédons à une projection des signaux caractéristiques hôtes sur un ensemble de signaux pseudo aléatoires de même loi. En vertu du théorème centrale limite, ces signaux projetés sont alors distribués selon une loi asymptotiquement gaussienne.

3.1.1 Implantation

Le schéma de l'implantation des modulations par étalement de spectre sur des images codées en niveaux de gris (1 octet/pixel) est présenté sur la figure 3.1. Nous voulons tatouer des images de $M \times N$ pixels. Après une transformée en 3 niveaux CDF 9/7 (Cohen, Daubechies et Fauveau) [23], nous arrangeons les neuf premières sous-bandes (N_t composantes hautes fréquences) de l'image hôte originale I_o dans un signal caractéristique $\mathbf{x_t} \in \mathbb{R}^{N_t}$. Nous avons vu précédemment que pour pouvoir appliquer le tatouage naturel NW et le tatouage naturel robuste RNW, une distribution gaussienne des contenus était à privilégier (distribution circulaire). Afin de respecter cette condition nous projetons notre signal hôte caractéristique sur des signaux uniformes afin d'obtenir un signal $\mathbf{x} \in \mathbb{R}^{N_v}$:

$$\forall i \in [N_v],\ \mathbf{x}(i) = \frac{2\sqrt{3}}{\sqrt{N_t}} \sum_{j=0}^{N_t-1} \mathbf{x_t}(j)\mathbf{a}_i(j). \tag{3.1}$$

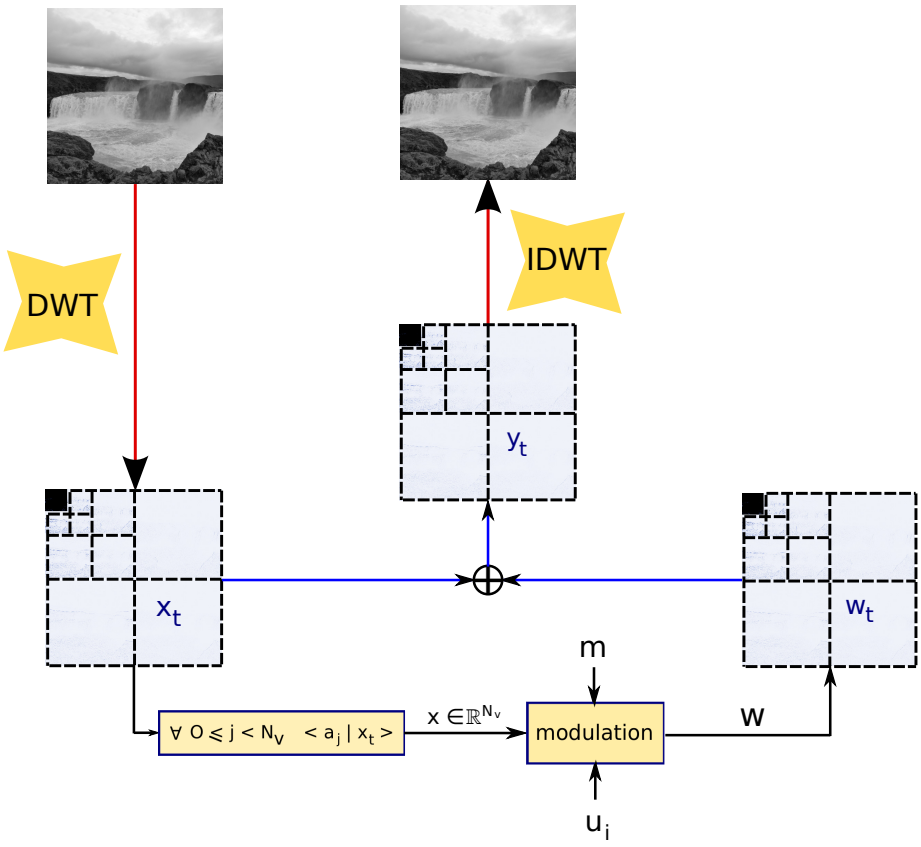


FIGURE 3.1 – Schéma de tatouage expérimental.

Les signaux $\mathbf{a}_i$ sont pseudo aléatoires, distribués uniformément ($\forall i \in [N_v],\ \mathbf{a}_i \sim \mathcal{U}[-0.5, 0.5]$), le rapport $2\sqrt{3}$ est utilisé pour compenser la variance d'une variable uniformément distribuée. La projection est orthogonale. Le signal de tatouage $\mathbf{w}$ est alors produit par tatouage par étalement de spectre à partir d'un message aléatoire $\mathbf{m} \in \mathbb{F}_2$ et d'un ensemble de porteuses secrètes $\{\mathbf{u}_i\}_{i\in[N_c]}$ (section 1.5).

La rétro-projection du signal $\mathbf{w}$ dans le domaine des ondelettes est définie par l'équation suivante :

$$\forall i \in [N_t],\ \mathbf{w_t}(i) = \frac{2\sqrt{3}}{\sqrt{N_t}} \sum_{j=0}^{N_v-1} \mathbf{w}(j)\mathbf{a}_j(i). \tag{3.2}$$

Finalement, nous construisons le signal caractéristique tatoué $\mathbf{y_t}$ dans le domaine des ondelettes (insertion constante) :

$$\mathbf{y_t} = \mathbf{x_t} + \mathbf{w_t}, \tag{3.3}$$

L'image tatouée I_t est produite par la transformée en ondelettes discrète inverse.

La distorsion provoquée par la tatouage est évaluée grâce au PSNR (*Peak Signal-to-Noise Ratio*) :

$$\mathrm{PSNR}_{[dB]}(I_o, I_t) = 10 \log_{10} \left(\frac{255^2}{\mathrm{EQM}(I_o, I_t)} \right), \tag{3.4}$$

où EQM désigne l'erreur quadratique moyenne entre deux matrices réelles.

Dans ce chapitre, la distorsion est souvent calculée en moyenne sur N_o images, nous calculerons alors le PSNR moyen comme une fonction de la moyenne des EQM entre images originales I_o et images tatouées I_t :

$$\mathbb{E}[\mathrm{PSNR}]_{[dB]}(I_o, I_t) = 10\log_{10}\left(\frac{255^2}{\mathbb{E}[\mathrm{EQM}(I_o, I_t)]}\right). \tag{3.5}$$

Lemme 1 *Dans le cas d'une insertion constante, le lien entre PSNR et WCR est donné par l'équation suivante :*

$$\mathrm{WCR} = 10\log_{10}\left(\frac{255^2}{\sigma^2_{\mathbf{x}}} \times \frac{M \times N}{N_v}\right) - \mathrm{PSNR}. \tag{3.6}$$

Preuve Le premier point est que, grâce à la normalisation de la rétro-projection du signal de tatouage dans le domaine des ondelettes donnée par l'équation (3.2), la distorsion reste constante dans le domaine des ondelettes et dans le domaine projeté (de dimension N_v), nous avons :

$$\|\mathbf{w}_t\|^2 = \|\mathbf{w}\|^2 = d^2. \tag{3.7}$$

Nous obtenons alors :

$$\sigma^2_{\mathbf{w}_t} = \frac{d^2}{N_t}, \tag{3.8}$$

$$\sigma^2_{\mathbf{w}} = \frac{d^2}{N_v}, \tag{3.9}$$

et donc :

$$\sigma^2_{\mathbf{w}} = \frac{N_t}{N_v}\sigma^2_{\mathbf{w}_t}. \tag{3.10}$$

L'erreur quadratique moyenne dans le domaine spatial (pixellique) est donnée par :

$$\mathrm{EQM} = \frac{N_t}{M \times N}\sigma^2_{\mathbf{w}_t}. \tag{3.11}$$

Toutefois, le PSNR vaut :

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{255^2}{\frac{N_t}{M\times N}\sigma^2_{\mathbf{w}_t}}\right). \tag{3.12}$$

D'après l'équation précédente, nous obtenons :

$$\sigma^2_{\mathbf{w}} = 255^2\frac{M \times N}{N_v}10^{-\frac{\mathrm{PSNR}}{10}}. \tag{3.13}$$

D'après l'équation du WCR donnée par :

$$\mathrm{WCR} = 10\log_{10}\left(\frac{\sigma^2_{\mathbf{w}}}{\sigma^2_{\mathbf{x}}}\right), \tag{3.14}$$

on obtient :

$$\mathrm{WCR} = 10\log_{10}\left(\frac{255^2}{\sigma^2_{\mathbf{x}}} \times \frac{M \times N}{N_v}\right) - \mathrm{PSNR}. \tag{3.15}$$

C.Q.F.D.

3.1.2 Contraintes psychovisuelles

Dans le cadre d'un tatouage sur des images naturelles, il est possible d'exploiter le Système Visuel Humain (SVH) pour améliorer l'insertion du signal de tatouage. Pratiquement :

- le SVH est moins sensible aux zones de forte activité (textures, contours),
- le SVH est plus sensible aux zones de faible activité (aplats).

L'exploitation de ces faiblesses dans notre algorithme de tatouage courant est faite par l'addition d'un masque psychovisuel dans le domaine des ondelettes. Nous utilisons une insertion multiplicative [71] :

$$\mathbf{y_t} = \mathbf{x_t} + \mathbf{w'_t} \text{ avec } \forall i \in [N_t],\ \mathbf{w'_t}(i) = \frac{1}{\mathbb{E}[|\mathbf{x_t}|]}|\mathbf{x_t}(i)|\mathbf{w_t}(i). \tag{3.16}$$

Le facteur $\frac{1}{\mathbb{E}[|\mathbf{x_t}|]}$ est utilisé pour conserver les corrélations entre les porteuses et le signal tatoué dans le sous-espace privé de dimension N_c et pour éviter les erreurs de détection [32].

Lemme 2 *Dans le cas d'une insertion multiplicative, le lien entre PSNR et WCR est donné par l'équation suivante :*

$$\mathrm{WCR} = 10\log_{10}\left(\frac{255^2}{\sigma^2_{\mathbf{x}}} \times \frac{M \times N}{N_v} \times \frac{\mathbb{E}[|\mathbf{x_t}|]^2}{\mathbb{E}[{\mathbf{x_t}}^2]}\right) - \mathrm{PSNR}. \tag{3.17}$$

Preuve D'après [71], le signal de tatouage varie avec la valeur absolue du coefficient d'ondelettes courant que l'on veut tatouer, en supposant que $\forall i \in [N_t],\ \mathbf{x_t}(i)$ est indépendant de $\mathbf{w_t}(i)$. Nous avons :

$$\begin{aligned}
\|\mathbf{w'_t}\|^2 &= \frac{1}{\mathbb{E}[|\mathbf{x_t}|]^2}\sum_{i=0}^{N_t-1}|\mathbf{x_t}(i)|^2\mathbf{w_t}(i)^2 && (3.18)\\
&\simeq \frac{1}{\mathbb{E}[|\mathbf{x_t}|]^2}\frac{1}{N_t}\sum_{i=0}^{N_t-1}\mathbf{x_t}(i)^2\sum_{i=0}^{N_t-1}\mathbf{w_t}(i)^2 && (3.19)\\
&= \frac{\mathbb{E}[{\mathbf{x_t}}^2]}{\mathbb{E}[|\mathbf{x_t}|]^2}\|\mathbf{w_t}\|^2. && (3.20)
\end{aligned}$$

L'équation (3.10) devient :

$$\sigma^2_{\mathbf{w}} = \frac{\mathbb{E}[{\mathbf{x_t}}^2]}{\mathbb{E}[|\mathbf{x_t}|]^2}\frac{N_t}{N_v}\sigma^2_{\mathbf{w'_t}}. \tag{3.21}$$

En reprenant la même démonstration que pour l'insertion constante : équations (3.11) à (3.14), nous obtenons :

$$\mathrm{WCR} = 10\log_{10}\left(\frac{255^2}{\sigma^2_{\mathbf{x}}} \times \frac{M \times N}{N_v} \times \frac{\mathbb{E}[|\mathbf{x_t}|]^2}{\mathbb{E}[{\mathbf{x_t}}^2]}\right) - \mathrm{PSNR}. \tag{3.22}$$

C.Q.F.D.

3.1.3 Valeurs numériques et hypothèses d'implantation

Dans ce chapitre, nous implantons notre schéma de tatouage d'image avec les paramètres suivants : $M = N = 512$, $N_t = 258048$, $N_v = 256$. Nous voulons cacher $N_c = 10$ bits dans chaque image. Le PSNR est fixé à $45\ dB$ pour les quatre modulations suivantes : SS, ISS, RNW, CW. À moins d'être clairement explicité, nous utiliserons une insertion multiplicative. L'implantation du NW sous-espace-sûr (avec $\eta = 1,\ N_v \neq N_c$, voir section 1.7.2) n'est pas évidente à cause de la fragilité de cette modulation. En effet, la variance du signal de tatouage est trop faible et produit une marque fragile (à cause de la quantification de l'image marquée sur 1 octet/pixel). Un moyen de résoudre ce problème est de produire une marque dans les sous-bandes d'ondelettes

correspondant aux fréquences moyennes de l'image hôte. La section 3.3.2 présente une implantation du NW sur des images naturelles.

Afin d'avoir une représentation en deux dimensions des nuages de points des corrélations entre porteuses et signaux, nous utiliserons $N_c = 2$ lorsque cela est nécessaire, permettant ainsi de raisonner intuitivement sur les constellations de points en BPSK (avec notre méthode il est naturellement possible d'insérer autant de bits que nécessaire dans chaque image). Les tests sont effectués sur $N_o = 2000$ images provenant de la base de données BOWS2-original [9].

3.2 Tests sur images naturelles

3.2.1 Distorsion et imperceptibilité

Nous avons implanté les quatre modulations SS, ISS, RNW et CW dans notre schéma de tatouage expérimental et exécuté plusieurs tests sur notre base d'images. La distorsion est calculée en espérance sur les contenus tatoués avec un PSNR cible de $45\ dB$. Cependant, la quantification qui consiste à coder l'image sur 1 octet/pixel produit une variation négligeable sur la distorsion cible.

Nous avons appliqué le masquage psychovisuel défini précédemment dans l'équation (3.16) sur nos quatre modulations. La figure 3.2 montre un exemple d'une image marquée avec une insertion constante (équation (3.6)) et multiplicative avec $\mathbb{E}[\mathrm{PSNR}] = 30\ dB$ (pour mettre en valeur les différences entre les deux approches) pour la modulation CW. La figure 3.3 montre des agrandissements (milieu droit) des images de la figure 3.2.

Nous remarquons alors qu'une insertion multiplicative donne un meilleur rendu visuel et améliore donc l'imperceptibilité du tatouage. Cependant, nous avons choisi de tatouer avec un PSNR cible de $45\ dB$, les différences entre les deux approches sont alors difficiles à cerner à l'œil nu.

La table 3.1 montre les résultats obtenus après tatouage sur les quatre modulations du point de vue de la distorsion pour un PSNR cible de $45\ dB$.

Modulation	$\mathbb{E}[\mathrm{EQM}]$	σ_{EQM}	$\mathbb{E}[\mathrm{PSNR}](\ dB)$
SS	2.18	8.55e-2	44.74
ISS	2.19	2.43e-1	44.72
RNW	2.17	1.03	44.77
CW	2.18	2.41e-1	44.74

TABLE 3.1 – Distorsion causée par l'insertion des messages. Le PSNR cible vaut $45\ dB$.

La modulation RNW est la modulation qui possède le plus grand écart type. Ceci s'explique par le faible degré de liberté accordé par l'insertion par étalement de spectre (limitée à la symétrie et la mise à l'échelle des corrélations hôtes afin de garder une distribution gaussienne). La qualité d'insertion étant correcte ($45\ dB$), la distorsion visuelle est négligeable.

3.2.2 Robustesse face à la compression

Nous avons testé la robustesse des quatre modulations face à la compression JPEG, un traitement classique que subit une image numérique (par exemple lors de sa diffusion sur des pages internet). Pour chaque image,

FIGURE 3.2 – Comparaison entre insertion constante (haut) et multiplicative (bas). Le PSNR cible vaut 30 dB. Modulation : CW ($\mathrm{NCR} = -10\ dB$).

nous insérons un message aléatoire, nous compressons l'image tatouée et nous mesurons le BER en fonction du facteur de qualité sur les 2000 images.

La figure 3.4 représente le BER moyen en fonction du facteur de qualité JPEG pour les quatre modulations. Cette figure montre la supériorité des techniques SS et ISS face aux techniques RNW et CW du point de vue de la robustesse. Cependant, les techniques SS et ISS ne gérant pas la contrainte de sécurité, nous remarquons une fois de plus le compromis qui doit être fait entre robustesse et sécurité. Nous remarquons que, lorsque que le facteur

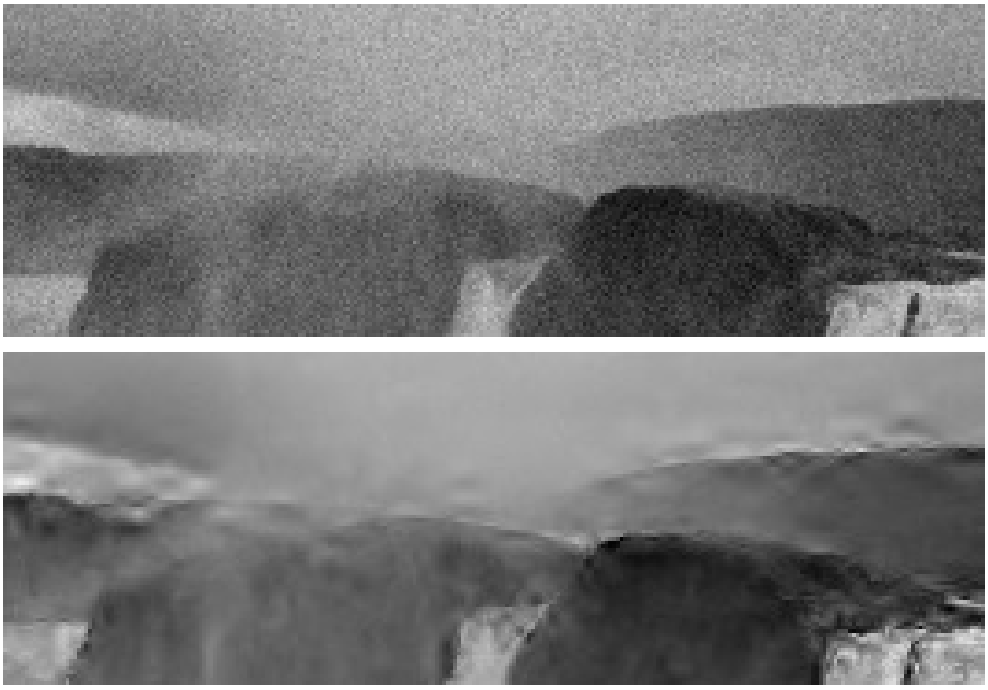


FIGURE 3.3 – Comparaison entre insertion constante (haut) et multiplicative (bas) (agrandissement des images de la figure 3.2). Le PSNR cible vaut 30 dB. Modulation : CW (NCR $= -10$ dB).

de qualité JPEG augmente, la modulation RNW est plus robuste que la modulation CW dans ce cadre de tatouage d'images.

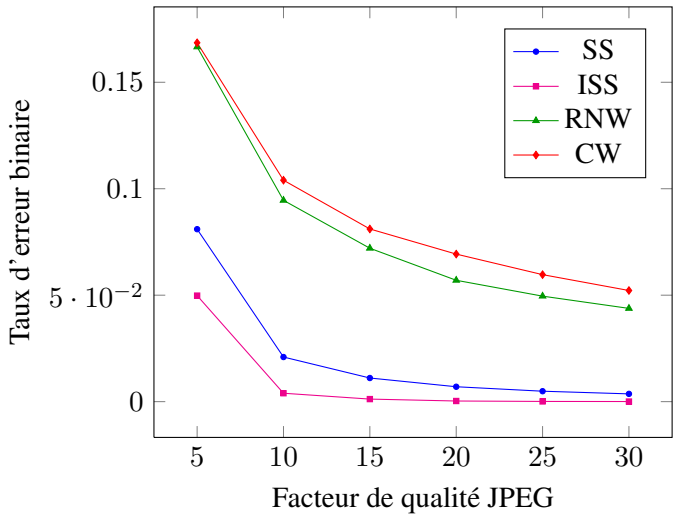


FIGURE 3.4 – Robustesse des modulations SS, ISS, RNW et CW face à la compression JPEG.

3.2.3 Étude de la clé-sécurité

3.2.3.1 Comparaison des distributions des contenus originaux et tatoués

La figure 3.5 représente la distribution des contenus hôtes dans le sous-espace engendré par deux porteuses secrètes. Comme prévu par l'équation (3.1), nous remarquons la gaussianité de la distribution. Nous présentons ensuite les distributions des signaux tatoués avec les modulations SS, ISS, RNW et CW après insertion multiplicative.

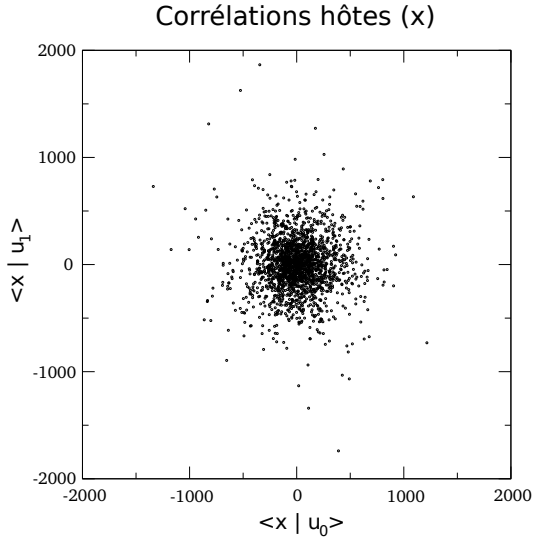


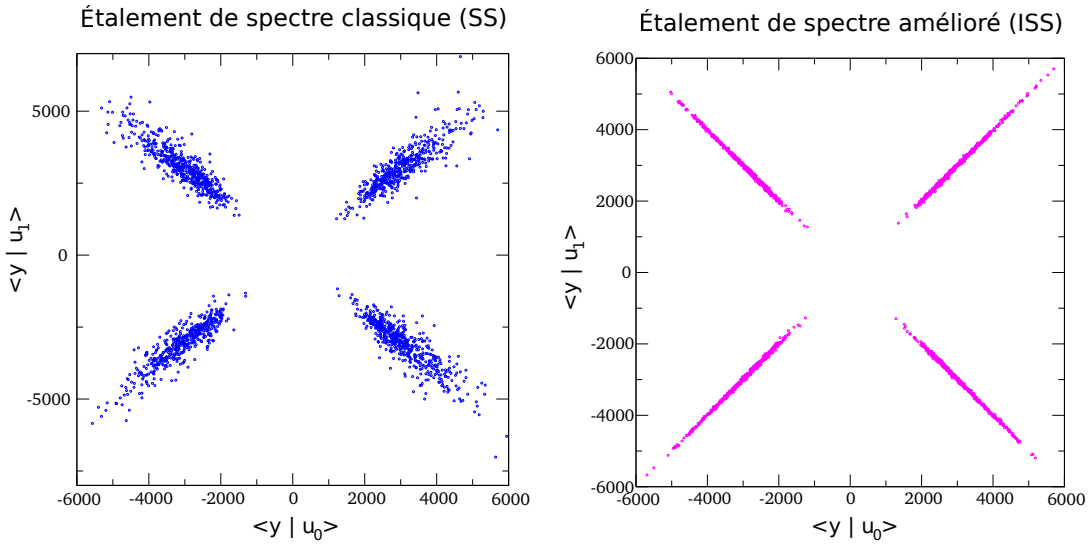
FIGURE 3.5 – Distribution des projections des signaux hôtes sur deux porteuses secrètes. Paramètres : $M = N = 512$, $N_t = 258048,\ N_v = 256,\ N_o = 2000$.

La figure 3.6 montre, respectivement, les distributions SS et ISS sur deux porteuses. Nous remarquons, comme pour les expériences sur signaux gaussiens de la section 1.6 la présence d'une constellation correspondant aux quatre messages possibles avec $N_c = 2 : (0\ 0),\ (0\ 1),\ (1\ 0)$ et $(1\ 1)$. Comme prévu, la modulation ISS diminue la variance des corrélations pour améliorer la robustesse. À l'aide d'une ACP, un adversaire peut estimer le sous-espace engendré par les porteuses ; à l'aide d'une ACI, il est capable d'estimer les régions de mots de code.

Les figures 3.7 et 3.8 montrent respectivement les distributions pour les modulations RNW et CW sur deux porteuses secrètes. Ces distributions sont circulaires. Nous pouvons alors conclure que, pour toutes bases $\{\hat{\mathbf{u}}_0, \hat{\mathbf{u}}_1\}$ de $\mathrm{vect}\,(\mathbf{u}_0, \mathbf{u}_1)$, la distribution $\mathrm{p}(\mathbf{y}_0, \ldots, \mathbf{y}_{N_o-1}|\hat{\mathbf{u}}_0, \hat{\mathbf{u}}_1)$ est identique (rotation des porteuses dans le sous-espace de dimension N_c). Cette propriété est cohérente avec la définition de circularité définie par l'équation (1.27). Par conséquent, ces schémas sont clé-sûrs : un adversaire peut accéder au sous-espace engendré par les porteuses mais n'a aucune information supplémentaire concernant les régions de décodage.

3.2.3.2 Estimation des porteuses

Dans le cadre WOA, d'après le principe de Kerckhoffs, un adversaire a accès au sous-espace projeté de dimension N_v dans lequel a lieu l'ajout de la marque. Cependant, il n'a pas accès aux messages insérés. Il peut seulement émettre l'hypothèse que ces messages sont indépendamment distribués. De plus, nous avons vu dans la section 1.6.1.2 que la technique de séparation de sources en composantes indépendantes (ACI) a ses limites :



(a) (b)

FIGURE 3.6 – Distribution des projections des signaux tatoués sur deux porteuses secrètes pour les modulations SS (a) et ISS (b) après insertion multiplicative. Paramètres : $M = N = 512$, $N_t = 258048$, $N_v = 256$, $N_o = 2000$ et un PSNR cible de $45\ dB$.

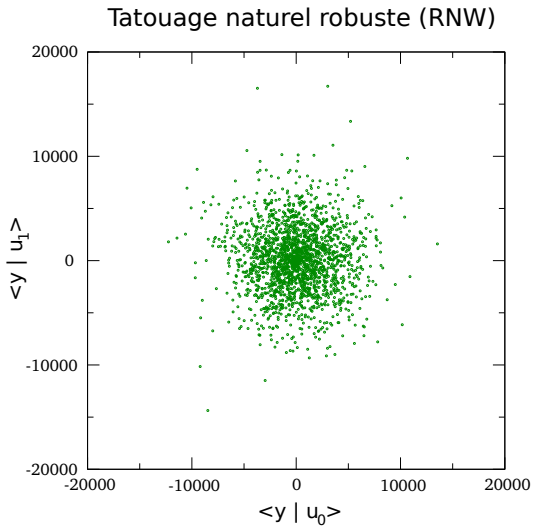


FIGURE 3.7 – Distribution des projections des signaux tatoués sur deux porteuses secrètes pour la modulation RNW après insertion multiplicative. Paramètres : $M = N = 512$, $N_t = 258048$, $N_v = 256$, $N_o = 2000$ et un PSNR cible de $45\ dB$.

1. les porteuses sont estimées au signe près,
2. l'ordre des porteuses n'est pas connu,
3. les porteuses ne peuvent être estimées si les sources ($\mathbf{S}$) sont distribuées selon une loi gaussienne ou sont dépendantes.

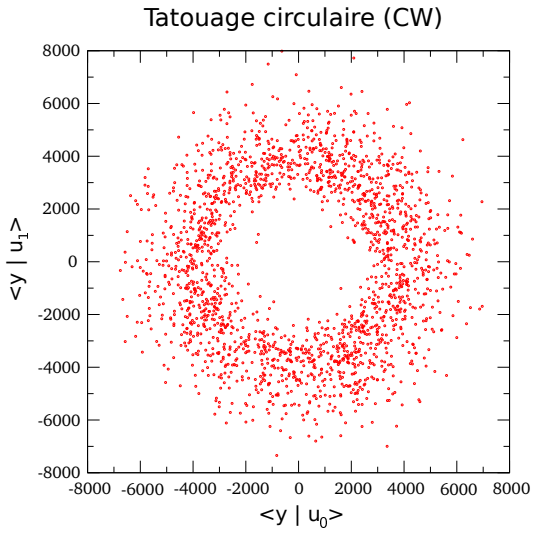


FIGURE 3.8 – Distribution des projections des signaux tatoués sur deux porteuses secrètes pour la modulation CW après insertion multiplicative. Paramètres : $M = N = 512$, $N_t = 258048$, $N_v = 256$, $N_o = 2000$ et un PSNR cible de $45\ dB$.

Nous utilisons alors le score $\mathrm{S}_{\mathrm{est}}$ défini par l'équation (1.23) pour quantifier le degré d'estimation des porteuses secrètes. Sur la figure 3.9, il est clair que les scores $\mathrm{S}_{\mathrm{est}}$ obtenus par un adversaire pour les quatre modulations SS, ISS, RNW et CW illustrent les différences entre la non-sécurité (SS, ISS) et la clé-sécurité (RNW, CW) : les modulations RNW et CW n'autorisent pas une estimation correcte des porteuses. De plus, le score d'estimation obtenu pour une modulation clé-sûre ne dépend pas du nombre de contenus observés N_o (contrairement à une modulation non-sûre où la continuité du score en fonction de N_o est clairement apparente).

3.3 Évaluation de la sous-espace-sécurité

3.3.1 Hypothèses

Jusqu'à présent, nous travaillions dans un espace transformé par ondelettes permettant d'implanter les quatre modulations SS, ISS, CW et RNW. L'implantation de la modulation NW (section 1.7.2) n'était pas possible du fait de la fragilité de la marque qui disparaissait lors de la quantification sur 1 octet/pixel de l'image tatouée.

Nous apportons alors les modifications suivantes au schéma expérimental (figure 3.1) :

- nous utilisons quatre niveaux lors de la transformée en ondelettes,
- nous arrangeons les sous-bandes HL4 et LH4 de notre image hôte dans le vecteur $\mathbf{x_t}$,
- nous choisissons ici d'utiliser une insertion constante (équation (3.3)) afin de simplifier les calculs.
- $N_t = 2048$.

La modulation NW ne permet pas de fixer une distorsion cible moyenne pour chaque image tatouée. Avec les nouveaux paramètres de notre schéma expérimental, nous obtenons un PSNR entre les images originales et tatouées qui vaut $\mathbb{E}[\mathrm{PSNR}] = 46.7\ dB$ en espérance pour la modulation NW. Notons que la modulation NW produit une insertion ne permettant pas de fixer la distorsion souhaitée (PSNR n'est pas fixé pour chaque image).

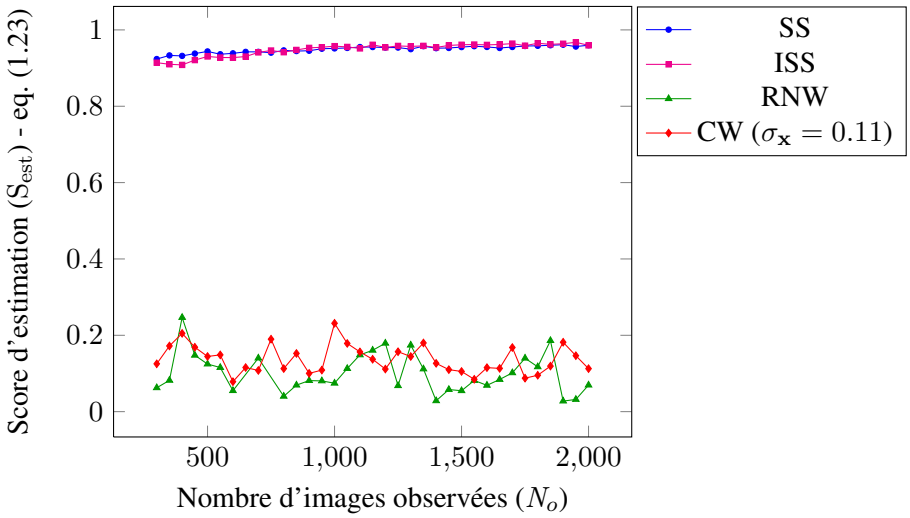


FIGURE 3.9 – Estimation des porteuses secrètes pour les quatre modulations par étalement de spectre en fonction du nombre de contenus observés par l'attaquant pour une attaque de sécurité liée à une technique de séparation de sources. Nous appliquons une ACI sur les signaux tatoués (matrice $N_v \times N_o$) afin d'obtenir une estimation $\{\hat{\mathbf{u}}_j\}_{j\in[N_c]}$ des porteuses. Ensuite, nous calculons le score d'estimation S défini par l'équation (1.23). Un score S proche de 1 correspond à une estimation correcte. Notons que l'ACI se fait en additionnant des signaux tatoués (colonnes de la matrice $N_v \times N_o$) lorsque N_o augmente.

3.3.2 Implantation du tatouage naturel

Le tatouage naturel (NW) permet un tatouage sous-espace-sûr lorsque nous avons $\eta = 1$, $N_v \neq N_c$. Nous avons testé le tatouage NW sur notre base d'images avec les nouveaux paramètres de notre schéma expérimental cités précédemment. La figure 3.10 montre, respectivement, la distribution des signaux hôtes et tatoués par NW sur deux porteuses. Comme nous pouvons le voir, la distribution des corrélations après tatouage est la même que la distribution des corrélations hôtes (à la variance des signaux près). Ceci est en accord avec la définition de sous-espace-sécurité.

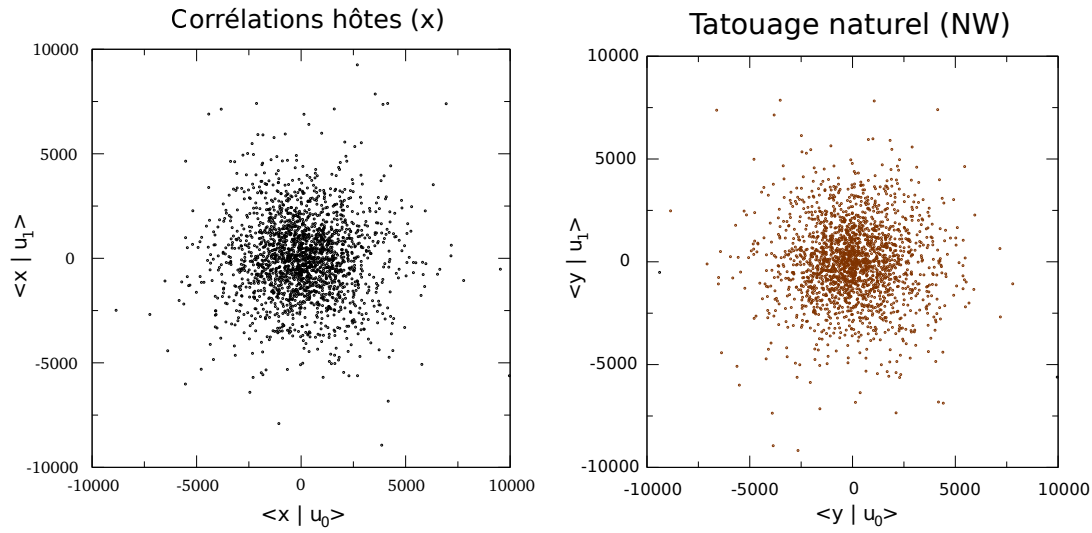
3.3.3 Attaque par effacement

Pour évaluer et comparer la sécurité des implantations NW et CW, nous proposons une **attaque par effacement** basée sur une ACP, qui permet de modifier le message inséré : la puissance d'attaque est entièrement dirigée dans le sous-espace privé de dimension N_c. Si l'adversaire construit une base de ce sous-espace, il est capable de rendre le message illisible. Concrètement, il récupère une base grâce à une ACP : $\{\hat{\mathbf{u}}_j\}_{j\in[N_c]}$ du sous-espace défini par les porteuses $\{\mathbf{u}_i\}_{i\in[N_c]}$. Pour chaque signal tatoué $\mathbf{y}$, l'adversaire construit le signal attaqué $\mathbf{r}$ par :

$$\mathbf{r} = \mathbf{y} - \sum_{j=0}^{N_c-1} \frac{\langle \mathbf{y}|\hat{\mathbf{u}}_j\rangle}{\langle \hat{\mathbf{u}}_j|\hat{\mathbf{u}}_j\rangle}\hat{\mathbf{u}}_j. \tag{3.23}$$

Le procédé d'effacement tend à annuler les composantes de $\mathbf{y}$ qui sont colinéaires avec chaque porteuse estimée $\hat{\mathbf{u}}_j$ afin de rendre le bit décodé aléatoire pendant la phase de décodage du contenu par un distributeur.

Nous avons implanté cette attaque avec les modulations NW et CW dans le nouveau schéma expérimental de tatouage défini dans cette section : pour CW, le PSNR cible est fixé à $46.7\ dB$ (correspondant à la distorsion en



(a) (b)

FIGURE 3.10 – Distribution des projections des signaux hôtes (a) et tatoués par la modulation NW (b) sur deux porteuses secrètes. Paramètres : $M = N = 512$, $N_t = 2048,\ N_v = 256,\ N_o = 2000,\ \mathbb{E}(\mathrm{PSNR}) = 46.7\ dB$.

	Avant attaque		Après attaque	
	$E(\mathrm{PSNR})$ original/tatoué	BER	$E(\mathrm{PSNR})$ tatoué/attaqué	BER
NW	46.73	0.01	45.9	0.05
CW	46.64	0.03	45.56	0.3

TABLE 3.2 – Attaque par effacement pour les modulations TNW et HCW.

moyenne provoquée par la modulation NW). Les résultats obtenus sur notre base d'images sont présentées dans la table 3.2.

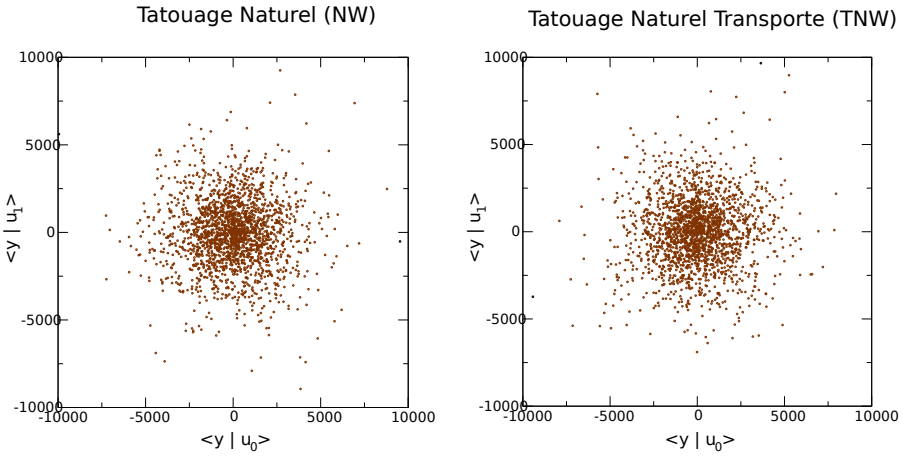
Dans le cas du tatouage CW, après attaque, la distorsion est correcte et 30% des bits ne sont pas lus correctement : nous avons vérifié que le CW n'est pas sous-espace-sûr. Pour NW, l'attaque n'est pas efficace, 95% des messages sont bien décodés. Dans ce cas, cette attaque devient une attaque classique par ajout de bruit gaussien (AWGN). En fait, l'attaque par ACP sur le tatouage naturel retourne une matrice de mélange (porteuses) et une matrice source (modulations) aléatoires. En accord avec les approches théoriques, la sous-espace-sécurité du NW est vérifiée.

3.4 Implantation des méthodes basées sur modèle

Dans cette section nous implantons deux schémas basés sur modèle : le tatouage naturel transporté TNW ainsi que le tatouage circulaire hongrois HCW. Nous évaluons le déficit en distorsion apportées par ces deux méthodes par rapport aux modulations classiques. Nous assurons aussi la validation de ces deux schémas des points de vue robustesse (ajout de bruit gaussien) et de sécurité (comparaison des distributions et calcul de la distance chordale après ACP).

3.4.1 Implantation du tatouage naturel transporté TNW

La figure 3.11 montre la distribution des signaux tatoués par NW (a) et TNW (b) sur deux porteuses. Comme nous pouvons le constater, la distribution des corrélations après tatouage est la même pour les deux techniques. Conformément à l'approche théorique, l'utilisation de la théorie du transport permet d'obtenir la distribution souhaitée.



(a) (b)

FIGURE 3.11 – Corrélations entre deux porteuses secrètes et 2000 signaux tatoués $\mathbf{y}$ pour les modulations NW (a) et TNW (b).

La figure 3.12 quantifie le déficit en distorsion apporté par la méthode TNW. Pour un PSNR moyen de $46.69\ dB$ pour la méthode TN, nous obtenons un PSNR de $50.15\ dB$ pour la méthode TNT.

Nous avons de plus calculé le taux d'erreur binaire moyen en fonction de la variance de bruit gaussien $\mathbf{n}$ ajouté dans le domaine pixellique pour la modulation NW ainsi que son amélioration TNW (figure 3.13). Nous remarquons que la méthode TNW ne modifie pas la robustesse du schéma initial NW.

3.4.2 Implantation du tatouage circulaire hongrois HCW

Dans cette sous-section, nous implantons d'une part le CW, d'autre part le HCW dans le nouveau schéma expérimental défini dans la section 3.3.1, le PSNR est ici fixé grâce à l'équation (3.6). Pour le HCW, nous construisons tout d'abord un N_m-atlas à l'aide de $N_m = 10000$ images naturelles codées en niveaux de gris à raison d'1 octet/pixel (images différentes des 2000 images BOWS2-original).

La figure 3.14 montre la distance chordale entre les porteuses secrètes et les porteuses estimées à l'aide d'une ACP sur nos 2000 images de référence (BOWS2-original) tatouées avec les méthodes CW et HCW (le NCR est fixé à $-10\ dB$ pour le CW ainsi que pour la construction du 10000-atlas pour le HCW). Nous remarquons que plus la distorsion est élevée (PSNR se rapprochant de $30\ dB$), plus il est facile d'estimer le sous-espace privé engendré par les porteuses (caractéristique de la sous-espace-sécurité de la modulation CW). De plus, cette distance chordale peut être utilisée comme un indicateur d'égalité entre les distributions générées par les modulations CW et HCW. Nous remarquons ici que la distance chordale est la même pour les deux modulations, indépendamment de la distorsion.

La figure 3.15 quantifie le déficit en distorsion apporté par la méthode HCW, le PSNR cible est fixé à $45\ dB$ pour la modulation CW ainsi que pour la construction du 10000-atlas pour le HCW. Pour un PSNR moyen de

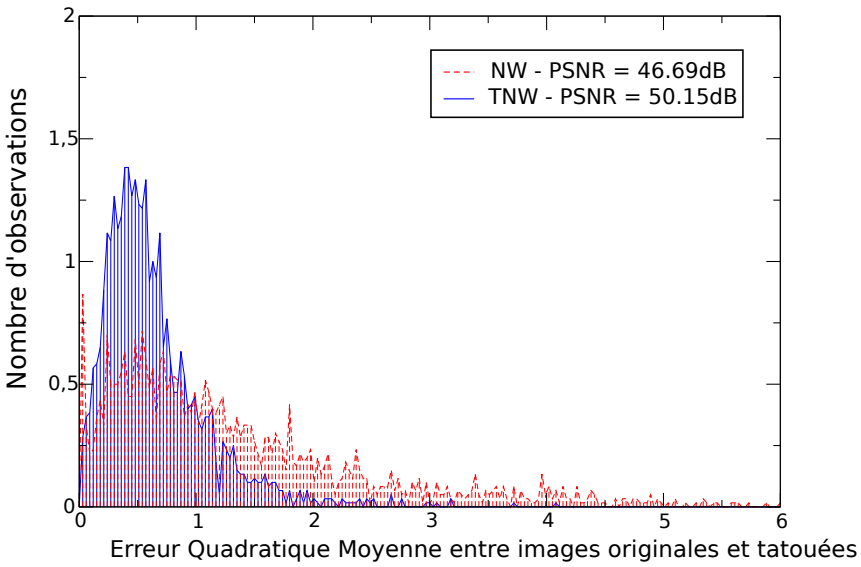


FIGURE 3.12 – Histogrammes des erreurs quadratiques moyennes entre images originales et tatouées pour les méthodes NW et TNW (cette dernière méthode étant appliquée avec la même distribution de corrélations entre signaux tatoués et porteuses secrètes que pour la méthode classique NW). Pour la méthode NW nous obtenons $\mathbb{E}(\text{EQM}) = 1.39,\ \text{std}(\text{EQM}) = 1.38$; pour la méthode TNW nous obtenons $\mathbb{E}(\text{EQM}) = 0.63,\ \text{std}(\text{EQM}) = 0.451$. Nous remarquons que l'utilisation de la méthode TNW permet un gain de $3.46\ dB$ en PSNR moyen. De plus cette méthode réduit considérablement la variance des distorsions sur nos 2000 images.

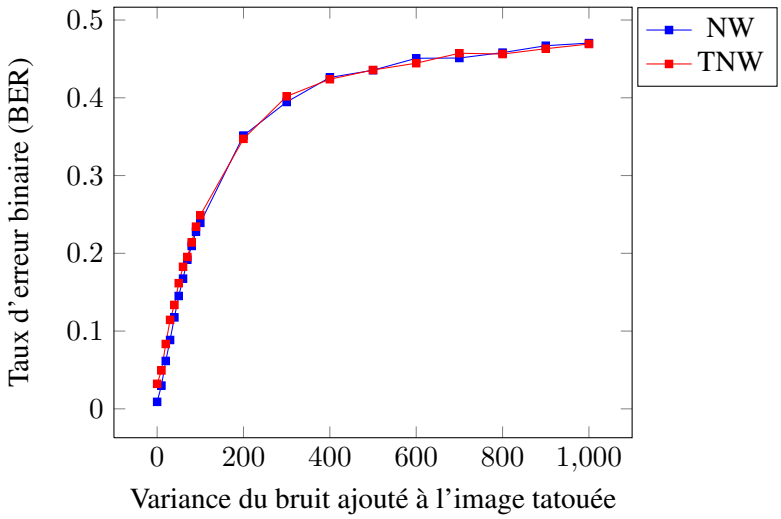


FIGURE 3.13 – Taux d'erreur binaire moyen en fonction de la variance du bruit $\mathbf{n}$ ajouté dans le domaine pixellique pour la modulation NW ainsi que son amélioration TNW. Nous remarquons que la méthode basée sur modèle ne modifie pas la robustesse du schéma initial.

$44.25\ dB$ pour la méthode CW, nous obtenons un PSNR de $45.9\ dB$ pour la méthode HCW.

La figure 3.16 quantifie le gain en PSNR apporté par la méthode HCW par rapport à la méthode CW en fonction du NCR (pour le CW ainsi que pour la construction du 10000-atlas pour le HCW). Nous remarquons

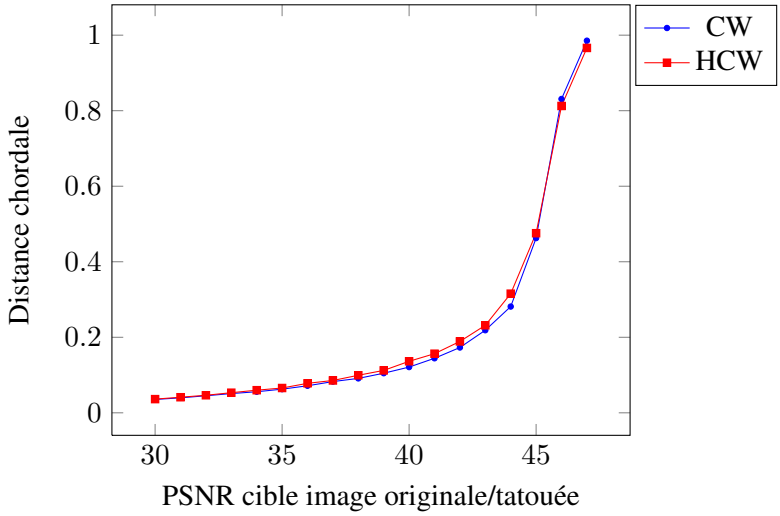


FIGURE 3.14 – Distance chordale entre les porteuses estimées par ACP sur $N_o = 2000$ images (BOWS2-original) tatouées et les porteuses secrètes pour les modulations CW et HCW. Nous remarquons que plus la distorsion est élevée, plus il est facile d'estimer le sous-espace privé engendré par les porteuses. De plus, pour chaque PSNR cible, la distance chordale est la même pour la modulation CW ainsi que son amélioration basée sur modèle HCW.

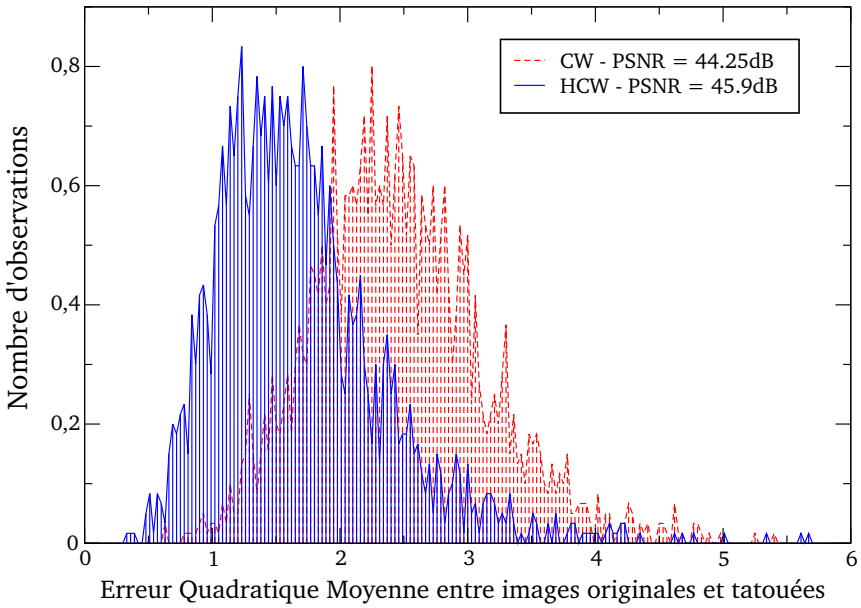


FIGURE 3.15 – Histogrammes des erreurs quadratiques moyennes entre images originales et tatouées pour les méthodes CW et HCW (cette dernière méthode étant appliquée avec la même distribution de corrélations entre signaux tatoués et porteuses secrètes que la méthode classique CW). Pour la méthode CW nous obtenons $\mathbb{E}(\text{EQM}) = 2.44,\ \text{std}(\text{EQM}) = 0.68$; pour la méthode HCW nous obtenons $\mathbb{E}(\text{EQM}) = 1.67,\ \text{std}(\text{EQM}) = 0.68$. Nous remarquons que l'utilisation de la méthode HCW permet un gain de $1.65\ dB$ en PSNR moyen.

que la méthode des Hongrois est plus efficace lorsque la variance des corrélations entre porteuses et signaux

tatoués est faible. La figure 3.17 mesure la robustesse des deux méthodes (taux d'erreur binaire) en fonction de la variance d'un bruit gaussien $\mathbf{n}$ ajouté dans le domaine pixellique de nos images tatouées. Ici, la distorsion cible est fixée avec un PSNR de $45\ dB$ pour le CW ainsi que pour la construction du 10000-atlas pour le HCW. Nous remarquons que la méthode HCW ne modifie pas la robustesse du schéma initial CW.

La figure 3.18 calcule la robustesse entre les deux schémas sous un autre angle. Nous avons fixé le PSNR cible à $45\ dB$ pour la méthode CW. Cependant, pour la construction du 10000-atlas pour le HCW, nous avons dévalué le PSNR cible afin d'obtenir une distorsion moyenne identique à la méthode CW lors des tests sur les 2000 images. Le PSNR moyen obtenu pour les deux modulations avoisine alors les $44.3\ dB$. Nous remarquons la possibilité de convertir le déficit en distorsion qu'offre la méthode basée sur modèle en gain de robustesse lorsque les deux distorsions moyennes finales sont identiques : la méthode HCW est alors plus robuste que la méthode CW.

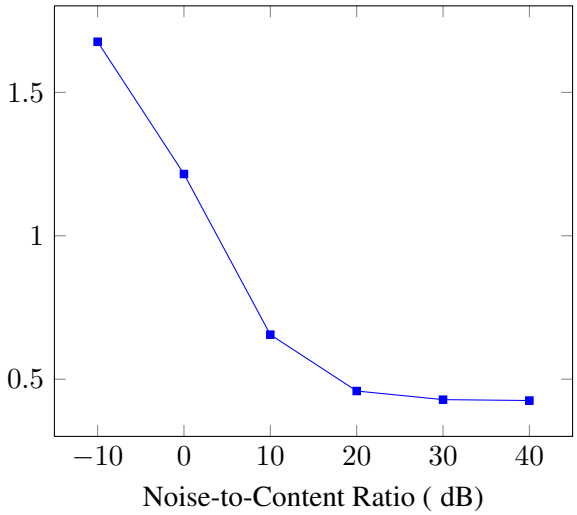


FIGURE 3.16 – Déficit en distorsion (différence entre PSNR moyens) apportée par la méthode basée sur modèle HCW sur la méthode classique CW en fonction du NCR cible. Nous remarquons que la méthode des Hongrois est plus efficace lorsque la variance des corrélations entre porteuses et signaux tatoués est faible.

3.5 Conclusion

Dans ce chapitre, nous nous sommes intéressé à l'implantation des méthodes de tatouage par étalement de spectre sur des images fixes codées en niveaux de gris. Nous avons tout d'abord comparé les méthodes non-sûres SS et ISS avec les méthodes sûres CW et RNW du point de vue de la robustesse (compression JPEG) et de la clé-sécurité (ACI). Notons que nous avons ici utilisé un masquage psychovisuel *a la* Piva [71] afin d'améliorer l'imperceptibilité de nos tatouages. Conformément aux approches théoriques, les techniques CW et RNW résistent moins à notre attaque de robustesse que les méthodes SS et ISS. Cependant, cette perte de robustesse est compensée par leur sécurité dans la mesure ou la séparation de source par ACI ne permet pas une estimation correcte des porteuses quelque soit le nombre de contenus possédés par l'adversaire (contrairement aux méthodes SS et ISS).

Nous avons ensuite mis en valeur la sous-espace-sécurité du NW grâce à une attaque par effacement consistant à annuler les corrélations dans le sous-espace estimé par ACP. En effet, une ACP sur la méthode CW (qui n'est que clé-sûre) est efficace. L'attaque par effacement est alors ciblée dans le sous-espace engendré par les porteuses et

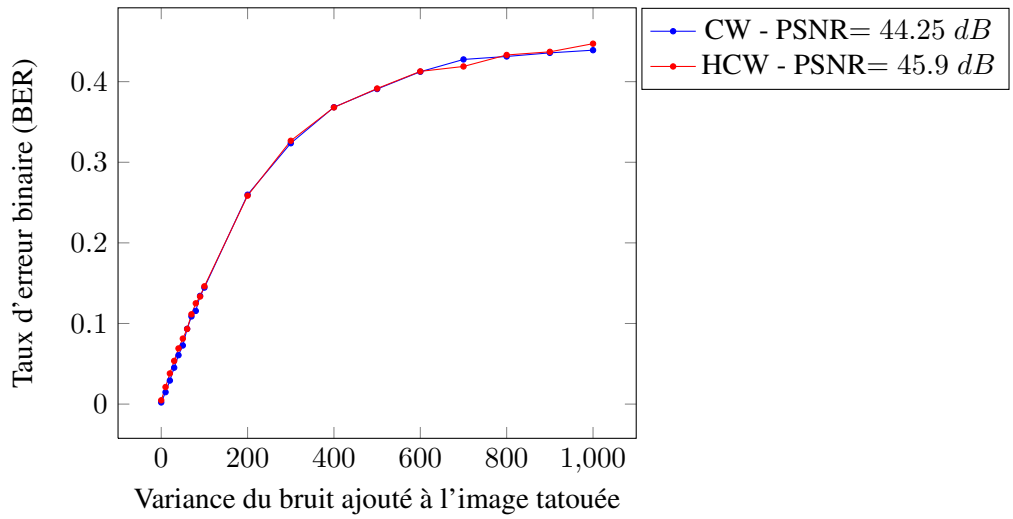


FIGURE 3.17 – Taux d'erreur binaire moyen en fonction de la variance du bruit $\mathbf{n}$ ajouté dans le domaine pixellique pour la modulation CW ainsi que son amélioration HCW basée sur une N_m-carte construite avec la même distorsion cible (distributions des corrélations entre signaux tatoués et porteuses secrètes identiques pour les deux méthodes). Nous remarquons que la méthode basée sur modèle ne modifie pas la robustesse du schéma initial.

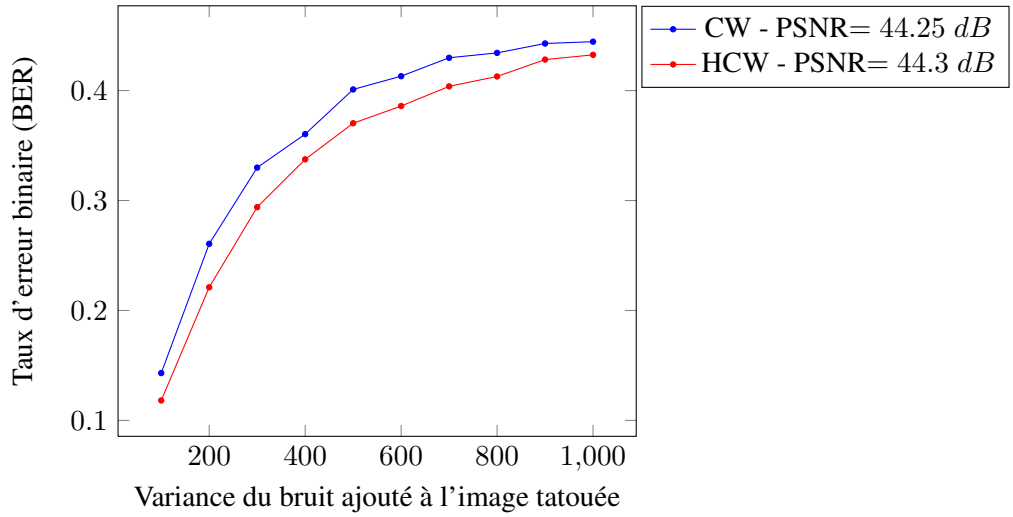


FIGURE 3.18 – Taux d'erreur binaire moyen en fonction de la variance du bruit $\mathbf{n}$ ajouté dans le domaine pixellique pour la modulation CW ainsi que son amélioration HCW basée sur une N_m-carte avec une distorsion cible inférieure de façon à obtenir le même PSNR final pour les deux modulations. Nous remarquons la possibilité de convertir le déficit en distorsion qu'offre la méthode basée sur modèle en gain de robustesse lorsque les deux distorsions moyennes finales sont identiques.

permet une modification de 30% des bits originaux contrairement à la méthode NW, où 95% des bits du message sont correctement décodés après attaque.

Finalement, nous implanté les méthodes basées sur modèle HCW et TNW dans notre schéma de tatouage d'images et avons montré que celles-ci sont utilisables en pratique. Conformément aux résultats obtenus sur signaux hôtes gaussien dans le deuxième chapitre, les méthodes basées sur modèle permettent de diminuer la

distorsion provoquée par les méthodes classiques sans modifier leur niveau de sécurité (gain de $3.46\ dB$ en PSNR pour le HNW).

Cette première partie de ce manuscrit de thèse concernait la sécurité en tatouage. Nous nous intéressons dans la prochaine partie à un problème particulier que les méthodes de tatouage peuvent résoudre : l'estampillage de contenus multimédia consistant à tracer les responsables d'une diffusion illégale d'une œuvre.

Deuxième partie

L’estampillage ou traçage de documents multimédia

Chapitre 4

Les codes traçants

Empreinte digitale d'un individu [1] : le caractère unique de cette empreinte en fait un outil biométrique très utilisé pour l'identification des individus en médecine légale et pour la police scientifique.

Sommaire

1. Auteur : Katpatuka, sous licence CC-BY-SA-3.0 (www.creativecommons.org/licenses/by-sa/3.0/).

L'ESTAMPILLAGE ou traçage de traitres (*fingerprinting* ou *traitor tracing*) consiste à marquer les copies d'un contenu numérique comme par exemple un film obtenu par VOD (*Video On Demand*) que le distributeur veut fournir à ses utilisateurs. Chaque copie est marquée par le distributeur avec la séquence qui identifie un utilisateur spécifique, le code d'estampillage est alors utilisé afin de tracer le responsable de la diffusion d'une copie non autorisée sur des réseaux d'échanges : téléchargement direct (Megaupload, Rapidshare, etc.) ou échange point à point (Emule, Torrents).

Les codes de traçage ont tout d'abord été étudiés en 1983 par Wagner [82]. Il définit les empreintes d'estampillage comme les *caractéristiques d'un objet qui tend à se distinguer d'autres objets similaires* et prend des exemples historiques d'empreintes utilisées à des fins de traçage pour des objets non numériques :

- empreintes digitales : estampillage naturel destiné à identifier de manière unique un individu,
- droits d'auteur sur les tables de logarithmes : celle-ci sont générées en changeant les chiffres les moins significatifs [63],
- explosifs fabriqués avec des particules codées qui peuvent être récupérées après explosion et servent à approximer la date de la vente [3].

Il devient utile de préciser les acteurs intervenant dans la problématique de l'estampillage. Nous reprenons la définition 1 de la section 1.3 à laquelle nous ajoutons la notion de **coalition** :

Définition 3 *: **Coalition (collusion)***
Groupement d'un (ou de plusieurs) adversaire(s), ces derniers mettent en commun leur copie de l'œuvre et tentent de détecter la localisation de l'empreinte afin de la modifier et ainsi de ne pas pouvoir être accusé par le distributeur (éventuellement, la coalition peut forger un contenu dont l'empreinte coïncide avec celle d'un utilisateur innocent).

Le distributeur cherchera alors à identifier le ou les utilisateurs ayant aidé un adversaire ou une coalition, ces derniers chercheront à empêcher l'identification par le distributeur des utilisateurs corrompus. Les techniques d'estampillage se distinguent selon quatre divisions orthogonales (classification de Wagner [82]) :

i) **Première division** :
   - estampillage *numérique*,
   - estampillage *physique* ;

ii) **Seconde division** :
   - estampillage *parfait* : tout altération de l'objet estampillé rendant l'empreinte illisible rendra nécessairement l'objet inutilisable,
   - estampillage *probabiliste* : selon la longueur (ou proportion) de l'empreinte, le distributeur pourra identifier un ou plusieurs adversaires (avec un certain niveau de confiance). Cette identification est cependant incertaine,
   - estampillage *déterministe* : le distributeur pourra identifier le ou les adversaires avec une probabilité de 100% ;

iii) **Troisième division** :
   - estampillage par *reconnaissance* : l'empreinte d'un objet estampillé fait déjà partie du contenu de cet objet,
   - estampillage par *suppression* : l'estampillage d'un objet se fait par l'omission d'une partie de cet objet,
   - estampillage par *addition* : l'estampillage se fait par en ajoutant de nouvelles portions à l'objet intéressé,
   - estampillage par *modification* : une partie de l'objet est modifiée pour permettre le traçage ;

iv) **Quatrième division** :

- estampillage *discret* : les empreintes possibles que l'on peut attacher à un contenu sont limitées en nombre (estampillage binaire ou q-aire pour $q > 2$),
- estampillage *continu* : le nombre d'empreintes possibles que l'on peut attacher à un contenu n'est pas fini (exemple : pourcentage de liquide ajouté dans une solution).

Dans ce manuscrit de thèse, nous nous intéressons à un estampillage numérique probabiliste discret par modification. Il est alors naturel pour un distributeur de vouloir cacher aux utilisateurs la localisation des données de l'empreinte dans son œuvre afin qu'un adversaire ne puisse les effacer ou les modifier. De plus, les données de l'empreinte ne doivent pas dénaturer le contenu original (un utilisateur doit pouvoir utiliser le contenu quel que soient les données liées à l'empreinte insérée).

Dans la plupart des systèmes d'estampillage probabilistes actuels, afin de pouvoir accuser efficacement au moins un des membres de la coalition (et ne pas accuser un utilisateur innocent de redistribution illégale de contenu), la séquence (empreinte) pirate doit respecter la **condition de marquage** :

Définition 4 *: **Condition de marquage** [15]*
Lors d'une attaque de coalition, les adversaires mettent en commun leurs contenus et tentent de localiser et de modifier l'empreinte. Ils modifient alors les données numériques de leurs contenus UNIQUEMENT aux positions où ils observent une différence : nous émettons l'hypothèse que les adversaires ne changeront pas la donnée de leurs contenus à une position i donnée lorsqu'ils ont tous la même donnée à cette position précise.

Cette condition est très restrictive dans les applications de l'estampillage : l'ajout de bruit gaussien ou autre attaque de robustesse sur le contenu pirate forgé produit une empreinte ne respectant a priori pas cette condition.

Les méthodes d'estampillage résistantes aux attaques de coalition d'adversaires ont été étudiées par Blackley *et. al.* [13]. Les codes IPP (*Identifiable Parent Property codes*) ont été introduits par Chor *et. al.* [22]. Ces codes déterministes sont efficaces uniquement lorsque le contenu pirate généré par la coalition contient, pour toute position i de cette empreinte, la i-ème position de l'empreinte d'un membre de la coalition. Notons que cette dernière condition est équivalente à la condition de marquage dans le cas d'un estampillage discret binaire (en effet, si les empreintes diffèrent à une position, c'est qu'il existe au moins un symbole "1" et un symbole "0" chez la coalition, l'empreinte modifiée contiendra alors forcément le symbole d'un des membres). Sous cette dernière condition, ces codes permettent d'accuser un adversaire de façon efficace.

Dans [79], Gabor Tardos émet la proposition suivante : sous l'hypothèse de la condition de marquage, il n'existe pas de système d'estampillage déterministe pour trois utilisateurs si deux d'entre eux forment une coalition (il n'existe pas d'algorithme déterministe permettant d'accuser un des adversaires sans risquer d'accuser l'utilisateur innocent).

Les techniques d'estampillage probabilistes ont été introduites par Boneh et Shaw [15], ils ont construit ces codes d'estampillage de longueur $m = O(c^4 \log(1/p_{\text{fa}}) \log(n/p_{\text{fa}}))$ avec c la taille maximale d'une coalition, n le nombre d'utilisateurs, p_{fa} la probabilité d'accuser un innocent (probabilité de fausse alarme). Dans ce manuscrit de thèse, nous nous intéressons particulièrement aux codes générés par Gabor Tardos [79] car ils ont la particularité d'atteindre la borne théorique inférieure de Peikert [68] : $m = O(c^2 \log(n/p_{\text{fa}}))$ pour les systèmes d'estampillage probabilistes.

4.1 Les codes probabilistes de Tardos

Cette section décrit la construction et le procédé d'accusation des codes discrets binaires probabilistes introduits par Tardos [79] pour n utilisateurs.

4.1.1 Construction

Nous considérons une matrice d'estampillage $\mathbf{M} \in \mathcal{M}_{n,m}(\mathbb{F}_2)$. Chaque ligne $\mathbf{m}_j$ de la matrice $\mathbf{X}$ est une séquence de m bits qui identifie l'utilisateur $j \in [n]$. Les colonnes de $\mathbf{M}$ (les i-èmes bits des utilisateurs) sont générés selon une loi de Bernoulli :

$$\forall j \in [n],\ \forall i \in [m],\ \mathbf{M}(j,i) \sim \mathcal{B}(p_i). \tag{4.1}$$

Les $\{p_i\}_{i\in[m]}$ sont distribués dans l'ensemble $[t, 1-t]$ avec $t = (300c)^{-1}$, selon une variable aléatoire P de densité de probabilité $\mathrm{f}_P(p)$:

$$\mathrm{f}_P(p) = \frac{1}{\pi\sqrt{p(1-p)}}. \tag{4.2}$$

La figure 4.1 montre l'histogramme des valeurs de p_i ainsi que la densité de probabilité théorique donnée par l'équation (4.2). Les p_i peuvent être générés de la manière suivante : nous générons m valeurs r_i uniformément dans l'intervalle $[t', 1-t']$ avec $t' = \sin^{-1}(\sqrt{t})$, nous construisons alors, pour tout $i \in [m]$, $p_i = \sin^2(r_i)$.

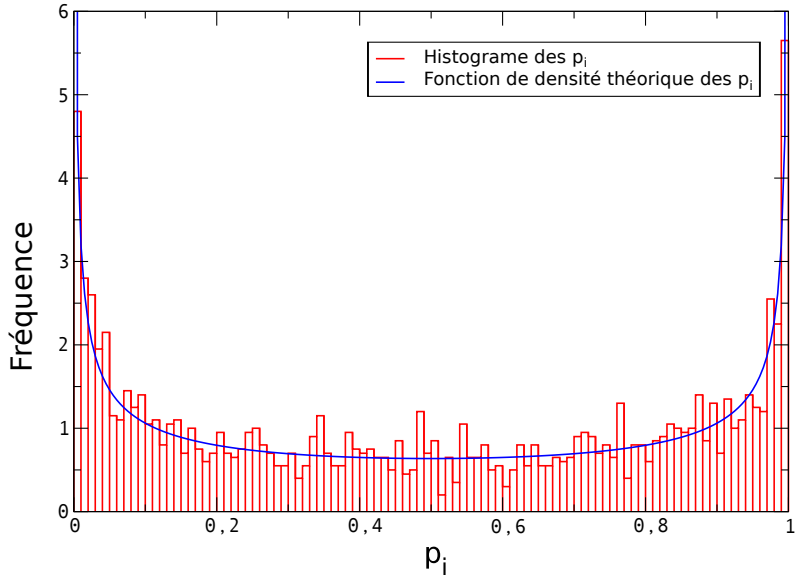


FIGURE 4.1 – Histogramme et densité de probabilité théorique des p_i, $c = 10, m = 10000$.

Dans le contexte des attaques de coalition, une coalition de c adversaires $\mathcal{C} = \{j_0 \dots j_{c-1}\} \subset [n]$ forge une séquence pirate $\mathbf{m}$ de m bits en combinant les bits de leurs séquences respectives (selon une stratégie prédéfinie, voir section 4.2.2 pour plus de détails). La séquence pirate $\mathbf{m}$ est obtenue de manière formelle par :

$$\mathbf{m} = \left(\mathbf{m}_{j'_0}(0) \dots \mathbf{m}_{j'_{m-1}}(m-1)\right), \tag{4.3}$$

avec :

$$(j'_0 \dots j'_{m-1}) \in \mathcal{C}^m, \tag{4.4}$$

et sont choisis d'après la stratégie d'attaque choisie par les adversaires. Cette propriété respecte la condition de marquage : si les adversaires ont tous le même symbole à une position $i \in [m]$ de leurs empreintes, alors la séquence pirate aura le même symbole.

4.1.2 Procédé d'accusation

Le but d'un distributeur est de pouvoir accuser au moins un des membres de la coalition avec une probabilité d'erreur p_{e} et une probabilité de fausse alarme p_{fa}. Pour identifier au moins un des adversaires, nous utilisons la fonction d'accusation proposée par G. Tardos et améliorée par Skoric *et. al.* [77] (cette amélioration consiste à donner aussi un poids d'accusation aux positions i des empreintes lorsque que le symbole extrait de la séquence pirate est "0", l'accusation classique de Tardos portait uniquement sur les symboles "1").

Cette accusation est basée sur la construction d'une matrice $\mathbf{U} \in \mathcal{M}_{n,m}(\mathbb{R})$:

$$\mathbf{U}(j,i) = \begin{cases} \mathrm{g}_1(p_i), & \text{si } \mathbf{m}(i) = 1, \mathbf{m}_j(i) = 1, \\ \mathrm{g}_0(p_i), & \text{si } \mathbf{m}(i) = 1, \mathbf{m}_j(i) = 0, \\ \mathrm{g}_0(1-p_i), & \text{si } \mathbf{m}(i) = 0, \mathbf{m}_j(i) = 1, \\ \mathrm{g}_1(1-p_i), & \text{si } \mathbf{m}(i) = 0, \mathbf{m}_j(i) = 0, \end{cases} \tag{4.5}$$

avec :

$$\mathrm{g}_1(p) = \sqrt{\frac{1-p}{p}},\ \mathrm{g}_0(p) = -\sqrt{\frac{p}{1-p}}. \tag{4.6}$$

Le score d'un utilisateur $j \in [n]$ est défini par S_j :

$$\mathrm{S}_j = \sum_{i=0}^{m-1} \mathbf{U}(j,i). \tag{4.7}$$

Un utilisateur $j \in [n]$ est accusé d'avoir participé à la création d'une séquence pirate $\mathbf{m}$ si $\mathrm{S}_j > \tau$ où τ est un seuil spécifique (Tardos utilise $\tau = 20c\lceil 1/p_{\mathrm{fa}} \rceil$). Une autre stratégie consiste à toujours accuser l'utilisateur ayant le plus gros score d'accusation.

Notons que les fonctions $\mathrm{g}_0,\ \mathrm{g}_1$ et $\mathrm{f}_P(p)$ sont celles qui maximisent le score d'accusation S_j en espérance lorsque celui-ci est voulu indépendant de la stratégie d'attaque (équations (4.3) et (4.4)). L'optimalité de ces fonctions données par Tardos est démontrée dans [34].

La figure 4.2 montre les valeurs des accusations d'un utilisateur j sur le i-ème bit conformément aux quatre cas présentés dans l'équation (4.5).

Quelques remarques sur la génération des codes et le processus d'accusation :

1. La p.d.f. des p_i, donnée par l'équation (4.2), permet de maximiser l'information mutuelle pour la pire des attaques de la coalition lorsque c tend vers l'infini [35, 39] (cette attaque correspond alors à un mixage des codes binaires de façon aléatoire 4.12) ;

2. Le choix de la distribution des valeurs de p_i est biaisé en direction des valeurs proches de 0 ou de 1. Ce choix est motivé par la condition de marquage, il permet aux adversaires d'avoir le moins d'information possible sur les localisations des empreintes (ils ne peuvent modifier les séquences qu'aux endroits où leurs symboles diffèrent) ;

3. Les bornes t et $1-t$ choisies pour la distribution permettent d'obtenir des valeurs de p_i qui ne sont pas trop proches des bornes 0 et 1 afin que les accusations définies dans l'équation (4.5) ne tendent pas rapidement vers l'infini (ce qui donnerait beaucoup plus d'influence à une position i donnée sur le score S_j d'un utilisateur $j \in [n]$).

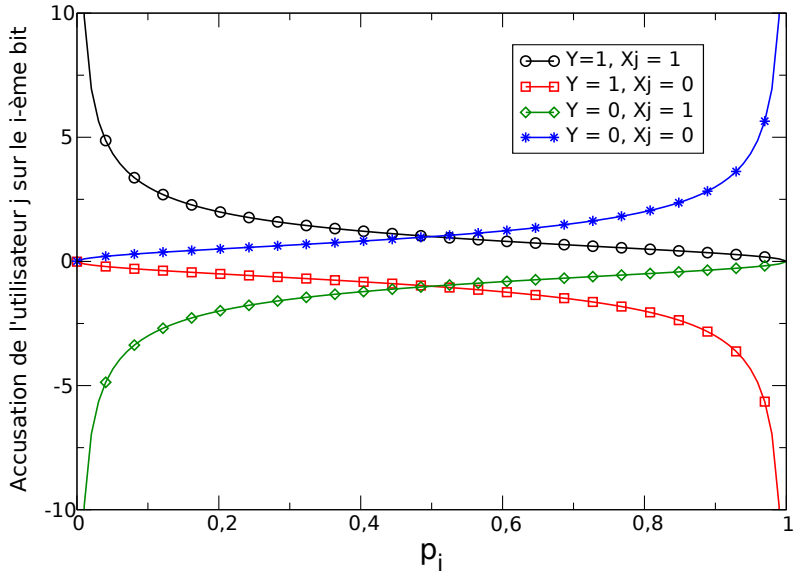


FIGURE 4.2 – Accusation d'un utilisateur j sur le i-ème bit d'une séquence d'estampillage $\mathbf{m}$.

4.2 Débits atteignables et stratégies d'attaque

4.2.1 Débits atteignables d'un schéma d'estampillage

Nous avons vu précédemment le cas des attaques par coalition, plusieurs adversaires (utilisateurs de contenus ou personnes ayant corrompu un ou plusieurs utilisateurs) mixent leurs séquences d'estampillage pour forger une nouvelle séquence (pirate) afin de pouvoir diffuser illégalement leur contenu sur des réseaux d'échange en limitant leur risque d'être accusés par le distributeur. Les codes de Tardos sont résistants aux attaques par coalition, c'est-à-dire qu'ils permettent d'accuser au moins un membre de la coalition avec une probabilité d'erreur p_e et une probabilité d'accuser un innocent p_{fa} lorsque la longueur des codes atteint la borne de Peikert $m = O(c^2 \log(n/p_{\text{fa}}))$. Toujours d'après [79], nous avons la relation suivante :

$$p_{\text{e}} = p_{\text{fa}}^{c/4}. \tag{4.8}$$

Les fonctions choisies pour Tardos (p.d.f. des $\{p_i\}_{i\in[m]}$ et fonctions d'accusation g_0 et g_1) sont optimales lorsque la moyenne et la variance des scores d'accusation sont voulues indépendantes de la manière dont est créée la séquence pirate forgée par c adversaires. La construction d'une nouvelle séquence est appelée stratégie d'attaque et dépend de la manière dont les symboles sont choisis parmi les membres de la coalition position par position, c'est-à-dire le choix du vecteur $\left(j'_0 \dots j'_{m-1}\right) \in \mathcal{C}^m$ dans l'équation (4.3). Dans [21], les auteurs proposent une manière d'estimer la stratégie d'attaque afin d'améliorer les performances d'accusation.

Notons que pour l'instant nous ne nous intéressons pas à la manière dont sont cachées les séquences de traçage dans les contenus, le tatouage des séquences (avec les contraintes de sécurité et de robustesse qui en découlent) sera abordé dans la section 4.3 suivante. Nous considérons pour le moment que le distributeur est capable de décoder parfaitement les séquences des utilisateurs ainsi que la séquence pirate forgée par la coalition (celle-ci respectant bien évidemment la condition de marquage).

Dans [34], les auteurs précisent le fonctionnement de la construction des scores d'accusation données par l'équation (4.5) : ceux-ci sont construits comme la somme de variables aléatoires indépendantes et identiquement distribuées et, d'après le théorème central limite, tendent vers une distribution gaussienne lorsque $m \to \infty$, in-

dépendante (par construction) de la stratégie d'attaque choisie par la coalition (vue alors comme un paramètre de nuisance, ne modifiant pas la variance et la moyenne des scores d'accusation). Cependant, l'utilisation du théorème central limite est à manier avec précaution car sa validité repose sur le nombre m de symboles de la séquence d'estampillage. Si celui-ci est faible, la distribution des scores n'est pas exactement gaussienne et l'influence du choix de la stratégie est alors visible, les débit atteignables permettent alors de mesurer cette influence. Pour la suite de ce manuscrit de thèse, nous notons M (respectivement M_j) la variable aléatoire représentant le symbole binaire à une position dans la séquence pirate (resp. d'un adversaire j), P la variable aléatoire de p.d.f. $\mathrm{f}_P(p)$ (équation (4.2)).

Dans [65] [39], les auteurs définissent deux types de décodeurs permettant de mesurer la capacité d'un schéma d'estampillage :

i) **décodeur simple** : le débit atteignable pour un décodeur simple R_s est défini comme l'information mutuelle entre la séquence pirate forgée par une coalition $\mathcal{C}$ et la séquence d'un utilisateur j_0 (en espérance sur les $\{p_i\}$) :

$$\mathrm{R}_s = \mathbb{E}_P[I(M; M_{j_0})|P = p], \tag{4.9}$$

ii) **décodeur joint** : le débit atteignable pour un décodeur joint R_j est défini comme l'information mutuelle entre la séquence pirate forgée par une coalition $\mathcal{C}$ et les séquences d'une coalition $\mathcal{C}'$ de taille c' (en espérance sur les $\{p_i\}$) :

$$\mathrm{R}_j = \frac{1}{c'}\mathbb{E}_P[I(M; \{M_j\}_{j\in\mathcal{C}'})|P = p]. \tag{4.10}$$

Les fonctions d'accusation données par Tardos correspondent à la classe des décodeurs simples : le score est calculé pour chaque utilisateur. C'est à cette classe que nous nous intéressons particulièrement dans cette seconde partie du manuscrit de thèse. Le calcul du débit atteignable nous permet de mesurer l'efficacité d'une attaque de coalition sur les codes de Tardos. En effet, le but d'une coalition sera de trouver une stratégie d'attaque permettant de diminuer la valeur de ce débit. La partie suivante formalise mathématiquement la notion de stratégie d'attaque et présente des exemples d'attaque en fonction de la connaissance qu'a une coalition sur les symboles de leurs séquences d'estampillage.

4.2.2 Stratégies d'attaque et attaque au pire cas

Une stratégie d'attaque définit le procédé utilisé par une coalition $\mathcal{C}$ pour générer une séquence d'estampillage $\mathbf{m}$ respectant la condition de marquage, elle consiste à sélectionner, pour chaque position $i \in [m]$, le symbole d'un adversaire "candidat", voir équations (4.3) et (4.4). Pour chaque position i, la valeur de $\mathbf{m}(i)$ dépend du nombre de symboles 1 (ou 0 par symétrie) que la coalition possède à cette position. Une stratégie d'attaque est entièrement définie par un vecteur $\theta = (\theta(0)\dots\theta(c))$ (notation de [36]). Si M (respectivement M_j) est la variable aléatoire représentant le symbole binaire à une position dans la séquence pirate (resp. d'un adversaire j), nous avons :

$$\theta(k) = \Pr\left(M = 1 \left| \sum_{j\in\mathcal{C}} M_j = k \right.\right). \tag{4.11}$$

Nous émettons l'hypothèse que la coalition adopte la même stratégie pour chaque bit de la séquence pirate. Pour toute stratégie, nous avons $\theta(0) = 0$ et $\theta(c) = 1$ afin de respecter la condition de marquage.

Les stratégies de coalition dépendent du degré d'information que les adversaires possèdent sur les symboles de leurs séquences. Nous nous intéressons ici à deux catégories de coalition : la coalition aveugle et la coalition voyante [34] :

$\#\mathcal{C}$	θ_{WCA}
$c = 2$	$(0.\ 0.5\ 1)$
$c = 3$	$(0.\ 0.651\ 0.349\ 1.)$
$c = 4$	$(0.\ 0.487\ 0.5\ 0.513\ 1.)$
$c = 5$	$(0\ 0.594\ 0.\ 1.\ 0.406\ 1.)$

TABLE 4.1 – Valeurs numériques des attaques θ_{WCA} pour des coalitions de taille $\#\mathcal{C} = 2, 3, 4, 5$.

i) **Coalition aveugle** : les adversaires n'ont aucune idée des valeurs de leurs symboles, pour chaque position $i \in [m]$, ils choisissent de façon aléatoire le symbole d'un des membres de la coalition, nous avons :

$$\forall k \in [c+1],\ \theta_{\mathrm{aléatoire}}(k) = k/c. \tag{4.12}$$

Nous remarquons que la condition de marquage est respectée : $\theta(0) = 0$ et $\theta(c) = 1$.

ii) **Coalition voyante** : les adversaires sont capables de comparer leurs symboles pour chaque position i et de déterminer si leurs symboles sont égaux ou non sans savoir la nature de chaque symbole (les adversaires se classent en deux sous-groupes, ceux ayant le symbole 0 et ceux ayant le symbole 1 sans connaître la nature du symbole de chaque sous-groupe). La coalition peut alors procéder à une stratégie de vote : vote minoritaire (respectivement vote majoritaire) où le symbole est choisi de façon aléatoire dans le sous-groupe ayant la taille la plus petite (resp. la plus grande) :

$$\forall k \in [c+1],\ \theta_{\mathrm{Minoritaire}}(k) = \begin{cases} 0 \text{ si } k = 0, \\ 1 \text{ si } k \in]\!]0, c/2[\!], \\ k/c \text{ si } k = c/2, \\ 0 \text{ si } k \in]\!]c/2, c[\!], \\ 1 \text{ si } k = c, \end{cases} \tag{4.13}$$

$$\forall k \in [c+1],\ \theta_{\mathrm{Majoritaire}}(k) = \begin{cases} 0 \text{ si } k \in [\![0, c/2[\!], \\ k/c \text{ si } k = c/2, \\ 1 \text{ si } k \in]\!]c/2, c]\!]. \end{cases} \tag{4.14}$$

Notons qu'un coalition voyante est capable des mêmes stratégies d'attaque qu'une coalition aveugle. Dans [36], les auteurs proposent une stratégie d'attaque dans le cadre d'une coalition voyante appelée attaque au pire cas : WCA (*Worst Case Attack*). Cette attaque est calculée en minimisant le débit atteignable d'un schéma d'estampillage pour décodeur simple R_s :

$$\theta_{\mathrm{WCA}} = \arg\min_{\theta} \{\mathrm{R}_s(\theta)\}. \tag{4.15}$$

Cette attaque permet donc de minimiser l'information mutuelle entre la séquence binaire d'un membre de la coalition et la séquence pirate forgée par cette même coalition. La table 4.1 donne quelques exemples de θ_{WCA} pour $c = 2, 3, 4, 5$ adversaires (repris de [36]).

Pour un seuil τ d'accusation fixé, l'attaque WCA permet alors d'augmenter la probabilité d'accuser un innocent : la probabilité de fausse alarme p_{fa} définie par :

$$p_{\mathrm{fa}} = \Pr(\mathrm{S}_j > \tau),\ j \in [n]\backslash\mathcal{C}, \tag{4.16}$$

ainsi que la probabilité d'accuser un des membres de la coalition : la probabilité d'erreur p_{e} :

$$p_{\mathrm{e}} = \Pr(\mathrm{S}_j < \tau),\ j \in \mathcal{C}. \tag{4.17}$$

La figure 4.3 illustre les p.d.f. des scores S_j des innocents et adversaires avec les probabilités p_{fa} et p_{e} associées.

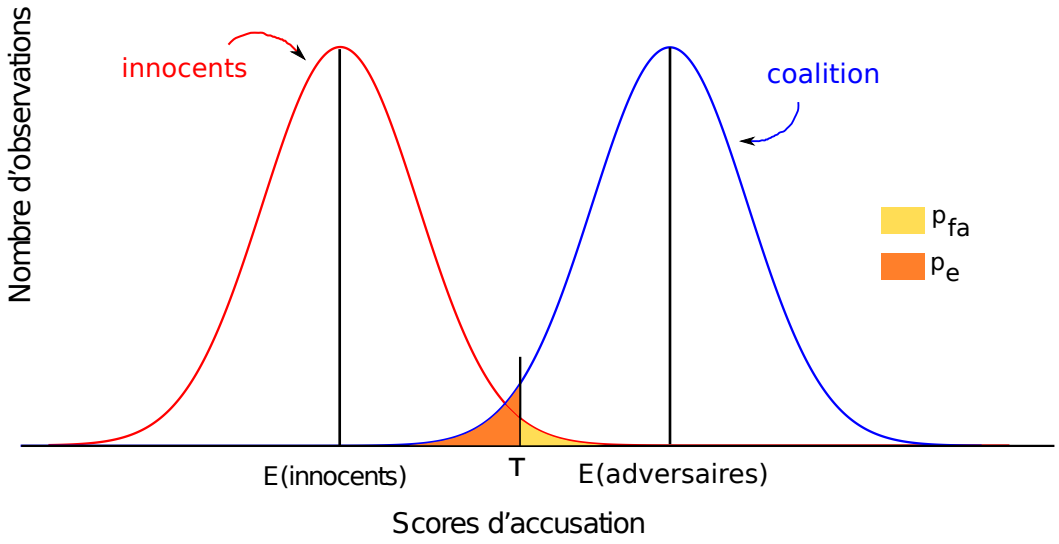


FIGURE 4.3 – Fonctions de densité des scores S_j des innocents et des adversaires et probabilités p_{fa} et p_{e} associées.

La WCA permet donc, pour une coalition d'adversaire, de forger la séquence pirate qui permettra de diminuer leur score d'accusation. Cependant, jusqu'à présent, nous nous placions dans un cadre théorique d'estampillage : pour utiliser la WCA, les adversaires doivent former une coalition au moins voyante. De plus, nous considérons le distributeur capable de décoder la séquence pirate sans erreur.

Le tatouage numérique est une solution permettant d'insérer chaque code d'estampillage dans un contenu. Cependant, nous avons vu dans la première partie de ce manuscrit que les contraintes liées au tatouage influencent d'une part la stratégie d'attaque de la coalition (le décodage des symboles binaire dépend de la **sécurité** du schéma de tatouage) et d'autre part l'accusation par le distributeur (la qualité de décodage de la séquence pirate dépend de la robustesse de l'insertion). Dans la prochaine section, nous prenons en compte ces deux contraintes afin de construire une attaque au pire cas dépendante cette fois-ci de l'estimation des symboles par la coalition et, par le biais des débit atteignables, nous répondons à la question suivante :

Dans le cadre de l'estampillage, un distributeur doit-il privilégier un tatouage très robuste et non-sûr ou au contraire un tatouage sûr mais moins robuste ?

4.3 Estampillage et tatouage sûr

4.3.1 Motivations

Nous avons vu dans le chapitre précédent que les récents travaux portant sur les codes de Gabor Tardos [36] proposent, pour une coalition, une stratégie d'attaque qui minimise l'information mutuelle entre une séquence d'estampillage pirate (produite par la coalition) et la séquence initiale d'un adversaire. Cette attaque, appelée attaque au pire cas (*Worst Case Attack* WCA) permet de maximiser non seulement la probabilité d'accuser un innocent (probabilité de fausse alarme : p_{fa}) mais aussi la probabilité qu'un adversaire ne soit pas accusé (probabilité d'erreur : p_{e}).

Il parait important de noter qu'un code d'estampillage (tel le code de Tardos [79]) utilisé à des fins de traçage d'un contenu multimédia doit être caché en utilisant des techniques de tatouage. La transparence des techniques d'insertion par tatouage numérique garantie que la qualité de l'œuvre numérique n'est pas dégradée par l'ajout d'une empreinte. La robustesse propre aux techniques de tatouage permet d'extraire l'empreinte d'estampillage quand le contenu multimédia subit des compressions (modification du débit de réencodage d'une vidéo par exemple) ou tout autre ajout de bruit. Des exemples d'implantations pratiques d'estampillage par tatouage numérique ont déjà été proposées dans le domaine des communications numériques [80, 83, 84]. De manière générale, l'estampillage de contenus numériques se divisent en deux étapes : le **codage** et l'**insertion**, la figure 4.4 illustre ces étapes.

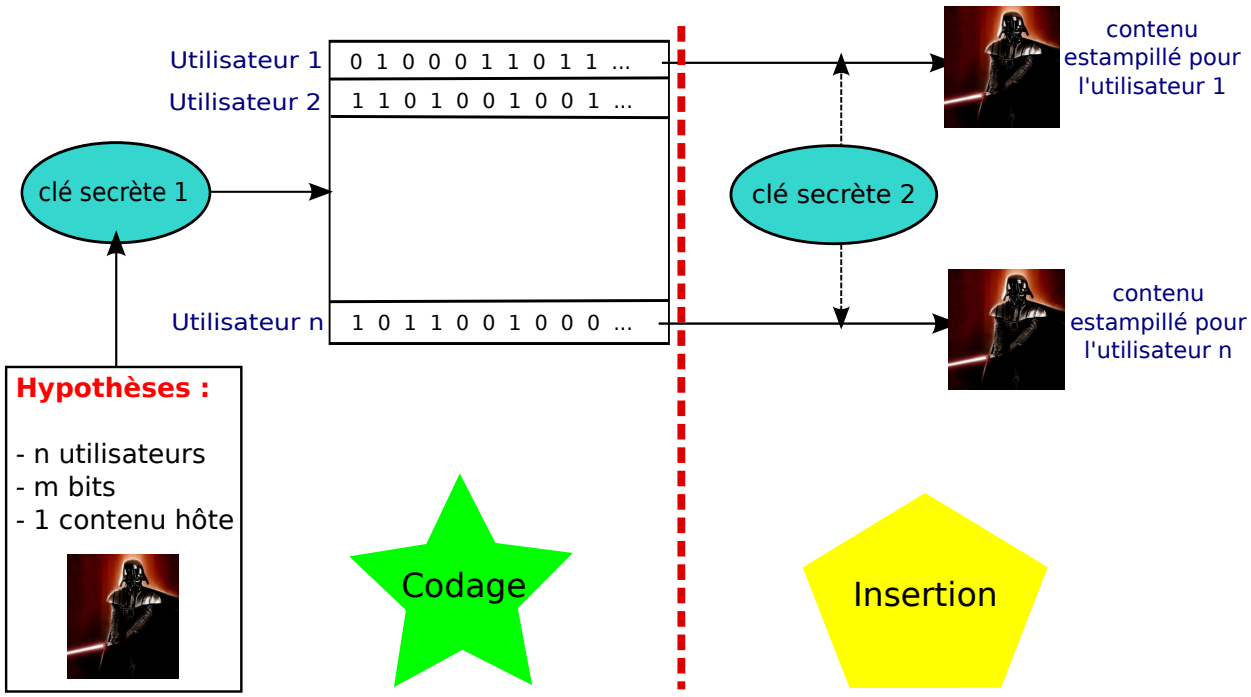


FIGURE 4.4 – Estampillage d'un contenu numérique. **Étape 1 (codage)** : le distributeur génère des codes (ici binaires) de longueur m pour n utilisateurs à l'aide d'une première clé secrète (par exemple le vecteur des $\{p_i\}_{i\in[m]}$ pour les codes de Tardos). **Étape 2 (insertion)** : chaque code d'estampillage est tatoué dans le contenu avant distribution aux utilisateurs à l'aide d'une deuxième clé secrète (par exemple les porteuses $\{\mathbf{u}_i\}_{i\in[N_c]}$ dans le cadre de l'étalement de spectre).

Un des aspects qui n'a pas encore été étudié dans le lien entre les techniques de tatouage et d'estampillage

est l'impact possible de la sécurité des schémas de tatouage [44] sur les stratégies d'attaque disponibles par une coalition d'adversaires. Les codes d'estampillage de Tardos ont été étudiés sous l'hypothèse que la coalition soit capable de connaître exactement les positions de leurs empreintes où les symboles insérés diffèrent. Néanmoins, l'utilisation de techniques de tatouage sûres comme celles proposées dans [20, 69] permettent de leurrer les adversaires en provoquant des erreurs d'estimation des symboles de leurs empreintes. Notons aussi que la sécurité d'un schéma de tatouage peut être évaluée en procédant à une estimation de la clé secrète de tatouage en utilisant des techniques d'apprentissage comme l'ACI [10, 20, 40, 41], l'ACP [11] (voir section 1.6) ou le groupement (*clustering*) [8].

Nous proposons dans ce chapitre de nouvelles stratégies d'attaque au pire cas sur les codes de Tardos destinés aux membres d'une coalition. Celles-ci dépendent alors d'une nouvelle contrainte : le niveau de sécurité du schéma de tatouage utilisé. Nous considérons une erreur d'estimation ϵ des empreintes (séquences) décodées par les adversaires et nous proposons alors de nouvelles attaques sous-jacentes.

Nous comparons ces attaques avec l'attaque au pire cas classique présentée dans la section 4.2.2 en calculant la probabilité de fausse alarme et l'information mutuelle entre la séquence forgée par une coalition et celle d'un de ses membres. De plus, nous simulons des attaques de robustesse, selon un taux d'erreur binaire η et nous quantifions, pour un même débit atteignable d'un schéma d'estampillage, le compromis existant entre robustesse et sécurité en comparant des schémas de tatouage offrant différents degrés de sécurité avec des schémas non-sûrs de manière théorique et pratique (méthode par étalement de spectre).

La figure 4.5 schématise les différentes étapes de la chaine d'estampillage : de la génération des codes par le distributeur à l'accusation des membres de la coalition.

4.3.2 Attaques au pire cas pour des schémas de tatouage sûrs

4.3.2.1 Erreurs d'estimation

La sécurité d'un schéma de tatouage peut être exprimée mathématiquement via l'estimation des symboles insérés par la coalition pour chaque position $i \in [m]$ avec une erreur ϵ. Plus le schéma est sûr, plus ϵ est proche de 0.5. Dans le cadre d'un schéma de tatouage non-sûr, nous avons alors $\epsilon = 0$ et en reprenant la terminologie des classes de coalition [34] (section 4.2.2), nous nous trouvons dans le cas d'une **coalition voyante**.

Au contraire, avec un schéma de tatouage complètement sûr ($\epsilon = 0.5$), les adversaires ne seront pas capables de déterminer si leurs symboles sont 0 ou 1 et la seule stratégie d'attaque qui leur est autorisée est celle qui consiste à choisir aléatoirement les symboles parmi les membres de la coalition pour chaque position $i \in [m]$, la coalition est alors une **coalition aveugle**.

Considérant cette nouvelle hypothèse liée à la sécurité du schéma de tatouage et appliquée à l'estampillage discret binaire, la coalition $\mathcal{C}$ sera capable de décoder de façon exacte l'ensemble des empreintes $\{\mathbf{m}_j\}_{j\in\mathcal{C}}$ uniquement si $\epsilon = 0$. De plus, chaque membre ne pourra savoir si son symbole est le même que celui d'un autre membre à une position $i \in [m]$. Notons que nous prenons comme hypothèse de travail un schéma de tatouage respectant le principe de Kerckhoffs, nous pouvons alors émettre l'hypothèse que chaque membre de la coalition connaît la sécurité du schéma de tatouage utilisé et, par conséquent, connaît l'erreur ϵ résultant de leur technique d'estimation.

Nous notons $\hat{\mathbf{m}}_j(i)$ le symbole décodé par l'adversaire $j \in \mathcal{C}$ à la position $i \in [m]$ et $\hat{M}_j$ la variable aléatoire associée avec la propriété suivante :

$$\Pr\left(\hat{M}_j = 1 | M_j = 0\right) = \Pr\left(\hat{M}_j = 0 | M_j = 1\right) = \epsilon. \tag{4.18}$$

La coalition forge alors une séquence pirate $\hat{\mathbf{m}} \in \mathbb{F}_2^m$ dans le domaine estimé en utilisant une stratégie θ définie

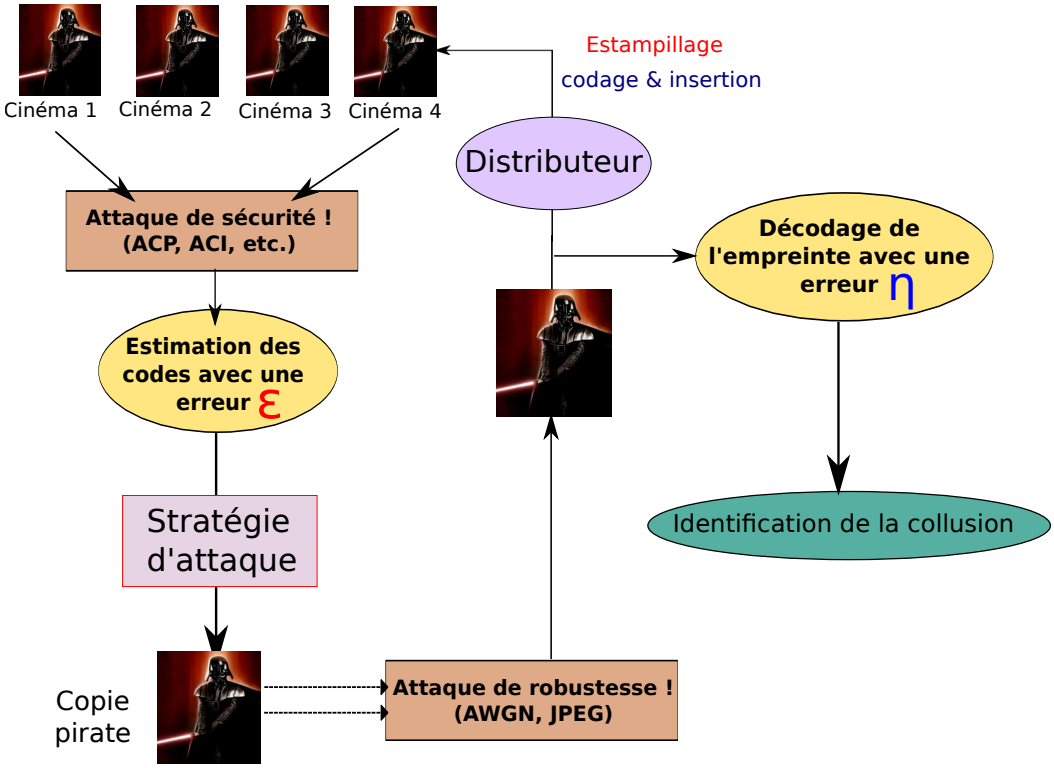


FIGURE 4.5 – Chaine complète d'estampillage. Le distributeur génère et insère les codes de traçage pour différents utilisateurs. Une coalition, après avoir effectué une attaque de sécurité pour estimer les symboles insérés, forge un nouveau contenu selon une stratégie d'attaque optimale dépendant de l'erreur d'estimation. La copie pirate, probablement sujette à des attaques de robustesse, est ensuite analysée par le distributeur pour cerner au moins un des membres de la coalition responsable de la diffusion illégale du contenu.

par :

$$\theta(k) = \Pr\left(\hat{M} = 1 \middle| \sum_{j \in \mathcal{C}} \hat{M}_j = k\right), \tag{4.19}$$

où $\hat{M}$ est la variable aléatoire associée au symbole de la séquence $\hat{\mathbf{m}}$ forgée par la coalition estimée à une position $i \in [m]$. La séquence pirate $\mathbf{m}$ est alors construite selon :

$$\forall i \in [m],\ \mathbf{m}(i) = \mathbf{m}_{j'}(i), \tag{4.20}$$

avec j' choisi de manière uniforme dans l'ensemble $\{j \in \mathcal{C} : \hat{\mathbf{m}}_j(i) = \hat{\mathbf{m}}(i)\}$. La figure 4.6 illustre le processus d'attaque pour $c = 5$ adversaires.

Nous remarquons que l'utilisation d'un tatouage sûr respecte la condition de marquage : lorsque la coalition possède le même symbole M_j à une position $i \in [m]$, le résultat de la stratégie d'attaque donnera exactement le même symbole $M = M_j$ de par l'équation (4.20).

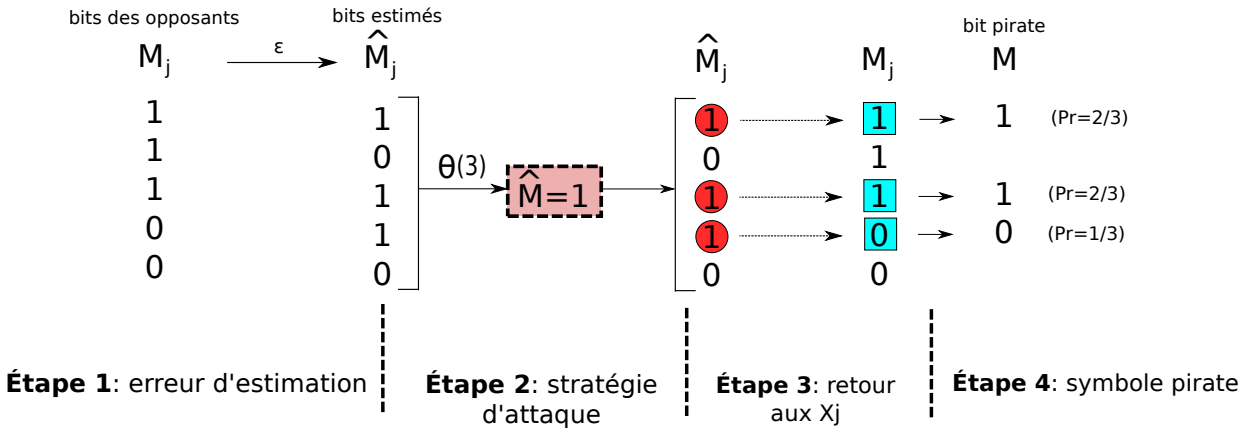


FIGURE 4.6 – Processus d'attaque pour un schéma de tatouage sûr avec $c = 5$ adversaires et $\theta(3) = 1$. **Étape 1** : les adversaires estiment trois symboles 1 ($\hat{M}_j$). **Étape 2** : nous avons $\theta(3) = 1$, la stratégie d'attaque donne $\hat{M} = 1$. **Étape 3** : la coalition cherche les M_j qui correspondent à $\hat{M}_j = \hat{M} = 1$. **Étape 4** : le symbole de la séquence pirate M est choisi uniformément parmi les M_j sélectionnés.

4.3.2.2 Débits atteignables prenant en compte la contrainte de sécurité

Nous redéfinissons le débit atteignable d'un schéma d'estampillage R_s initialement donné par l'équation (4.9) (celui-ci dépend dorénavant de ϵ) :

$$\begin{aligned}\mathrm{R}_s(\theta,\epsilon) &= \mathbb{E}_P[I(M;M_{j_0})|P=p] \\ &= \mathbb{E}_P[H(M) - H(M|M_{j_0})|P=p] \\ &= \mathbb{E}_P[H(M) - (pH(Y|M_{j_0}=1) + (1-p)H(M|M_{j_0}=0))|P=p], \\ &= \mathbb{E}_P[H_b(p_1) - (pH_b(p_2) + (1-p)H_b(p_3)|P=p].\end{aligned} \tag{4.21}$$

où $H(.)$ et $H_b(.)$ représentent respectivement l'entropie et l'entropie binaire, les probabilités p_1, p_2 et p_3 sont données par :

$$p_1 = \Pr(M=1), \tag{4.22}$$

$$p_2 = \Pr(M=1|M_{j_0}=1), \tag{4.23}$$

$$p_3 = \Pr(M=1|M_{j_0}=0), \tag{4.24}$$

avec : M le bit de l'empreinte pirate et M_{j_0} le bit d'un adversaire j_0.

Nous calculons maintenant les expressions analytiques de p_1, p_2 et p_3.

Calcul de p_1 La définition de la probabilité conditionnelle nous donne :

$$p_1 = \sum_{l=0}^{c}\sum_{k=0}^{c}\left(\Pr\left(\Sigma_M = l, \Sigma_{\hat{M}} = k\right) \times \Pr\left(M=1|\Sigma_M = l, \Sigma_{\hat{M}} = k\right)\right), \tag{4.25}$$

avec $\hat{M}$ le bit estimé résultant de la stratégie d'attaque, $\sum_{j\in\mathcal{C}} M_j = \Sigma_M$ et $\sum_{j\in\mathcal{C}} \hat{M}_j = \Sigma_{\hat{M}}$,

Nous introduisons la variable aléatoire V, qui correspond au nombre de $M_j = 1$ qui ont été estimés par la coalition comme $\hat{M}_j = 0$, $V = \#\{j \in \mathcal{C} : M_j = 1, \hat{M}_j = 0\}$. Pour $l, k \in [c+1]$; V prend ses valeurs dans l'ensemble $\Omega = \{i \in \mathbb{N} : i \leq l; i \leq c-k; i \geq l-k\}$. Toujours grâce aux probabilités conditionnelles, nous obtenons :

$$p_1 = \sum_{l=0}^{c} \Big(\Pr(\Sigma_M = l) \sum_{k=0}^{c} \Big(\Pr(\Sigma_{\hat{M}} = k | \Sigma_M = l) \times \sum_{i \in \Omega} \big(\Pr(M = 1 | V = i, \Sigma_M = l, \Sigma_{\hat{M}} = k) \times \Pr(V = i | \Sigma_M = l, \Sigma_{\hat{M}} = k) \big) \Big) \Big), \quad (4.26)$$

où les quatre probabilités concernées dans l'équation précédente sont calculées en utilisant l'analyse combinatoire et la définition des probabilités conditionnelles :

$$\Pr(\Sigma_M = l) = \binom{c}{l} p^l (1-p)^{c-l}, \quad (4.27)$$

$$\Pr(\Sigma_{\hat{M}} = k | \Sigma_M = l) = \sum_{i \in \Omega} \binom{l}{i} \binom{c-l}{k-l+i} \epsilon^i (1-\epsilon)^{l-i} \epsilon^{k-l+i} (1-\epsilon)^{c-k-i}, \quad (4.28)$$

$$\Pr(M = 1 | V = i, \Sigma_M = l, \Sigma_{\hat{M}} = k) = \theta(k) \frac{l-i}{k} + (1 - \theta(k)) \frac{i}{c-k}, \quad (4.29)$$

$$\Pr(V = i | \Sigma_M = l, \Sigma_{\hat{M}} = k) = \frac{\binom{l}{i}\binom{c-l}{k-l+i} \epsilon^i (1-\epsilon)^{l-i} \epsilon^{k-l+i} (1-\epsilon)^{c-k-i}}{\sum_{t \in \Omega} \binom{l}{t}\binom{c-l}{k-l+t} \epsilon^t (1-\epsilon)^{l-t} \epsilon^{k-l+t} (1-\epsilon)^{c-k-t}}. \quad (4.30)$$

où $\theta(k) = \Pr(\hat{M} = 1 | \Sigma_{\hat{M}} = k)$ désigne la stratégie d'attaque dans le domaine des symboles estimés par la coalition.

Calculs de p_2 et p_3 Nous cherchons maintenant à calculer :

$$p_2 = \Pr(M = 1 | M_{j_0} = 1) = \Pr_1(M = 1), \quad (4.31)$$

et :

$$p_3 = \Pr(M = 1 | M_{j_0} = 0) = \Pr_0(M = 1), \quad (4.32)$$

avec $\Pr_1(.) \equiv \Pr(. | M_{j_0} = 1)$ et $\Pr_0(.) \equiv \Pr(. | M_{j_0} = 0)$. À nouveau, en utilisant l'analyse combinatoire, nous obtenons :

$$p_2 = \sum_{l=1}^{c} \Big(\Pr_1(\Sigma_M = l) \sum_{k=0}^{c} \Big(\Pr(\Sigma_{\hat{M}} = k | \Sigma_M = l) \times \sum_{i \in \Omega} \big(\Pr(M = 1 | V = i, \Sigma_M = l, \Sigma_{\hat{M}} = k) \times \Pr(V = i | \Sigma_M = l, \Sigma_{\hat{M}} = k) \big) \Big) \Big), \quad (4.33)$$

et :

$$p_3 = \sum_{l=0}^{c-1} \Big(\Pr_0(\Sigma_M = l) \sum_{k=0}^{c} \Big(\Pr(\Sigma_{\hat{M}} = k | \Sigma_M = l) \times \sum_{i \in \Omega} \big(\Pr(M = 1 | V = i, \Sigma_M = l, \Sigma_{\hat{M}} = k) \times \Pr(V = i | \Sigma_M = l, \Sigma_{\hat{M}} = k) \big) \Big) \Big), \quad (4.34)$$

	$c=3$	$c=4$
$\epsilon=0.$	(0. 0.651 0.349 1.)	(0. 0.487 0.5 0.513 1.)
$\epsilon=0.05$	(0. 0.726 0.274 1.)	(0. 0.543 0.5 0.457 1.)
$\epsilon=0.1$	(0. 0.830 0.170 1.)	(0. 0.620 0.5 0.379 1.)
$\epsilon=0.15$	(0. 0.982 0.018 1.)	(0. 0.734 0.5 0.266 1.)
$\epsilon=0.2$	(0. 1. 0. 1.)	(0. 0.908 0.5 0.091 1.)
$\epsilon>0.2$	(0. 1. 0. 1.)	(0. 1. 0.5 0. 1.)

TABLE 4.2 – Valeurs de $\theta_{\epsilon\text{-WCA}}$ en fonction de ϵ pour $c=3,4$. Pour $c=2$, pour tout ϵ, $\theta_{\epsilon\text{-WCA}}=(0.0.51.)$.

avec :

$$\Pr_1(\Sigma_M=l)=\binom{c-1}{l-1}p^{l-1}(1-p)^{c-l}, \tag{4.35}$$

et :

$$\Pr_0(\Sigma_M=l)=\binom{c-1}{l}p^{l}(1-p)^{c-l-1}. \tag{4.36}$$

Les équations (4.26), (4.33), (4.34) sont par conséquent utilisées pour le calcul des fonctions d'entropie binaire de l'équation (4.21) moyennées en utilisant des techniques d'intégration numérique fournies par la librairie GNU/GSL [2].

4.3.3 Comparaisons entre WCA et ϵ-WCA

Nous calculons dans cette section la ϵ-attaque au pire cas : ϵ-WCA (ϵ *Worst Case Attack*), la stratégie $\theta_{\epsilon\text{-WCA}}$ qui minimise le débit atteignable donné par l'équation 4.21. L'étape de minimisation est assurée par l'utilisation de l'algorithme du simplexe [67]. Pour $c=2$, pour tout ϵ, $\theta_{\epsilon\text{-WCA}}$ =(0. 0.5 1.). La table 4.2 montre la ϵ-WCA pour $c=3,4$ et plusieurs valeurs de ϵ. Nous remarquons ici que les deux stratégies convergent quand l'erreur d'estimation augmente.

La figure 4.7 montre une estimation de la probabilité de fausse alarme p_{fa} (probabilité d'accuser un innocent) en fonction de ϵ pour trois stratégies : aléatoire, WCA et ϵ-WCA pour un seuil d'accusation τ fixé ici à 80. Pour la stratégie d'accusation aléatoire, chaque composante de $\mathbf{m}$ est choisie uniformément parmi les membres de la coalition. Nous estimons la p_{fa} par une moyenne sur 1000 observations en utilisant des techniques d'analyse d'évènements rares comme dans [36]. Comme prévu, les performances de la WCA et de la ϵ-WCA sont meilleures par rapport à la stratégie aleatoire quand $\epsilon=0$. De plus, les performances de la ϵ-WCA sont meilleures par rapport à la stratégie WCA quand ϵ est proche de 0.25.

Nous pouvons voir les performances sur la figure 4.11 qui montre les valeurs du débit atteignable $\mathrm{R}_s(\theta)$ en fonction de ϵ pour les trois stratégies (aléatoire, WCA et ϵ-WCA). Comme prévu, l'information mutuelle entre l'empreinte pirate et l'empreinte d'un adversaire est plus faible pour la ϵ-WCA que pour la WCA et l'aléatoire.

4.4 Compromis robustesse/sécurité

Nous considérons à présent les effets des attaques de robustesse comme la compression ou l'ajout de bruit gaussien sur le document multimédia résultant de l'attaque par coalition. Au lieu de décoder chaque symbole

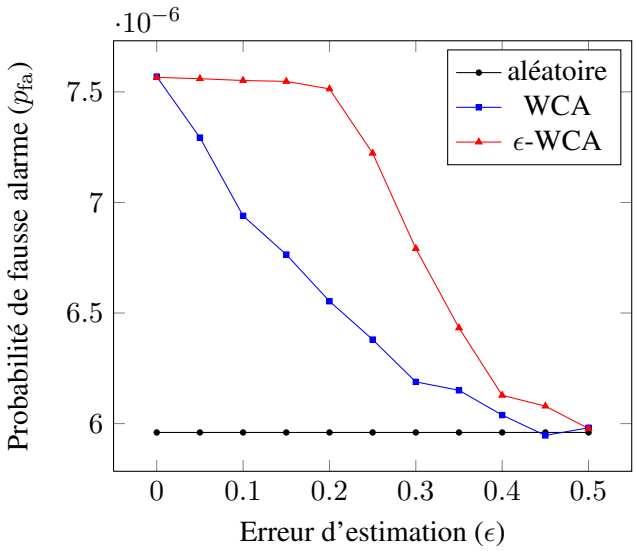


FIGURE 4.7 – Estimation de la probabilité de fausse alarme en fonction de l'erreur d'estimation ϵ pour trois attaques : WCA, ϵ-WCA et aléatoire. Paramètres : $m = 400, \tau = 80, c = 4$.

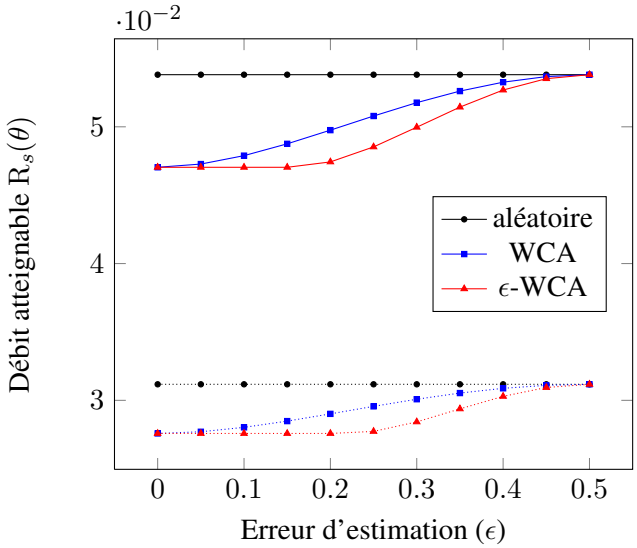


FIGURE 4.8 – Valeurs pour $\mathrm{R}_s(\theta) = \mathbb{E}_P(I(X,Y))$ en fonction de l'erreur d'estimation ϵ pour $c = 3$ (courbes pleines) et $c = 4$ (courbes en pointillés).

$\mathbf{m}(i)$ $(i \in [m])$, le distributeur décode $\mathbf{m}'(i)$ avec un taux d'erreur binaire η modélisant un canal symétrique binaire BSC (*Binary Symmetric Channel*). La variable aléatoire correspondante M' est définie par :

$$\Pr\left(M' = 1 | M = 0\right) = \Pr\left(M' = 0 | M = 1\right) = \eta. \tag{4.37}$$

Le but de cette section est de comparer les débit atteignables des schémas d'insertion des empreintes non-sûrs ($\epsilon = 0$) avec les schémas sûrs ($\epsilon > 0$) en incluant un canal BSC de caractéristique η qui prend en compte la robustesse du schéma de tatouage. Nous calculons à présent le débit atteignable d'un schéma d'estampillage

couplé à un modèle de tatouage R'_s défini par :

$$\mathrm{R}'_s(\theta, \epsilon, \eta) = \mathbb{E}_P[I(M'; M_{j_0})|P = p]. \tag{4.38}$$

Nous calculons $\mathrm{R}'_s(\theta, \epsilon, \eta)$ avec la même méthode que celle proposée pour l'équation (4.21) avec :

$$p'_1 = \Pr(M' = 1) = (1 - \eta)p_1 + \eta(1 - p_1), \tag{4.39}$$

$$p'_2 = \Pr(M' = 1|M_{j_0} = 1) = (1 - \eta)p_2 + \eta(1 - p_2), \tag{4.40}$$

$$p'_3 = \Pr(M' = 1|M_{j_0} = 0) = (1 - \eta)p_3 + \eta(1 - p_3). \tag{4.41}$$

4.4.1 Approche expérimentale

Dans la figure 4.9, pour $c = 4$ adversaires d'une coalition, pour une valeur de $\epsilon > 0$ fixée, nous calculons le BER η_1 tel que le débit atteignable R'_s après une ϵ-WCA soit le même que pour un schéma de tatouage non-sûr $\epsilon = 0$ après attaque de robustesse induisant un BER η_2. η_1 est équivalent à la probabilité d'erreur maximale de décodage que peuvent subir les schémas de tatouage non-sûrs afin d'offrir le même taux de transmission qu'un schéma sûr (dépendant de ϵ). Formellement, nous recherchons la racine η_1 satisfaisant :

$$\mathrm{R}'_s(\theta_{\epsilon\text{-WCA}}, \epsilon, \eta_1) == \mathrm{R}'_s(\theta_{\text{WCA}}, 0., \eta_2). \tag{4.42}$$

η_1 est calculé grâce à l'algorithme de Brent-Dekker [16, 17].

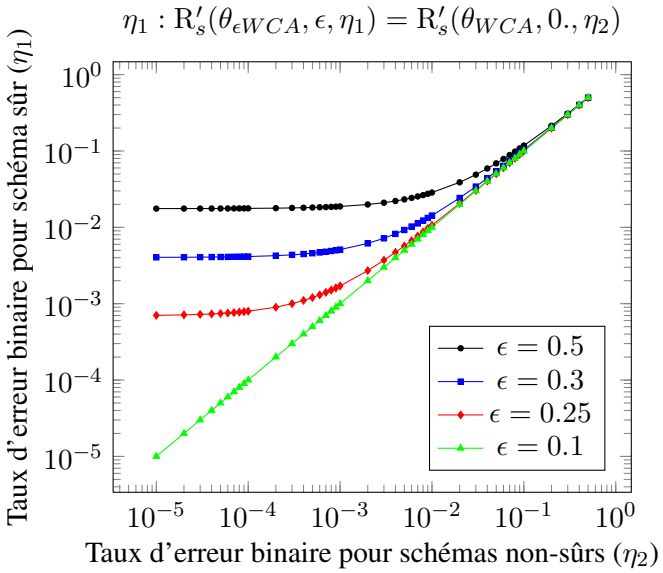


FIGURE 4.9 – Taux d'erreur binaire η_1 (schémas sûrs) en fonction du taux d'erreur binaire η_2 (schémas non-sûrs) pour $\epsilon = 0.5, 0.3, 0.25, 0.1$, $c = 4$.

Cette figure permet de quantifier le compromis entre la sécurité et la robustesse d'un système d'estampillage utilisant des techniques d'insertion par tatouage numérique. Quand ϵ augmente, nous remarquons qu'un schéma de tatouage sûr sera plus résistant aux erreurs qu'un schéma non-sûr. Pour la même information mutuelle entre

une empreinte pirate forgée par une coalition et une empreinte d'un de ses membres, un BER η_2 de $1.e-05$ pour un schéma non-sûr est équivalent à un schéma sûr avec $\epsilon = 0.5$ (correspondant, d'après la classification des schémas de tatouage énoncée dans la section 1.4, à une technique clé-sûre) avec un BER $\eta_1 = 1.761e-02$.

Notons cependant que la différence entre schéma sûr et schéma non-sûr devient négligeable quand la sécurité du schéma sûr n'est pas très élevée ($\epsilon < 0.1$) ou quand le BER augmente.

La figure 4.10 illustre la différence Δ_R entre les débit atteignables des schémas de tatouage sûrs et non-sûrs ayant subi le même BSC de paramètre η pour $c = 4$ adversaires. À nouveau, nous remarquons que la différence entre ces deux schémas est seulement significative pour des schémas très sûrs (ϵ proche de 0.5) et des schémas très robustes (η proche de 0).

D'après les résultats obtenus, nous soulignons l'importance de la sécurité en tatouage pour l'estampillage de contenus. Un schéma de tatouage très sûr et moins robuste offre de meilleures performance qu'un schéma seulement très robuste.

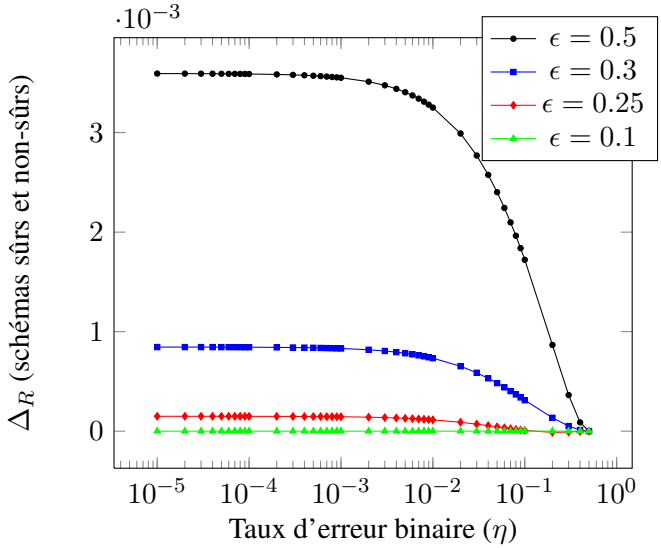


FIGURE 4.10 – $\Delta_R = \mathrm{R}'_s(\theta_{\epsilon WCA}, \epsilon, \eta) - \mathrm{R}'_s(\theta_{WCA}, 0., \eta)$ en fonction du taux d'erreur binaire η pour $\epsilon = 0.5, 0.3, 0.25, 0.1$, $c = 4$.

4.4.2 Application à l'étalement de spectre

Nous nous plaçons maintenant dans un cadre pratique et associons un schéma d'insertion par tatouage au schéma d'estampillage. Nous reprenons ici le même mode opératoire que dans la section précédente en utilisant deux modulations par étalement de spectre :

i) une méthode non-sûre : l'étalement de spectre amélioré **ISS** (équations (1.18) et (1.19)),

ii) une méthode assurant une flexibilité entre schéma non-sûr et schéma totalement sûr : le tatouage ρ circulaire ρ**-CW** (équation (2.1).

Nous mesurons la robustesse à l'aide d'un ajout de bruit gaussien sur les deux schémas.

Dans le scénario d'estampillage, le distributeur produit différentes copies d'un contenu hôte pour plusieurs utilisateurs. Le signal à tatouer est divisé selon N_t tronçons de N_v composantes. Chaque empreinte d'estampillage est un code de Tardos de m bits qui sera inséré dans le signal hôte. Dans cette sous section, nous avons fait le choix de cacher $N_c = 16$ bits dans chaque tronçon de taille $N_v = 512$ selon les hypothèses définies dans la

section 1.5. Pour nos expériences, les signaux hôtes sont distribués selon une loi gaussienne. Les paramètres utilisés pour les deux modulations ISS et ρ-CW sont $WCR = -10\ dB,\ NCR = -10\ dB$.

Dans la figure 4.11, nous comparons la quantité de bruit que les schémas ISS et ρ-CW peuvent subir pour atteindre la même capacité d'estampillage (quantifié par le débit atteignable R'_s). Formellement, nous résolvons de nouveau l'équation 4.42 ici dans un cadre pratique : le WNR appliqué à l'ISS sur 10^6 signaux nous fournit une valeur moyenne pour η_2 avec $\epsilon = 0.3$ estimé par ACI avec $N_t = 1300, \rho = 0.04$ (même mode opératoire que pour la figure 2.2.

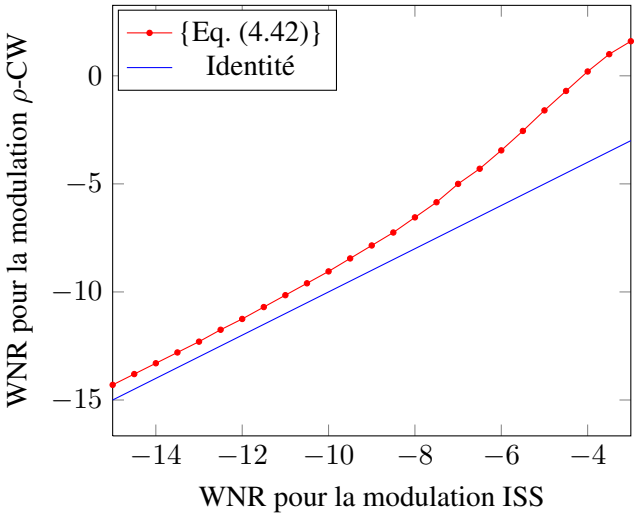


FIGURE 4.11 – WNR pour le ρ-CW (de BER η_1) en fonction du WNR pour l'ISS (de BER η_2). Paramètres : $N_c = 16\ N_v = 512,\ WCR = -10\ dB,\ NCR = -10\ dB$. La relation entre η_1 et η_2 est donnée par : $\mathrm{R}'_s(\theta_{\epsilon WCA}, \epsilon, \eta_1) == \mathrm{R}'_s(\theta_{WCA}, 0., \eta_2)$. $\epsilon = 0.3$ est obtenu par ACI avec $N_t = 1300, \rho = 0.04$.

Ensuite, nous trouvons η_1 à l'aide de l'équation 4.42 (ce calcul utilise de nouveau l'algorithme de Brent-Dekker [16, 17]) ainsi que le WNR correspondant pour le ρ-CW. La figure 4.11 montre le WNR pour la modulation ρ-CW en fonction du WNR pour la modulation ISS pour des capacités d'estampillage équivalentes. Contrairement à nos attentes et aux résultats théoriques des figures 4.9 et 4.10, pour des capacités d'estampillage équivalentes, la modulation ISS supporte mieux les attaques de robustesse que la modulation ρ-CW (par exemple un WNR environ égal à $1\ dB$ pour le ρ-CW équivaut à un WNR d'environ $-2\ dB$ pour l'ISS. Cette perte de robustesse peut s'expliquer par la sécurité du ρ-CW qui est relative. Celle-ci dépend du nombre de contenus utilisés pour l'ACI qui est une donnée non prise en compte dans nos calculs théoriques.

4.5 Conclusion

Dans ce chapitre, nous avons défini les codes de traçage permettant de contrôler les fuites ou redistributions illégales de contenus numériques. Nous avons utilisé les codes de Tardos, performants de par leur longueur réduite pour parer les attaques dites de coalition. Les codes d'estampillage étant couplés à des méthodes de tatouage, nous avons alors défini des attaques au pire cas dépendantes de la sécurité de la méthode d'insertion, permettant aux membres de la coalition de minimiser leur culpabilité au sein de la production d'une copie pirate. De plus nous avons réussi à quantifier de façon théorique le compromis robustesse/sécurité des méthodes de tatouage dans le cadre de l'estampillage, les méthodes sûres et peu robustes étant plus fiables pour un distributeur de contenus que des méthodes non-sûres mais très robustes.

Dans le chapitre suivant nous implantons les codes de Tardos dans un schéma de tatouage de flux vidéo par étalement de spectre agissant image par image dans la matrice de luminance. Nos tests sont effectués sur une bande annonce d'une durée de 90 secondes issue d'un DVD. Nous utilisons les méthodes ISS et CW et testons alors la robustesse sur une attaque pratique (qualité de réencodage des images tatouées) ainsi que la sécurité par le biais des distributions des corrélations entre les signaux et porteuses secrètes. Pour la partie traçage, nous montrons l'efficacité de notre schéma expérimental en comparant les scores d'accusation sur des vidéos générées par une coalition.

Chapitre 5

Estampillage de contenu vidéo

San Francisco [1] (CA, USA).

Sommaire

1. Auteur : Jeremy Brooks, licence CC-BY-NC-3.0 (http://creativecommons.org/licenses/by-nc/3.0/).

DANS le précédent chapitre, nous avons étudié l'impact des attaques par coalition sur les codes d'estampillage. Nous nous posons maintenant la question : le traçage de traitres peut-il s'effectuer de manière efficace en pratique ? En effet, la priorité pour un distributeur de contenus sera ici d'obtenir une très faible probabilité d'accuser un utilisateur innocent tout en désignant au moins un des adversaires responsables de la diffusion illégale d'une œuvre. Le distributeur exigera alors une forte robustesse lors de l'insertion des codes : seules les méthodes de tatouage peuvent alors répondre à cette demande. Nous proposons dans ce dernier chapitre une implantation de codes à la Tardos sur un contenu vidéo image par image à l'aide de méthodes de tatouage par étalement de spectre. Pour la partie **codage** : nous générons des codes d'estampillage discrets binaires *à la* Tardos [79] pour plusieurs utilisateurs. Pour la partie **tatouage**, nous utilisons des techniques par étalement de spectre avec d'une part une modulation non-sûre (ISS) et d'autre par une modulation clé-sûre (CW). Nous considérons ensuite une coalition de quelques utilisateurs, mixant leurs contenus respectifs conformément à la condition de marquage. Ce chapitre est organisé de la façon suivante : dans la section 5.1 nous décrivons notre algorithme de tatouage en explicitant de quelle manière seront insérés les codes d'estampillage dans le flux vidéo, ce dernier étant préalablement décomposé en une suite d'images fixes. Dans la section 5.2, nous donnons les détails techniques au niveau des chaines d'insertion et de décodage ainsi que les valeurs numériques utilisées, les données sont tatoués dans le domaine pixellique à l'aide de méthodes par étalement de spectre. Toujours dans cette même section, nous testons notre algorithme sur une bande originale extraite d'un DVD et nous comparons les méthodes par étalement de spectre utilisées en terme de sécurité et de robustesse. Enfin dans la dernière section 5.3 nous nous intéressons aux attaques dites de coalition : plusieurs adversaires mixent leurs contenus afin de forger une version pirate. Après décodage de la vidéo pirate (sujette à attaques de robustesse lors du réencodage de la vidéo après tatouage image par image) par le distributeur, nous vérifions la pertinence des scores d'accusation pour chaque membre de la coalition en les comparant avec le score maximal d'un utilisateur innocent. Ces travaux ont été réalisés à l'aide de Floran Devillez, étudiant en Master professionnel de sécurité et cryptologie de l'université Joseph Fourier de Grenoble et en stage au laboratoire GIPSA en 2010.

5.1 Schéma de tatouage d'un flux vidéo

5.1.1 Tatouage image par image

Afin de pouvoir associer un contenu vidéo et un utilisateur précis, le distributeur doit pouvoir cacher un certain nombre de bits dans le contenu tout en respectant les contraintes d'imperceptibilité, de robustesse et de sécurité imposées par les techniques de tatouage. Nous considérons le contenu vidéo comme une suite d'images de $M \times N$ pixels, chaque image est découpée en tuiles de $S_z \times S_z$ pixels. Le code de Tardos de N_T bits correspondant à l'identifiant de l'utilisateur sera découpé en tronçons de tailles égales (N_c bits), une tuile permettra alors de cacher un tronçon. Le format d'image qui sera utilisé dans cette implantation sera le Portable Pixel Map (*PPM*) : une image de taille $M \times N$ est stockée sous la forme d'une matrice $\mathcal{M}_{M,N\times 3}(\mathbb{R})$, chaque pixel est codé par un triplet d'échantillons rouge, vert et bleu dans cet ordre. Nous avons choisi de tatouer les identifiants dans les composantes de luminance de l'image afin d'assurer une meilleure robustesse face aux traitements que peut subir la vidéo une fois marquée. Si l'on note R, G, B l'intensité de chaque couleur dans un pixel, la luminance Y est donnée par $Y = 0.299R + 0.587G + 0.114B$. La matrice de ces composantes est ensuite découpée en tuiles carrées de taille $S_z \times S_z = N_v$.

Dans chaque tuile d'une image, nous cachons un message $\mathbf{m} \in \mathbb{F}_2^{N_c}$ dans le domaine pixellique. Chaque tuile à tatouer est convertie en un vecteur $\mathbf{x} \in \mathbb{R}^{N_v}$. Ce vecteur est alors tatoué à l'aide d'une clé secrète K en utilisant les modulations par étalement de spectre ISS ou CW. La figure 5.1 illustre les mécanismes du tatouage de la vidéo

image par image.

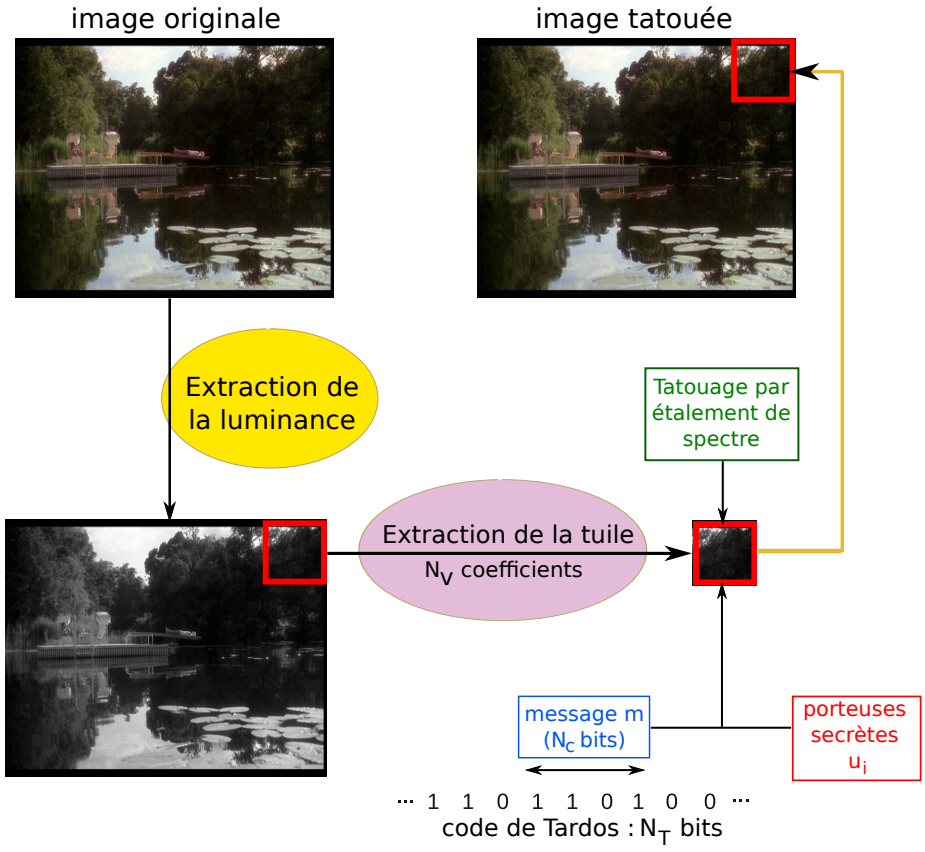


FIGURE 5.1 – Tatouage d'une image au format *PPM* : la luminance de l'image est découpée en tuiles de taille $S_z \times S_z = N_v$, chaque tronçon de N_c bits de l'identifiant est caché dans une tuile par tatouage par étalement de spectre.

5.1.2 Construction du bitstream

Nous détaillons maintenant les informations qui seront tatouées dans le flux vidéo. La quantité d'images (donc de tuiles) présentes dans une vidéo (1 seconde $\equiv$ 25 images) nous permet d'obtenir un espace de stockage des données de l'identifiant assez grand pour que ce dernier puisse être répété : la robustesse en sera améliorée. Cependant, nous voulons rester dans un cadre proche du WOA : un adversaire possède plusieurs signaux tatoués (tuiles de taille $S_z \times S_z$) avec des messages de N_c bits différents. Pour éviter de se retrouver dans le cadre d'une CMA (*Constant Message Attack*) à cause des répétitions du code de Tardos, nous proposons de permuter chaque répétition d'un code de Tardos complet tout au long du flux vidéo. Contrairement aux analyses de sécurité en WOA, les coefficients de chaque porteuse secrète ne sont ici plus indépendants les uns des autres, mais les relations entre les coefficients de portion de la porteuse d'origine et de la portion de la porteuse permutée sont liées par la clé secrète.

Le schéma général du flux vidéo qui sera tatoué image par image est illustré figure 5.2. La partie en vert correspond aux données publiques : c'est-à-dire les données qui seront connues par le ou les adversaires. Il s'agit ici d'un identifiant propre à la vidéo que l'on souhaite tatouer. Celui-ci permet la resynchronisation par le distributeur du début de chaque code de Tardos permuté si la vidéo est rognée par les adversaires, permettant

ainsi de limiter les erreurs de décodage du tatouage. La contrepartie de cet ajout de bits publics réside dans une meilleure estimation des porteuses secrètes par les adversaires si la méthode de tatouage est non-sûre : une ACI permet d'estimer les porteuses secrètes au signe et à l'ordre près. Nous avons choisi de cacher une petite quantité de bits par tuile ($N_c = 4$) : les indéterminations précédentes sur les porteuses peuvent être levées avec un nombre de bits connus de l'ordre de $\log_2(N_c)$. Néanmoins, ceci n'influe pas sur l'efficacité des codes de Tardos, cet identifiant public propre à la vidéo ne rentre pas dans le processus d'accusation des adversaires.

La partie en rouge correspond aux répétitions du code de Tardos propre à l'utilisateur, cette partie est (a priori) inconnue de la coalition. Les permutations permettent de sortir du cadre CMA et ainsi d'empêcher une estimation correcte des porteuses par les adversaires. La permutation que nous avons choisie ici est le shuffle de Fisher-Yates [29] (aussi connu sous le nom de shuffle de Knuth [49]) présenté dans l'algorithme 2. S'appliquant sur des vecteurs d'entiers, le shuffle permet de créer une permutation à partir d'un générateur de nombres pseudo-aléatoires.

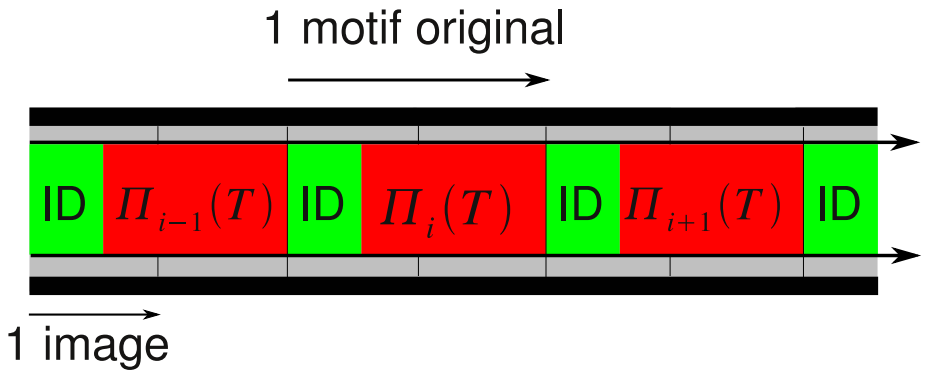


FIGURE 5.2 – Tatouage du flux vidéo image par image. La partie en rouge correspond aux permutations du code de Tardos propre à l'utilisateur, la partie en vert est un identifiant public propre à la vidéo.

Algorithme 2 Shuffle de Fisher-Yates (shuffle de Knuth).

ENTRÉE : $\mathbf{a} \in \mathbb{R}^n$.
SORTIE : $\mathbf{a} \in \mathbb{R}^n$ permuté.
  Pour $i = n - 1$ à 0 **Faire**
    $j \leftarrow$ nombre aléatoire avec $0 \leq j \leq i$;
    échanger $\mathbf{a}(i)$ et $\mathbf{a}(j)$;
    $i \leftarrow i - 1$;
  Fin Pour

Le distributeur, à l'aide d'une clé secrète (graine) K, génère un ensemble de clés $\{K_i\}_i$ de cardinal le nombre de répétitions du code de Tardos possibles pour une vidéo donnée. Ces K_i permettent alors de réinitialiser le GNPA pour chaque permutation π_i. De cette façon le distributeur peut retrouver la séquence initiale en appliquant les permutations inverses et ainsi améliorer le décodage des identifiants.

5.2 Implantation

Dans cette section, plus technique, nous présentons les techniques employées pour tatouer le flux vidéo ainsi que les grandeurs utilisées. De plus, nous évaluons les résultats obtenus sur la sécurité et la robustesse des mé-

thodes de tatouage par étalement de spectre utilisées (ISS & CW). Le programme d'estampillage de la vidéo (insertion / décodage) est réalisé à l'aide des langage C (librairie LIBIT [43]) et Perl.

5.2.1 Mise en place du tatouage

Nous détaillons les scripts d'insertion (tatouage, figure 5.3) et de décodage (figure 5.4).

5.2.1.1 Chaîne d'insertion

Les tests sont réalisés sur une bande annonce d'une durée de 90 secondes récupérée sur un DVD au format *VOB*. Nous ne nous intéressons pas à la bande sonore. Si on note N_c le nombre de bits à insérer dans chaque tuile de taille $S_z \times S_z$, N_b le nombre de bits cachés dans chaque image, N_T la longueur d'un code de Tardos complet, N_{id} la longueur de l'identifiant public, N_i le nombre d'images dans un motif (nombre d'images nécessaires pour cacher un code de Tardos) et enfin $M \times N$ les dimensions des images, nous obtenons le système d'équation suivant :

$$\begin{cases} N_b = \lfloor \frac{M}{S_z} \rfloor \times \lfloor \frac{N}{S_z} \rfloor \times N_c, \\ N_T = N_i \times N_b - N_{id}. \end{cases} \tag{5.1}$$

N_T et N_{id} sont des variables fixées au préalable (dépend du nombre de vidéos que possède le distributeur ainsi que du nombre d'utilisateurs, de la taille maximale de la coalition à laquelle nous voulons résister et de la probabilité d'accuser un innocent p_{fa}). Les variables S_z et N_c sont choisies en fonction des besoins en robustesse et imperceptibilité (des tuiles de petite taille permettent une meilleure resynchronisation de l'image en cas d'attaque géométrique, un faible nombre de bits par tuile sera plus robuste aux attaques de type compression).

Dans un premier temps, nous générons, pour chaque utilisateur, un identifiant (code de Tardos) et l'ensemble des permutations à l'aide d'une graine de GNPA K_s. Ensuite, nous décomposons le flux vidéo en images de format *PPM* à raison de **25 images par seconde** à l'aide du logiciel *MPlayer*. Le tatouage s'effectue par étalement de spectre pour chaque tuile et chaque tronçon de l'identifiant à tatouer en fixant un PSNR pour le budget de distorsion autorisé. L'image tatouée au format *PPM* est convertie au format *PNG* (compression sans perte) à l'aide de la commande Unix *pnmtopng*.

L'ensemble des images tatouées est réencodé sous la forme d'un fichier vidéo par le logiciel *MEncoder* en spécifiant le débit (en kbits/sec).

5.2.1.2 Chaîne de décodage

Le processus de décodage doit pouvoir, à partir d'une vidéo tatouée (probablement attaquée) récupérée sur un réseau d'échange, sortir un identifiant binaire de taile N_T (code de Tardos). À l'aide de *MPlayer*, nous sectionnons la vidéo en images *PPM* à raison de 25 images par seconde. Nous remplissons ensuite un vecteur avec les corrélations entre chaque tuile et les porteuses secrètes $\{\mathbf{u}_i\}_{i\in[N_c]}$. Nous connaissons les permutations qui ont été effectuées sur chaque répétition d'un code de Tardos (les π_i faisant office ici d'une clé secrète), nous sommes alors capable d'ordonner le vecteur de corrélations. Il nous suffit alors d'additionner les corrélations correspondant au même bit de l'identifiant original. Augmenter les corrélations en valeur absolue nous permet alors d'augmenter la robustesse du tatouage : si une erreur s'est produite pendant le décodage d'un symbole, celle-ci sera compensée par les répétitions du code dans tout le flux vidéo.

Une fois les N_T corrélations obtenues, nous regardons leur signe bit à bit pour décoder l'identifiant grâce à l'équation 1.14.

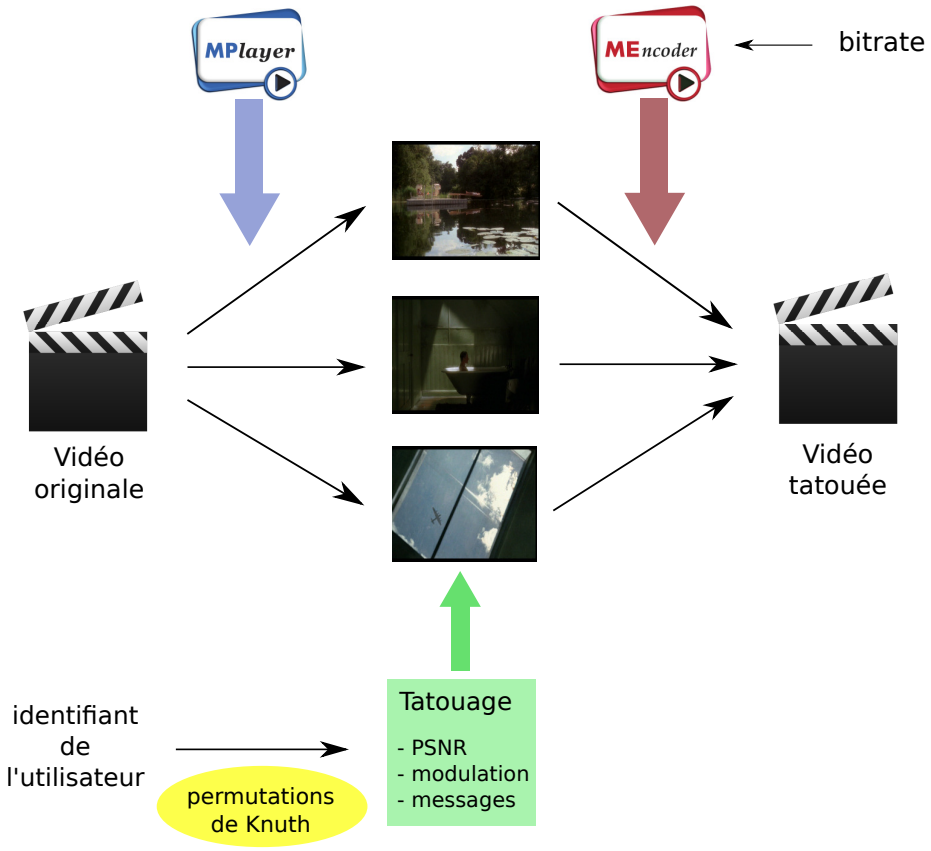


FIGURE 5.3 – Chaîne d'insertion de l'identifiant de l'utilisateur dans le flux vidéo.

5.2.1.3 Dimensionnement

Nous cette sous-section, nous fixons les valeurs des grandeurs utilisée pour le tatouage d'une vidéo. Les tests sont effectués sur une bande annonce d'un DVD. La vidéo est décomposée en images de hauteur $M = 576$ et de largeur $N = 720$. La matrice de luminance de chaque image est alors découpée en tuiles carrées de taille $S_z \times S_z = 32 \times 32 = N_v$. Nous cachons $N_c = 4$ bits dans chaque tuile de l'image, soit un total de $N_b = 1584$ bits par image.

L'identifiant public de la vidéo est composé de $N_{id} = 128$ bits. Le code de Tardos que nous utilisons pour chaque utilisateur est construit selon l'équation (4.1). Afin d'assurer une probabilité d'accuser un innocent (p_{fa}) et une probabilité de n'accuser aucun membre d'une coalition (p_e) faibles, nous cherchons à obtenir un code de Tardos d'une longueur $N_T \geq 3000$ bits, ce qui est largement suffisant pour $n = 20$ utilisateurs et une coalition $\mathcal{C}$ d'au plus $c = 4$ adversaires. Quand on sait qu'une seconde de vidéo est décomposée en 25 images, nous pouvons alors facilement augmenter la taille de l'identifiant pour une utilisation à plus grande échelle (en augmentant le nombre d'utilisateurs n ainsi que la taille de la coalition c à laquelle nous voulons résister.

Nous cherchons aussi à tatouer un code de Tardos complet sur un multiple entier d'images (afin d'éviter des problèmes de synchronisation). D'après les équations données par (5.1), un motif complet (identifiant public + code de Tardos) sera codé sur $N_i = 2$ images avec $N_T = 2 \times 1584 - 128 = 3040$ bits.

Les modulations utilisées sont l'ISS et le CW avec une valeur de $\mathrm{NCR} = 0\ dB$. Le WCR est calculé pour

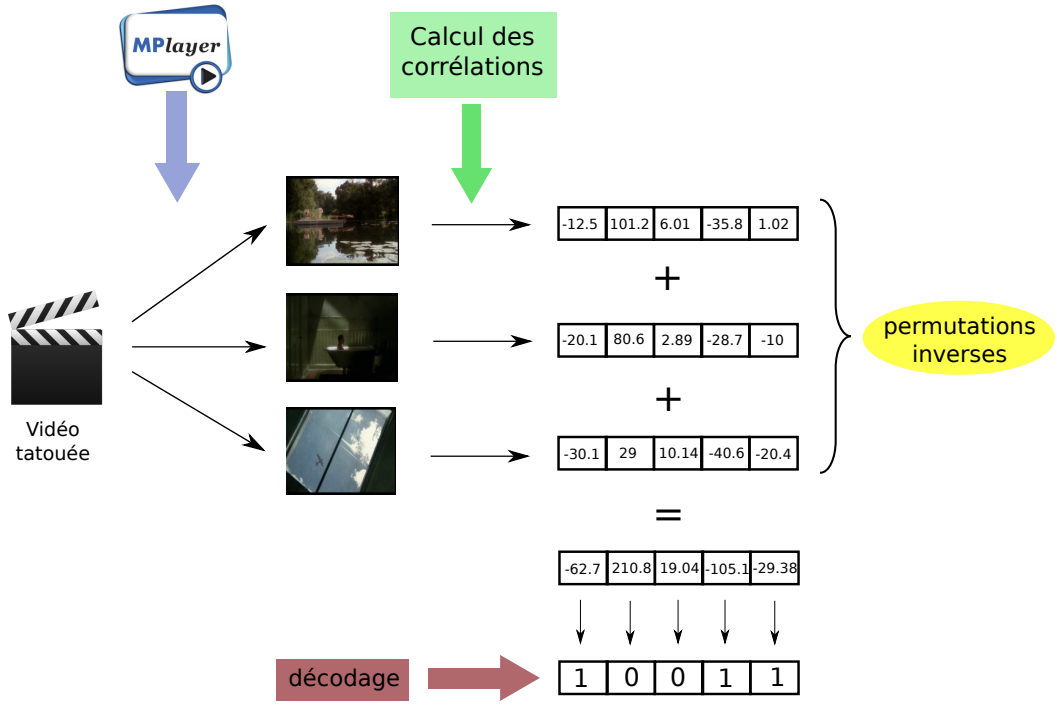


FIGURE 5.4 – Chaîne de décodage de l'identifiant de l'utilisateur dans le flux vidéo. Nous proposons ici un code correcteur basé sur la répétition dont la sécurité est améliorée en utilisant des fonctions de permutations difficiles à inverser.

chaque tuile $\mathbf{T}$ selon la formule :

$$\mathrm{WCR} = 10 \log_{10} \left(\frac{255^2}{\sigma^2_{\mathbf{T}}} - \mathrm{PSNR} \right). \tag{5.2}$$

Le but de cette partie ne concerne pas un rendu optimal de l'imperceptibilité. En effet, nous avons choisi un tatouage dans le domaine pixellique sans masquage psychovisuel. Une application dans le domaine industriel choisira un domaine transformé de l'image plus apte à recevoir le tatouage. La figure 5.5 montre une image de notre bande annonce tatouée par ISS avec des PSNR de 33 dB et 44 dB.

5.2.2 Validation du schéma de tatouage pour les contraintes de robustesse et de sécurité

5.2.2.1 Robustesse

Nous avons tatoué notre flux vidéo avec un PSNR de 46 dB par tuile en faisant varier trois paramètres :

1. la modulation : ISS ou CW,
2. le nombre de répétitions d'un motif (identifiant public + code de Tardos),
3. le débit (kbits par seconde) de réencodage de la vidéo après tatouage des images.

Nous mesurons la robustesse de notre schéma en mesurant le taux d'erreur binaire (BER) des images obtenues après décomposition de la vidéo tatouée. La figure 5.6 montre le BER obtenu en fonction du débit de réencodage. Les légendes se lisent ainsi : **modulation - nombre de répétitions**.

Image originale

Image tatouée ($33\ dB$)

Image tatouée ($44\ dB$)

FIGURE 5.5 – Image originale et images obtenues après tatouage dans le domaine de la luminance pour des PSNR par tuile de $33\ dB$ et $44\ dB$. Modulation : ISS.

Nous remarquons que l'ISS est plus robuste que le CW. Cependant, cette différence entre les deux modulations devient négligeable lorsque le nombre de répétitions du motif original augmente. À partir de 10 répétitions (soit 20 images, 31680 bits, 8/10e seconde), la robustesse est très acceptable en pratique.

5.2.2.2 Sécurité

Nous utilisons deux modulations ISS et CW. L'ISS est qualifiée de **non-sûre**, la distribution des contenus marqués dans le sous-espace engendré par les porteuses est différente pour chaque clé, il est donc possible d'estimer les porteuses car les corrélations entre les signaux tatoués et les porteuses secrètes à cause de la présence de clusters (constellation) dans le sous-espace privé de dimension N_c.

Au contraire, le CW est **clé-sûr**, il existe un sous-ensemble de clés (ensemble des bases du sous-espace engendré par les porteuses) donnant la même distribution (pour $N_c = 2$, il s'agit de toutes les bases obtenues par rotations centrales des porteuses secrètes).

Nous allons alors nous assurer que les propriétés statistiques des contenus tatoués correspondent aux propriétés théoriques des modulations. Lors de la phase du décodage, nous récupérons les corrélations entre chaque tuile et les porteuses secrètes $\{\mathbf{u}_i\}_{i \in N_c}$. Afin de pouvoir illustrer la distribution de ces corrélations, nous cachons $N_c = 2$ bits par tuile de taille 32×32 pixels.

La figure 5.7 montre la distribution des corrélations entre chaque tuile et les porteuses secrètes pour 19800 tuiles (50 images, 2 secondes de vidéo) pour la modulation ISS avec un PSNR de $44\ dB$. Nous remarquons bien

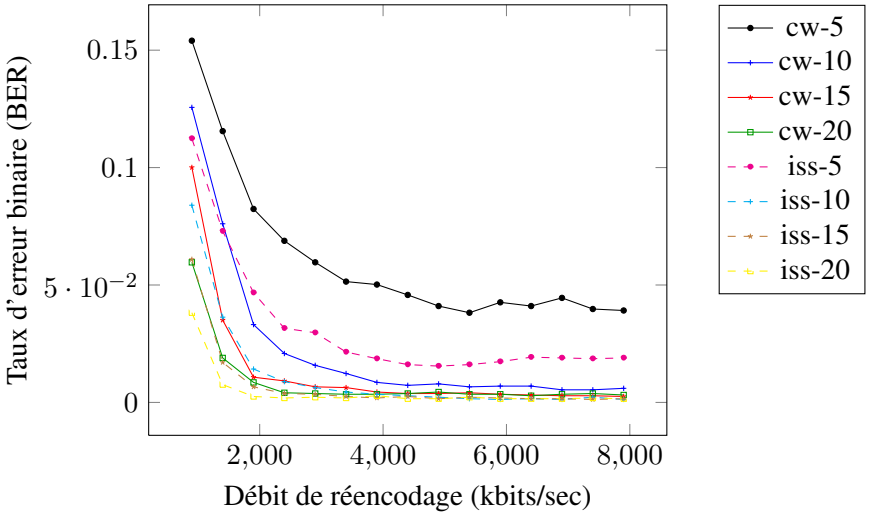


FIGURE 5.6 – Robustesse du schéma de tatouage vidéo : BER en fonction du débit de réencodage de la vidéo après tatouage image par image. PSNR $= 46\ dB$. Les légendes se lisent ainsi : **modulation - nombre de répétitions**. Conformément aux approches théoriques, l'ISS est plus robuste que le CW. Toutefois, la faible robustesse du CW est compensée par le nombre de répétitions du motif original lorsque celui-ci augmente.

la forme caractéristique de la modulation ISS sous forme de constellation, une estimation des porteuses par ACI est alors possible (voir figure 3.9).

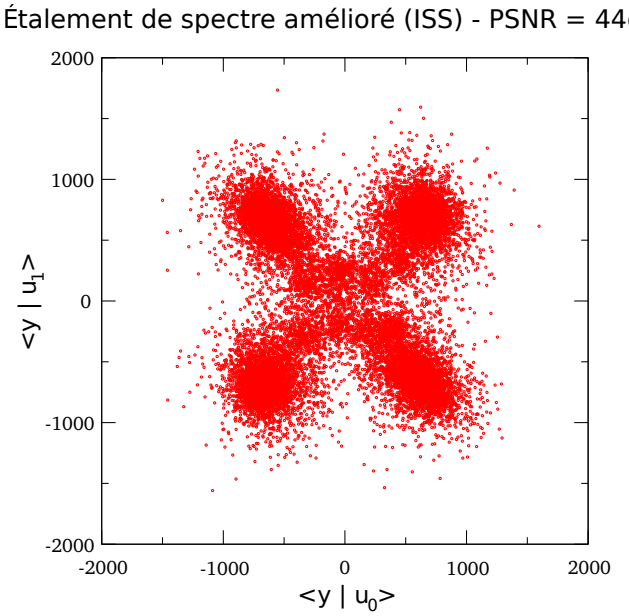


FIGURE 5.7 – Corrélations entre 19800 tuiles ($\equiv$ 50 images, 2 sec de vidéo) tatouées par ISS et les porteuses secrètes. PSNR $= 44\ dB$.

La figure 5.8 montre la distribution des corrélations entre chaque tuile et les porteuses secrètes pour 19800 tuiles pour un PSNR de $44\ dB$ par tuile pour la modulation CW (NCR = $0\ dB$), la distribution est circulaire (équation (1.27)), ce qui vérifie le clé-sécurité de la modulation CW. Nous remarquons mieux la forme caracté-

ristique du CW en "anneau" sur la figure 5.9 qui montre cette même distribution avec cependant une puissance d'insertion des messages plus forte, i.e. $\mathrm{PSNR} = 30\ dB$ par tuile.

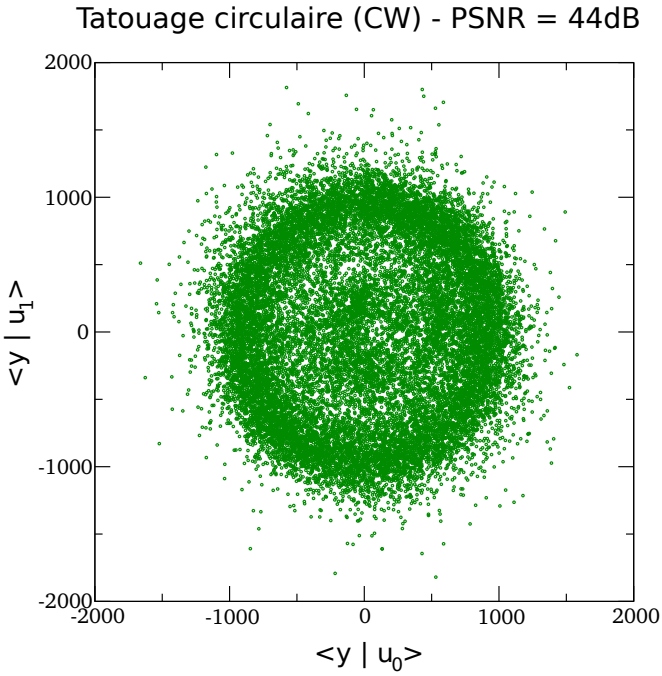


FIGURE 5.8 – Corrélations entre 19800 tuiles ($\equiv$ 50 images, 2 sec de vidéo) tatouées par CW et les porteuses secrètes. $\mathrm{PSNR} = 44\ dB$.

Pour les deux modulations, nous avons de nombreuses corrélations proches de 0. Ce résultat s'explique par la nature des images que nous tatouons. En effet, il est fréquent d'apercevoir dans une vidéo des images noires ou très sombres. Par exemple, nous pouvons apercevoir des fondus entre les différentes séquences d'un film (ce qui est encore plus fréquent dans notre bande annonce test). Nous avons également au sein d'une vidéo des images plus colorées, les tuiles les composants sont parfois de variance proches de 0, comme présenté sur la figure 5.10.

5.3 Attaques de coalition

Nous nous intéressons à présent à l'efficacité de notre algorithme d'insertion et de décodage face aux attaques de coalition.

5.3.1 Mise en place de la stratégie d'attaque

Nous partons de vidéos tatouées par la modulation CW pour $n = 20$ utilisateurs avec les paramètres cités dans la section 5.2.1.3, un PSNR de $44\ dB$, un NCR de $0\ dB$ ainsi qu'un débit de 9800 Kbps pour le réencodage. Nous créons une coalition à partir de $c = 4$ utilisateurs pour forger une version pirate du contenu.

Dans un premier temps notre coalition décompose chacune des séquences vidéos en images ainsi qu'en tuiles. En accord avec le principe de Kerckhoffs, les seules données inconnues des adversaires sont : le vecteur des $\{p_i\}_{i\in[N_T]}$ pour la génération des codes, les permutations π_i de ces codes dans chaque flux vidéo ainsi que les porteuses $\{\mathbf{u}_i\}_{i\in[N_c]}$.

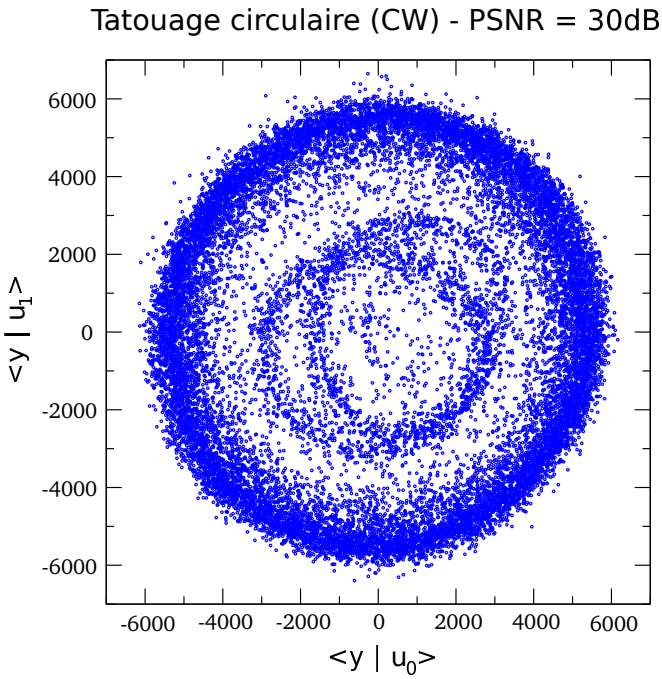


FIGURE 5.9 – Corrélations entre 19800 tuiles ($\equiv$ 50 images, 2 sec de vidéo) tatouées par CW et les porteuses secrètes. PSNR $= 30\ dB$.

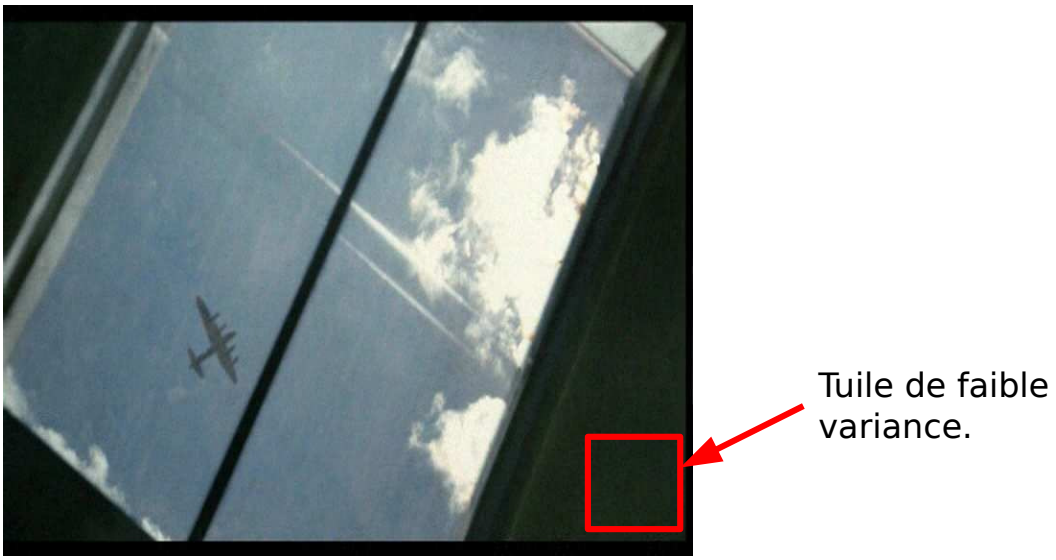


FIGURE 5.10 – Exemple d'une image tirée de notre bande annonce test possédant des tuiles de faible variance, empêchant un tatouage correct des données.

Un point important dans la construction des empreintes avec les codes de Tardos est le respect de la condition de marquage. Nous avons vu précédemment que la stratégie d'attaque choisie par les adversaires respecte la condition de marquage si le symbole candidat pour la séquence pirate est choisi parmi les symboles des membres de la coalition. Dans notre modèle d'estampillage utilisant les techniques de tatouage par étalement de spectre,

les stratégies d'attaque définies par l'équation 4.19 ne peuvent être correctement appliquées avec $N_c > 1$.

Ce problème ne se pose pas lorsque $N_c = 1$ car les images pirates seraient forgées en concaténant les tuiles des adversaires) et respecteraient la stratégie d'attaque symbole par symbole. Cependant la modulation CW nous oblige à utiliser au moins deux porteuses secrètes pour l'insertion.

Pour résoudre ce problème, nous utilisons la même technique que celle proposée par Hartung et Kutter [37] en utilisant des techniques d'insertion sûres : en considérant les tuiles de taille N_v, la clé secrète utilisée pour cacher les N_c bits sont N_c porteuses $\mathbf{u}_i \in \mathbb{R}^{N_v}$ construites selon :

$$\forall\, j \in [N_v],\ \mathbf{u}_i(j) \begin{cases} \sim \mathcal{N}(0,1) \text{ si } j \in [\![i\tfrac{N_v}{N_c}; (i+1)\tfrac{N_v}{N_c} - 1]\!], \\ = 0 \text{ sinon.} \end{cases} \tag{5.3}$$

Ces porteuses sont ensuite standardisées. Ce procédé implique que, pour une tuile $\mathbf{y}_j,\ j \in \mathcal{C}$ construite selon l'équation (1.6), le premier bit est seulement inséré dans les $\frac{N_v}{N_c}$ premières composantes du signal, le second bit dans les $\frac{N_v}{N_c}$ composantes suivantes, etc. Par conséquent, les adversaires peuvent forger un signal pirate en mixant leurs fragments de $\frac{N_v}{N_c}$ coefficients selon une stratégie prédéfinie tout en respectant la condition de marquage. Ainsi, les membres de la coalition sont capables de mixer leurs tuiles selon une stratégie donnée par l'équation (4.19).

La stratégie que nous utilisons ici est la ϵ-WCA pour $c = 4$ adversaires (table 4.2). La valeur de ϵ obtenue par ACI est ici de 0.3 [2], l'ordre des porteuses estimées dans $[N_c]$ étant déterminé grâce aux bits de l'identifiant public. La connaissance du signe des porteuses n'est pas une donnée utile pour la coalition pour appliquer une ϵ-WCA (symétrie des symboles "0" et "1"). Pour chaque tuile et pour chaque image, la coalition sélectionne la portion de tuile (ici un quart de tuile car $N_c = 4$) chez un de ses membres choisi aléatoirement parmi ceux possédant le symbole fourni par la stratégie 0.3-WCA. La figure 5.11 illustre ce principe. En mettant bout à bout les portions de tuiles, la coalition obtient les images piratées. Cette construction respecte alors la condition de marquage.

5.3.2 Scores d'accusation

Une fois les images piratées construites selon la stratégie ϵ-WCA, la coalition procède donc à un réencodage des images attaquées en flux vidéo. Cependant, la qualité de réencodage dépend alors d'un débit spécifié par la coalition et fait ici office d'attaque de robustesse. Il est important de noter qu'avec une qualité très basse (≈ 500 Kbps), le BER obtenu lors du décodage peut être fort (voir figure 5.6), les bits insérés dans la vidéo peuvent alors être modifié entrainant un non-respect de la condition de marquage !

Nous avons calculé les scores d'accusation pour chaque membre de la coalition ainsi que pour les autres utilisateurs innocents en fonction de la qualité de réencodage choisie par la coalition ainsi que du nombre de répétitions du motif original (code de Tardos + identifiant public). Notons que la vidéo originale a été réencodée avec un débit de 9800 Kbps après tatouage par le distributeur.

La figure 5.12 montre les scores d'accusation pour les $c = 4$ adversaires ainsi que pour l'utilisateur ayant le plus grand score chez les innocents en fonction de la qualité de réencodage en Kbps fixée par la coalition. Nous remarquons que les scores entre utilisateurs innocents et adversaires peuvent être séparés par un seuil pouvant être spécifié par le distributeur, même lorsque la qualité de la vidéo est faible (≈ 300 Kbps).

Nous observons maintenant ces mêmes scores d'accusation, cette fois-ci en fonction du nombre de répétitions du motif original (code de Tardos + identifiant public) choisi pour le décodage par le distributeur pour un débit

2. Nous pourrions nous attendre à $\epsilon = 0.5$ car le CW est *clé-sûr*. Cependant, lorsque N_c est proche de 1, les porteuses estimées sont en moyenne légèrement corrélées avec les porteuses réelle, même si celles-ci sont choisie aléatoirement. Pour obtenir $\epsilon = 0.5$, il faudrait prendre des valeurs de N_c beaucoup plus grandes.

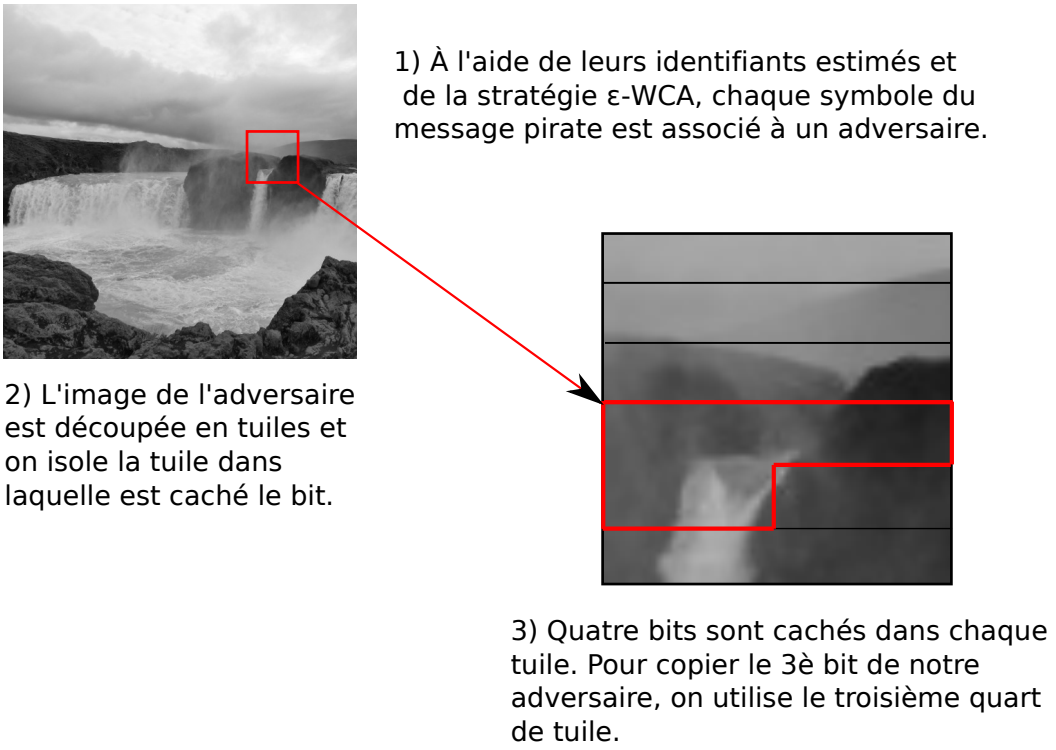


FIGURE 5.11 – Sélection d'un quart de tuile (cachant 1 bit) chez un adversaire correspondant au bit estimé choisi par la coalition.

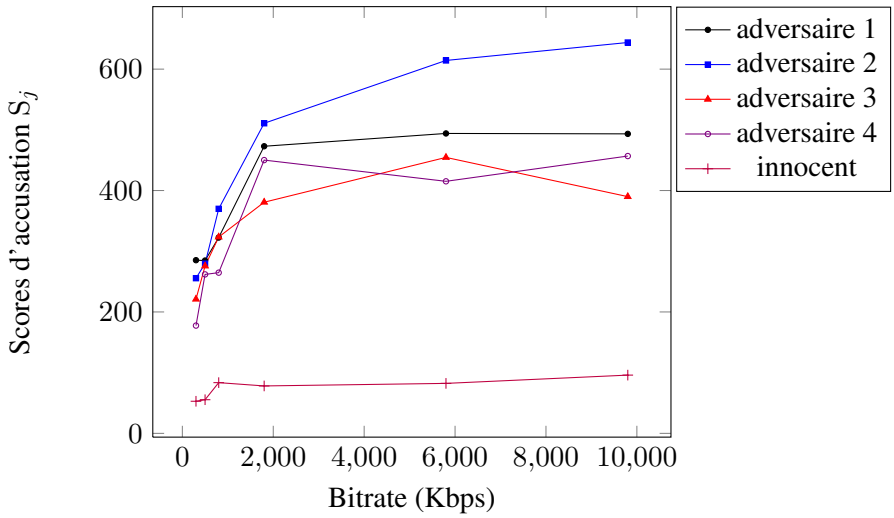


FIGURE 5.12 – Scores d'accusation S_j, $j \in [n]$ pour $c = 4$ adversaires et pour l'utilisateur ayant le plus grand score chez les innocents (au nombre de 16).

de 9800 Kbps (5.14), 800 Kbps (5.16) et 300 Kbps (5.18). Les figures 5.13, 5.15 et 5.17 montrent des images extraites des vidéos pour ces trois débits.

Nous remarquons alors que les scores des adversaires se rapprochent de ceux des innocents lorsque la qualité de réencodage baisse mais peuvent toujours être séparés par un seuil τ donné même avec 2 répétitions ($\equiv$ 4

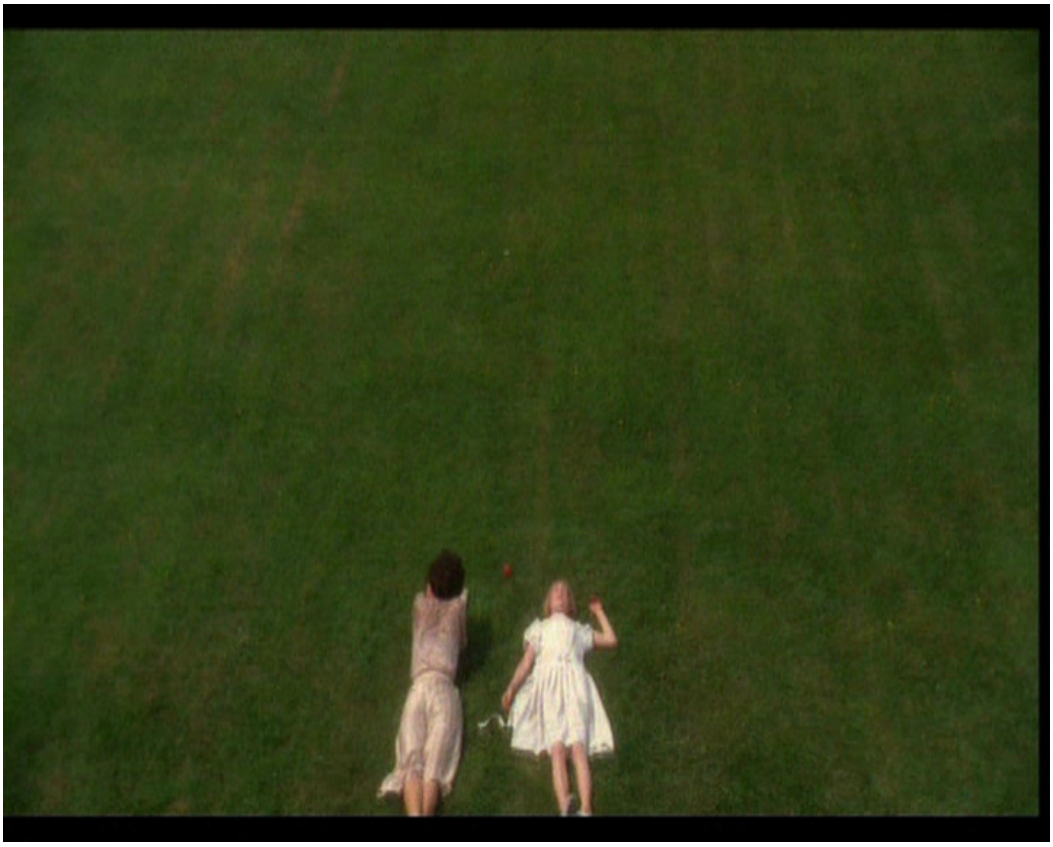


FIGURE 5.13 – Échantillon d'image pour un débit de réencodage de 9800 Kbps.

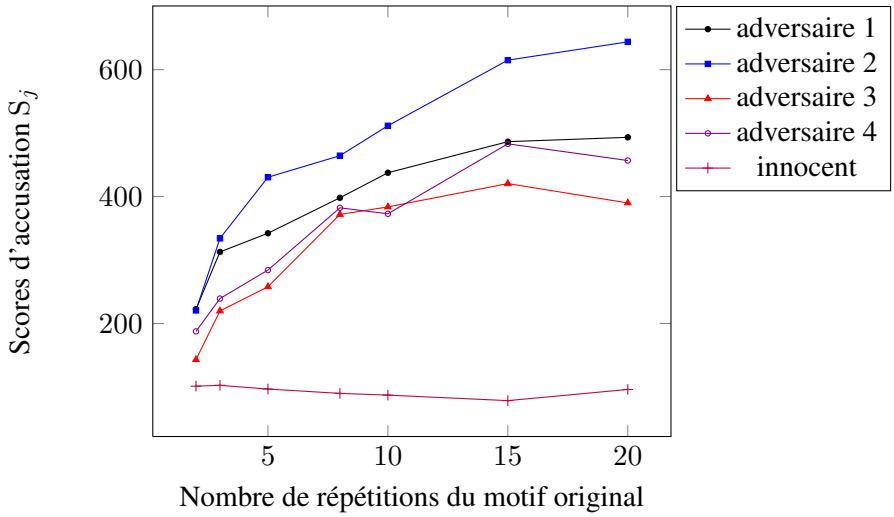


FIGURE 5.14 – Scores d'accusation $S_j,\ j \in [n]$ pour $c = 4$ adversaires et pour l'utilisateur ayant le plus grand score chez les innocents (au nombre de 16) en fonction du nombre de répétitions du motif original (code de Tardos + identifiant public) pour une qualité de réencodage de 9800 Kbps après attaque par la coalition.

images).

FIGURE 5.15 – Échantillon d'image pour un débit de réencodage de 800 Kbps.

5.4 Débits atteignables en pratique

Le but de cette sous-section est de mettre en valeur l'utilisation d'attaques optimales pour une coalition grâce aux débits atteignables. Nous calculons $\mathrm{R}_s(\theta, \epsilon)$ (équation 4.21) pour $\epsilon = 0.3$ (valeur correspondant à l'erreur estimé par ACI dans la sous-section précédente) pour trois stratégies d'attaque : $\theta_{\text{Aléatoire}}$ (équation (4.12)), θ_{WCA} (équation (4.15)) et $\theta_{\epsilon-\text{WCA}}$ (table 4.2).

Pour le calcul pratique d'un débit atteignable, nous procédons de la manière suivante : nous générons tout d'abord m valeurs de p_i ainsi que des codes de Tardos de longueur m pour c utilisateurs (formant ici une coalition). Après simulation d'une erreur d'estimation des codes de $\epsilon = 0.3$, la coalition forge alors une séquence pirate $\mathbf{m}$ par l'équation 4.20 et selon une stratégie donnée. Cette opération est alors répétée N_o fois afin d'obtenir une estimation des probabilités p_1, p_2 et p_3 pour chaque $i \in [m]$ (section 4.3.2.2) par la méthode de Monte-Carlo. Le débit atteignable est alors calculé comme la moyenne des informations mutuelles entre séquences (calculées pour chaque $i \in [m]$ à l'aide de p_1, p_2 et p_3).

Il est important de noter que le calcul du débit atteignable sert ici de fonction d'accusation (décodeur simple) par le distributeur qui n'a a priori aucune information sur la stratégie adoptée par la coalition (contrairement au calcul théorique des débits de la section 4.3.2.2). La figure 5.19 illustre ce processus d'accusation.

La table 5.1 montre les résultats obtenus pour $c = 3, 4$ adversaires. Ces calculs ont ici été obtenus avec $m = 1e4$ et $N_o = 1e5$. Nous remarquons que, conformément aux calculs théoriques des débits atteignables calculés dans la figure 4.11, l'utilisation d'attaques au pire cas permet, pour une coalition, de minimiser l'information mutuelle entre leur séquence pirate et leurs identifiants originaux.

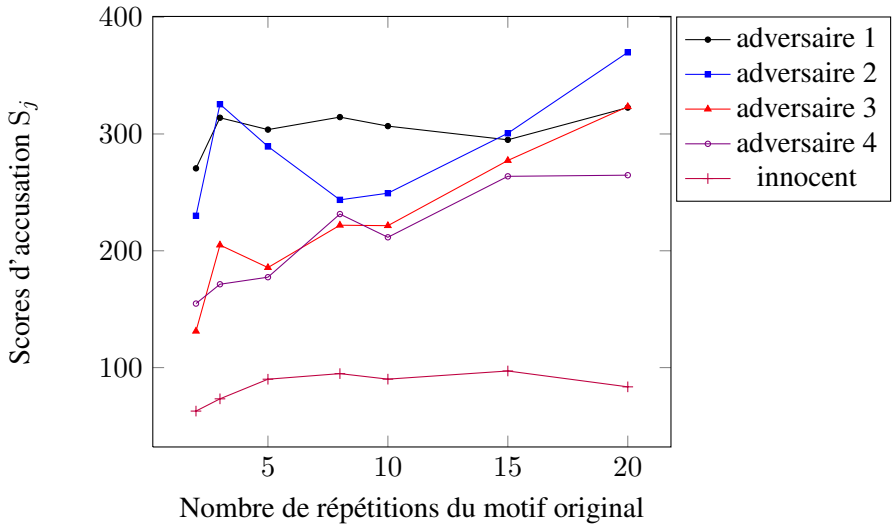


FIGURE 5.16 – Scores d'accusation S_j, $j \in [n]$ pour $c = 4$ adversaires et pour l'utilisateur ayant le plus grand score chez les innocents (au nombre de 16) en fonction du nombre de répétitions du motif original (code de Tardos + identifiant public) pour une qualité de réencodage de 800 Kbps après attaque par la coalition.

FIGURE 5.17 – Échantillon d'image pour un débit de réencodage de 300 Kbps.

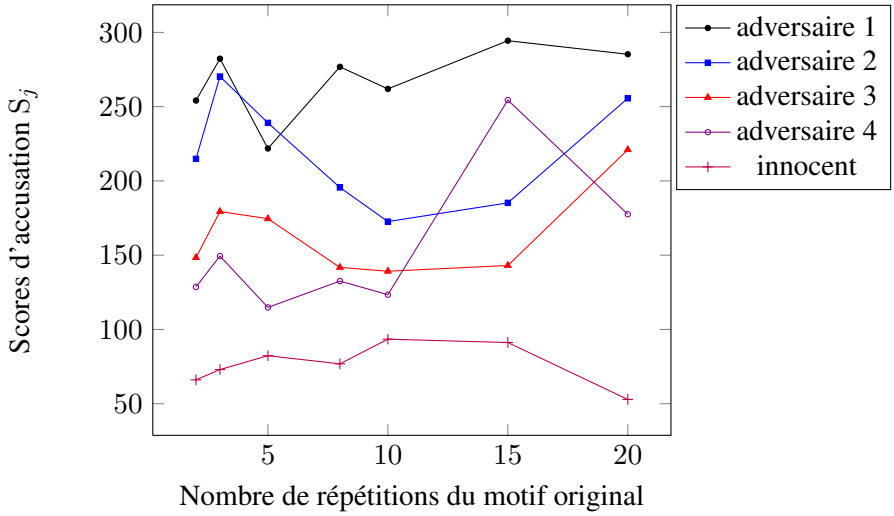


FIGURE 5.18 – Scores d'accusation S_j, $j \in [n]$ pour $c = 4$ adversaires et pour l'utilisateur ayant le plus grand score chez les innocents (au nombre de 16) en fonction du nombre de répétitions du motif original (code de Tardos + identifiant public) pour une qualité de réencodage de 300 Kbps après attaque par la coalition.

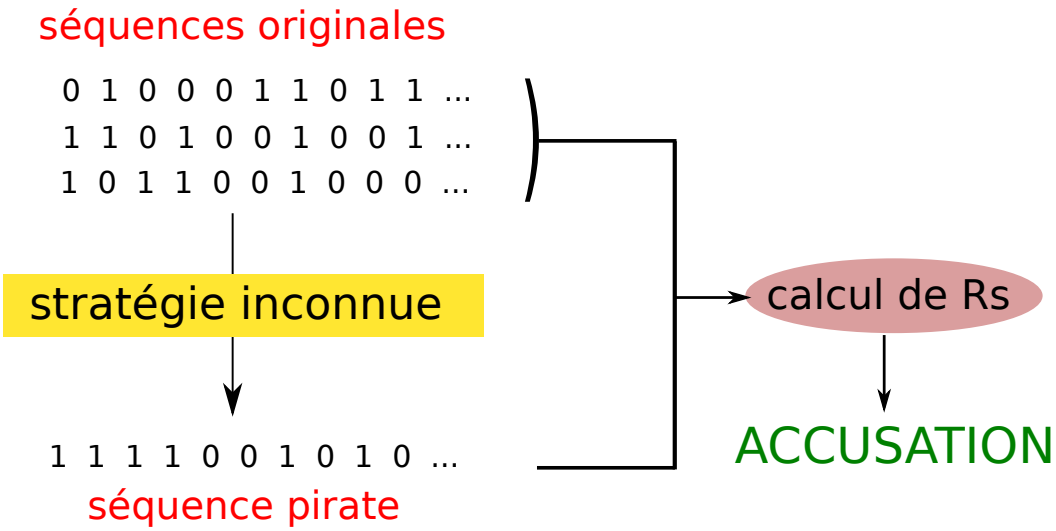


FIGURE 5.19 – Processus d'accusation de la coalition par le distributeur. L'accusation est effectuée par le biais du calcul du débit atteignable simple. Ce calcul ne dépend pas de la stratégie d'attaque employée par la coalition pour forger leur séquence pirate, le distributeur n'a a priori aucune information sur cette stratégie.

5.5 Conclusion

Dans ce chapitre, nous avons montré que les codes définis par Tardos couplés à des techniques de tatouage sûres par étalement de spectre sont utilisables et performants (robustesse et accusation d'une coalition) à des fins pratiques. Afin d'évaluer l'efficacité de l'attaque au pire cas, nous avons proposé d'estimer le débit atteignable directement à partir de l'observation du produit de la coalition. Cela nous a permis de nous assurer que la ϵ-WCA est efficace en pratique. De plus, les stratégies d'attaque au pires cas sont définies pour un alphabet binaire, il serait alors intéressant d'étudier les codes de traçage sur des alphabets plus grands (voir [14, 76]) permettant de

	$c = 3$	$c = 4$
$\theta_{\text{Aléatoire}}$	$5.58e-2$	$3.24e-2$
θ_{WCA}	$5.42e-2$	$3.12e-2$
$\theta_{0.3-\text{WCA}}$	$5.23e-2$	$2.96e-2$

TABLE 5.1 – Débits atteignables calculés pratiquement pour $c = 3, 4$ adversaires. Paramètres : $m = 10^4,\ N_o = 10^5$.

faire facilement varier la taille de notre sous-espace secret engendré par les porteuses en étalement de spectre.

Conclusion et perspectives

Conclusion

Dans cette thèse, nous avons abordé la problématique de la sécurité des œuvres numériques, problème qui a pris de plus en plus d'importance depuis le développement d'Internet et de réseaux d'échange et la dématérialisation des supports multimédia. Le tatouage numérique, technique permettant d'insérer de façon discrète et robuste une information dans un support, a été proposé comme une solution à la piraterie sur la propriété intellectuelle des œuvres numériques. Dans une première partie, nous nous sommes intéressés à la contrainte de sécurité en tatouage, définie comme l'incapacité pour des personnes non autorisées d'avoir accès au message transmis dans le signal de tatouage [44]. Nous nous sommes placés dans un cadre WOA, cadre dans lequel un adversaire a accès à plusieurs contenus tatoués et tente d'estimer la clé secrète. Nos contributions dans ce domaine ont été les suivantes :

1. Nous avons proposé deux nouvelles méthodes de tatouage sûres : le tatouage par la loi du χ^2 ainsi que le tatouage ρ circulaire ;
2. Nous avons amélioré la distorsion de schémas sûrs existants grâce aux solutions apportées par l'algorithme des Hongrois et les éléments de la théorie du transport. Ces méthodes, basées sur la minimisation de déplacements entre distributions, nous ont permis d'améliorer la distorsion d'insertion des tatouages ;
3. Nous avons montré que les méthodes de tatouage sûres par étalement de spectre (méthodes classiques et nouvelles méthodes) peuvent être utilisables en pratique en prenant comme exemple ici le tatouage d'images fixes.

Nous nous sommes ensuite, dans une seconde partie, intéressés à la problématique de l'estampillage, discipline permettant de tracer les redistributeurs de copies illégales d'œuvres numériques. Les codes d'estampillage que nous avons choisis de développer ici sont ceux développés par Gabor Tardos [79] car ils sont résistants aux attaques de coalition (groupement de plusieurs adversaires tentant de modifier leurs empreintes de traçage) pour une longueur de code optimale. Notons qu'il est nécessaire que ces codes traçants soient insérés de façon robuste dans les contenus multimédia, l'utilisation de méthodes de tatouage est alors pertinente. Nous avons ici contribué à l'étude de la contrainte de sécurité pour l'estampillage grâce aux points suivants :

1. Nous avons développé une stratégie d'attaque permettant, pour une coalition, de diminuer l'information mutuelle entre leur code de traçage et le code retrouvé sur la copie du contenu distribuée de façon illégale. Cette attaque, nommée ϵ-WCA, est optimale pour la coalition (minimisation du score d'accusation) et dépend de la sécurité du schéma de tatouage, paramétrée par ϵ ;
2. Nous avons quantifié le compromis existant entre les contraintes de sécurité et de robustesse pour l'estampillage. À robustesse égale, le distributeur aura intérêt à choisir la technique la plus sûre ;
3. Nous avons implanté les codes de Tardos dans un schéma de tatouage vidéo et montré qu'une méthode sûre (ici le CW) est utilisable en pratique et permet d'identifier le ou les responsables de la diffusion illégale d'une œuvre.

Perspectives

Le travail réalisé pendant cette thèse laisse entrevoir de nouvelles perspectives dans les domaines de l'estampillage et du tatouage numérique. En effet :

i) Le tatouage ρ circulaire, méthode hybride entre la clé-sécurité et la non-sécurité est finalement peu robuste et moins adapté au contexte de l'estampillage comme nous l'avons vu dans la section 4.4.2. La recherche de méthodes sûres et robustes reste toujours un défi pour la communauté des tatoueurs. Nous pouvons prendre l'exemple du *Scalar-Costa-Scheme* (SCS) [28] dont une version étendue permet un tatouage sûr applicable dans le contexte WOA [5] ;

ii) Les stratégies d'attaque qu'une coalition utilise en estampillage sont définies ici pour un alphabet binaire. Cette contrainte nous a posé quelques problèmes lors de notre implantation de la méthode CW pour le traçage vidéo. L'utilisation de stratégies pour un alphabet de taille $q > 2$ [14] serait beaucoup plus flexible et correspondrait plus à la vision que nous nous faisons d'une attaque de coalition ;

iii) Nous savons que l'insertion de codes par tatouage n'est sûre en WOA que si les bits d'informations sont équiprobables et indépendants, ce qui empêche l'utilisation de codes correcteurs d'erreurs. Cependant, dans le dernier chapitre de ce manuscrit, nous avons utilisé une répétition du code traçant (avec permutations) comme code correcteur. Deux questions viennent alors à l'esprit : quel est l'impact d'un code correcteur d'erreur sur la sécurité ? Existe-t-il des codes correcteurs plus sûrs que d'autres ? Certain codes comme les codes convolutifs ont montré des failles de sécurités potentielles [8], mais une étude systématique de la sécurité des codes correcteurs fait actuellement défaut ;

iv) En cryptographie, la sécurité est souvent calculatoire. C'est-à-dire que l'on considère qu'un système de communications est assez sûr si le nombre d'opérations mathématiques nécessaires pour le mettre à mal est supérieur à la puissance actuelle des ordinateurs (exemple du système RSA en introduction). La question que nous nous posons actuellement est la suivante : comment mesurer la sécurité calculatoire des méthodes de tatouage actuelles ? Quelle doit être la taille réelle de la clé secrète ? Des travaux comparant la sécurité en tatouage et en cryptographie sont récemment apparus [27] et sont à approfondir.

Nous concluons ce manuscrit sur un cadre théorique en tatouage qui fait cruellement défaut dans ce domaine. Il s'agit de la question de la capacité : quelle est la capacité sûre en tatouage ? C'est-à-dire quelle est la quantité d'information sûre maximale que l'on peut transmettre dans un canal donné et quels sont les codes qui permettent de se rapprocher de cette capacité ?

Bibliographie

[1] Cinavia. http://www.cinavia.com/languages/french/pages/technology.html.

[2] Gnu scientific library. http://www.gnu.org/software/gsl/.

[3] After the blast : fingerprinting bombs. *Science 80*, 1,6:96–99, sept. 1980.

[4] G. BAILLY, V. ATTINA, C. BARAS, P. BAS, S. BAUDRY, D. BEAUTEMPS, R. BRUN, J.-M. CHASSERY, F. DAVOINE, F. ELISEI, G. GIBERT, L. GIRIN, D. GRISON, J.-P. LÉONI, J. LIÉNARD, N. MOREAU et P. NGUYEN : Artus : synthèse et tatouage audiovisuel des mouvements d'un personnage animé virtuel pour l'accessibilité d'émissions télévisuelles aux téléspectateurs sourds comprenant la langue française parlée complétée. *Handicap*, 2006.

[5] P. BAS : Soft-SCS : improving the security and robustness of the Scalar-Costa-Scheme by optimal distribution matching. *In Proc. of the thirteenth Int. Workshop on Information Hiding*, 2011.

[6] P. BAS et F. CAYRE : Achieving Subspace or Key Security for WOA using Natural or Circular Watermarking. *In MM&Sec '06 : Proceedings of the 8th workshop on Multimedia and security*, p. 80–88, New York, NY, USA, 2006. ACM.

[7] P. BAS et F. CAYRE : Natural Watermarking : A Secure Spread Spectrum Technique for WOA. *In Information Hiding*, vol. 4437 de *Lecture Notes in Computer Science*, p. 1–14. Springer Berlin / Heidelberg, 2007.

[8] P. BAS et G. DOËRR : Practical Security Analysis of Dirty Paper Trellis Watermarking. *In Information Hiding*, vol. 4567 de *Lecture Notes in Computer Science*, p. 174–188. Springer Berlin / Heidelberg, 2007.

[9] P. BAS et T. FURON : Break Our Watermarking System 2nd edition, 2007. http://bows2.gipsa-lab.inpg.fr.

[10] P. BAS et J. HURRI : Vulnerability of DM watermarking of non-iid host signals to attacks utilising the statistics of independent components. *Information Security, IEE Proceedings*, 153(3):127 –139, sept. 2006.

[11] P. BAS et A. WESTFELD : Two key estimation techniques for the broken arrows watermarking scheme. *In MM&Sec '09 : Proceedings of the 11th ACM workshop on Multimedia and security*, p. 1–8, New York, NY, USA, 2009. ACM.

[12] E. BERLEKAMP : *Nonbinary BCH decoding*. University of North Carolina. Dept. of Statistics, 1966.

[13] G. R. BLAKLEY, C. MEADOWS et G. B. PURDY : Fingerprinting long forgiving messages. *In Proc. of Crypto'85*, p. 180–189. Springer Berlin / Heidelberg, 1985.

[14] D. BOESTEN et B. SKORIC : Asymptotic fingerprinting capacity for non-binary alphabets. *Arxiv preprint arXiv :1102.0445*, 2011.

[15] D. BONEH et J. SHAW : Collusion-secure fingerprinting for digital data. *Information Theory, IEEE Transactions on*, 44(5):1897 –1905, sep 1998.

[16] R. P. BRENT : An algorithm with guaranteed convergence for finding a zero of a function. *Computer Journal*, 14:422–425, 1971.

[17] J. C. P. BUS et T. J. DEKKER : Two Efficient Algorithms with Guaranteed Convergence for Finding a Zero of a Function. *ACM Trans. Math. Softw.*, 1(4):330–345, 1975.

[18] C. CACHIN : An information-theoretic model for steganography. *In Information Hiding*, p. 306–318. Springer, 1998.

[19] F. CAYRE et P. BAS : Kerckhoffs-Based Embedding Security Classes for WOA Data Hiding. *Information Forensics and Security, IEEE Transactions on*, 3(1):1–15, mars 2008.

[20] F. CAYRE, C. FONTAINE et T. FURON : Watermarking Security : Theory and Practice. *Signal Processing, IEEE Transactions on*, 53(10):3976–3987, oct. 2005.

[21] A. CHARPENTIER, C. FONTAINE et T. FURON : Décodage EM du code de Tardos pour le fingerprinting= EM decoding of Tardos fingerprinting code. *TS. Traitement du signal*, 27(2):127–146, 2010.

[22] B. CHOR, A. FIAT, M. NAOR et B. PINKAS : Tracing traitors. *In Proc. of Crypto'94*, vol. 839 de *Lecture Notes in Computer Science*, p. 257–270. Springer Berlin / Heidelberg, 1994.

[23] A. COHEN, I. DAUBECHIES et J.-C. FEAUVEAU : *Biorthogonal bases of compactly supported wavelets*, vol. 45 de *Communications on Pure and Applied Mathematics*. 1992.

[24] J. H. CONWAY, R. H. HARDIN et N. J. A. SLOANE : Packing lines, planes, etc. : packings in grassmanian spaces. *In Experimental Mathematics*, vol. 5(2), p. 139–159, 1996.

[25] I. COX, J. KILIAN, F. LEIGHTON et T. SHAMOON : Secure spread spectrum watermarking for multimedia. *Image Processing, IEEE Transactions on*, 6(12):1673–1687, déc. 1997.

[26] I. COX, M. MILLER et A. MCKELLIPS : Watermarking as communications with side information. *Proceedings of the IEEE*, 87(7):1127–1141, 2002.

[27] I. J. COX, G. DOËRR et T. FURON : *Watermarking Is Not Cryptography*, vol. 4283 de *Lecture Notes in Computer Science : Digital Watermarking*. Springer Berlin / Heidelberg, 2006.

[28] J. EGGERS, R. BAUML, R. TZSCHOPPE et B. GIROD : Scalar costa scheme for information embedding. *Signal Processing, IEEE Transactions on*, 51(4):1003–1019, 2003.

[29] R. A. FISHER et F. YATES : *Statistical tables for biological, agricultural and medical research (3rd ed.)*. London : Olivier & Boyd, 1948.

[30] J. FRIDRICH : *Steganography in Digital Media : Principles, Algorithms, and Applications*. Cambridge University Press, New York, NY, USA, 1st édn, 2009.

[31] T. FURON : Le traçage de traîtres. *Symposium sur la Sécurité des Technologies de l'Information et des Communications*, 2009.

[32] T. FURON et P. BAS : Broken Arrows. *EURASIP Journal of Information Security*, (ID 597040), 2008.

[33] T. FURON et P. DUHAMEL : An asymmetric public detection watermarking technique. *In* A. PFITZMANN, éd. : *Proc. of the third Int. Workshop on Information Hiding*, p. 88–100, Dresden, Germany, sep 1999. Springer Verlag.

[34] T. FURON, A. GUYADER et F. CÉROU : On the design and optimisation of Tardos probabilistic fingerprinting codes. *In Information Hiding*, vol. 5284 de *Lecture Notes in Computer Science*, p. 341–356. Springer Berlin / Heidelberg, 2008.

[35] T. FURON et L. PÉREZ-FREIRE : Worst case attacks against binary probabilistic traitor tracing codes. *IEEE Transactions on Information Forensics and Security*, 2009.

[36] T. FURON, L. PÉREZ-FREIRE, A. GUYADER et F. CÉROU : Estimating the minimal length of Tardos code. *In Information Hiding*, vol. 5806 de *Lecture Notes in Computer Science*, p. 176–190. Springer Berlin / Heidelberg, 2009.

[37] F. HARTUNG et M. KUTTER : Multimedia watermarking techniques. *Proceedings of the IEEE*, 87(7):1079–1107, July 1999.

[38] I. D. HILL et M. C. PIKE : Algorithm 299. *ACM TOMS*, p. 185, juin 1985.

[39] Y. W. HUANG et P. MOULIN : Saddle-point solution of the fingerprinting capacity game under the marking assumption. *In Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, p. 2256–2260, 2009.

[40] A. HYVÄRINEN : Fast and robust fixed-point algorithms for independent component analysis. *Neural Networks, IEEE Transactions on*, 10(3):626–634, mai 1999.

[41] A. HYVÄRINEN, J. KARHUNEN et E. OJA : *Independent Component Analysis*. John Wiley & Sons, 2001.

[42] D. IBBETSON : Algorithm 209. *Collected Algorithms of the CACM*, p. 616, 1963.

[43] H. JÉGOU, V. CHAPPELIER et F. CAYRE : Libit : Information theory and signal processing library. http://libit.sourceforge.net/.

[44] T. KALKER : Considerations on watermarking security. *In Multimedia Signal Processing, 2001 IEEE Fourth Workshop on*, p. 201–206, 2001.

[45] L. KANTOROVITCH : On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[46] L. KANTOROVITCH : On a problem of Monge (in Russian). *Uspekhi Math. Nauk.*, 3:225–226, 1948.

[47] A. KERCKHOFFS : La Cryptographie militaire. *Journal des Sciences militaires*, IX:5–38, jan. 1883.

[48] M. KNOTT et C. S. SMITH : On the optimal mapping of distributions. *Journal of Optimization Theory and Applications*, 43:39–49, mai 1984.

[49] D. E. KNUTH : *The Art of Computer Programming volume 2 : Seminumerical algorithms*. Reading, MA : Addison-Wesley, 1969.

[50] H. W. KUHN : The Hungarian Method of Solving the Assignment Problem. *Naval Res. Logistics Quart.*, 2:83–97, 1955.

[51] H. MALVAR et D. FLORÊNCIO : Improved spread spectrum : a new modulation technique for robust watermarking. *Signal Processing, IEEE Transactions on*, 51(4):898–905, avr. 2003.

[52] G. MARSAGLIA et W. TSANG : The ziggurat method for generating random variables. *Journal of Statistical Software*, 5(8):1–7, 2000.

[53] J. MASSEY : Shift-register synthesis and BCH decoding. *Information Theory, IEEE Transactions on*, 15(1):122–127, 1969.

[54] B. MATHON, P. BAS et F. CAYRE : Practical Performance Analysis of Secure Modulations for WOA Spread-Spectrum based Image Watermarking. *In MM&Sec '07 : Proceedings of the 9th workshop on Multimedia & security*, p. 237–244, New York, NY, USA, 2007. ACM.

[55] B. MATHON, P. BAS, F. CAYRE et B. MACQ : Comparison of Secure Spread-Spectrum Modulations Applied to Still Image Watermarking. *Annals of Telecommunications*, 64(11-12):801–813, déc. 2009.

[56] B. MATHON, P. BAS, F. CAYRE et B. MACQ : Optimization of Natural Watermarking using Transportation Theory. *In MM&Sec '09 : Proceedings of the 11th ACM workshop on Multimedia and security*, p. 33–38, New York, NY, USA, 2009. ACM.

[57] B. MATHON, P. BAS, F. CAYRE et B. MACQ : Considering security and robustness constraints for watermark-based tardos fingerprinting. *In IEEE SPS Multimedia Signal Processing Conference (MMSP 2010)*, Saint Malo, France, oct. 2010.

[58] B. MATHON, P. BAS, F. CAYRE et B. MACQ : Security and robustness constraints for spread-spectrum tardos fingerprinting. *In IEEE Workshop on Information Forensics and Security (WIFS 2010)*, Seattle, US, oct. 2010.

[59] B. MATHON, P. BAS, F. CAYRE et B. MACQ : Théorie du transport appliquée au tatouage sûr d'images naturelles. *In GRETSI Symposium on Signal and Image Processing (soumis à)*, Bordeaux, France, 2011.

[60] B. MATHON, P. BAS, F. CAYRE et F. PÉREZ-GONZÁLEZ : Distortion Optimization of Model-Based Secure Embedding Schemes for Data-Hiding. *In Information Hiding*, vol. 5284 de *Lecture Notes in Computer Science*, p. 325–340. Springer Berlin / Heidelberg, 2008.

[61] B. MATHON, F. CAYRE et P. BAS : Techniques Sûres de Tatouage pour l'Image. *In Proc. CORESA2007*, Montpellier, France, nov. 2007.

[62] M. MATSUMOTO et T. NISHIMURA : Mersenne twister : a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30, 1998.

[63] J. K. MILEN : *Discussion*. Foundations of Secure Computations. R. A. DeMillo et. al., Academic Press, 1978.

[64] G. MONGE : Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, p. 666–704, 1781.

[65] P. MOULIN : Universal fingerprinting : Capacity and random-coding exponents. *In Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, p. 220–224, juil. 2008.

[66] P. MOULIN et A. BRIASSOULI : A stochastic QIM algorithm for robust, undetectable image watermarking. *In ICIP*, p. 1173–1176, 2004.

[67] J. A. NELDER et R. MEAD : A Simplex Method for Function Minimization. *Computer Journal*, 7(4):308–313, 1965.

[68] C. PEIKERT, A. SHELAT et A. SMITH : Lower bounds for collusion-secure fingerprinting. *In SODA '03 : Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, p. 472–479, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics.

[69] L. PÉREZ-FREIRE et F. PÉREZ-GONZÁLEZ : Spread-Spectrum Watermarking Security. *Information Forensics and Security, IEEE Transactions on*, 4(1):2–24, mars 2009.

[70] L. PÉREZ-FREIRE, F. PÉREZ-GONZÁLEZ et T. FURON : On achievable security levels for lattice data hiding in the Known Message Attack scenario. *In Proceedings of the 8th workshop on Multimedia and security*, p. 68–79. ACM, 2006.

[71] A. PIVA, M. BARNI, F. BARTOLINI et V. CAPPELLINI : Dct-based watermark recovering without resorting to the uncorrupted original image. *In Image Processing, 1997. Proceedings., International Conference on*, vol. 1, p. 520 –523 vol.1, 26-29 1997.

[72] S. RACHEV et L. RÜSCHENDORF : *Mass Transportation Problems : Theory*. Springer, 1998.

[73] R. RIVEST, A. SHAMIR et L. ADLEMAN : A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[74] P. SALLEE : *Model-Based Steganography*, vol. 2939 de *Lecture Notes in Computer Science : Digital Watermarking*. Springer Berlin / Heidelberg, 2004.

[75] C. E. SHANNON : A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, juil. 1948.

[76] A. SIMONE et B. SKORIC : Asymptotically false-positive-maximizing attack on non-binary Tardos codes. *Arxiv preprint arXiv :1102.0451*, 2011.

[77] B. SKORIC, T. VLADIMIROVA, M. CELIK et J. TALSTRA : Tardos Fingerprinting is Better Than We Thought. *Information Theory, IEEE Transactions on*, 54(8):3663–3676, août 2008.

[78] K. TANAKA, Y. NAKAMURA et K. MATSUI : Embedding secret information into a dithered multi-level image. *In Military Communications Conference, 1990. MILCOM'90, Conference Record, A New Era. 1990 IEEE*, p. 216–220. IEEE, 1990.

[79] G. TARDOS : Optimal probabilistic fingerprint codes. *J. ACM*, 55(2):1–24, 2008.

[80] W. TRAPPE, M. WU, Z. WANG et K. LIU : Anti-collusion fingerprinting for multimedia. *Signal Processing, IEEE Transactions on*, 51(4):1069–1087, avr. 2003.

[81] C. VILLANI : *Topics in optimal transportation*. Amer Mathematical Society, 2003.

[82] N. R. WAGNER : Fingerprinting. *In SP '83 : Proceedings of the 1983 IEEE Symposium on Security and Privacy*, p. 18, Washington, DC, USA, 1983. IEEE Computer Society.

[83] M. WU et Z. J. WANG : Collusion resistance of multimedia fingerprinting using orthogonal modulation. *In IEEE Trans. on Image Proc*, p. 804–821, 2005.

[84] F. XIE, T. FURON et C. FONTAINE : On-off keying modulation and Tardos fingerprinting. *In MM&Sec '08 : Proceedings of the 10th ACM workshop on Multimedia and security*, p. 101–106, New York, NY, USA, 2008. ACM.

Développement de méthodes de tatouage sûres pour le traçage de contenus multimédia

Résumé

Dans cette thèse, nous étudions dans une première partie l'impact de la contrainte de sécurité en tatouage. Dans le contexte WOA (*Watermarked contents Only Attack*), un adversaire possède plusieurs contenus tatoués et cherche à estimer la clé secrète d'insertion afin d'accéder aux messages cachés. Une nouvelle manière de tatouer en étalement de spectre est présentée ici. Celle-ci est basée sur la construction de distributions circulaires dans le sous-espace secret de tatouage. Cette technique permet de minimiser la distorsion en moyenne provoquée par l'ajout de la marque dans le contexte WOA en utilisant l'algorithme d'optimisation des Hongrois et la théorie du transport. Nous vérifions ensuite qu'un tatouage sûr est utilisable en pratique en prenant comme exemple le tatouage d'images naturelles.

Dans une seconde partie, nous nous intéressons au cadre de l'estampillage d'œuvres numériques permettant de tracer les redistributeurs de copies illégales. Les codes traçants utilisés sont ceux proposés par Gabor Tardos et sont résistants aux attaques de coalition, c'est-à-dire au groupement d'adversaires mettant en commun leurs contenus numériques afin de forger une version pirate. Puisque les techniques de tatouage permettent l'insertion de codes traçants dans un contenu numérique, nous avons conçu une attaque "au pire cas" qui dépend du niveau de sécurité et qui permet, pour les adversaires, de baisser leur accusation. Nous montrons que pour le cas particulier de l'estampillage un tatouage sûr sera plus efficace qu'un tatouage non-sûr (à robustesse équivalente). Finalement, une implantation des codes traçants dans un contenu vidéo utilisant des méthodes sûres par étalement de spectre est proposée. Nous montrons alors l'efficacité de l'accusation des adversaires dans ce cadre pratique.

Mots-clés : Tatouage numérique ; Étalement de spectre ; Sécurité ; Transport Optimal ; Méthode des Hongrois ; Estampillage ; Multimédia.

Secure watermarking methods for fingerprinting of multimedia contents

Abstract

In this thesis, we first study the constraint of security in watermarking. In the WOA (Watermarked contents Only Attack) framework, an adversary owns several marked contents and try to estimate the secret key used for embedding in order to have access to the hidden messages. We present a new mean for spread-spectrum watermarking based on circular distributions in the private watermarking subspace. Thanks to this technique, we are able to minimise the distortion (on expectation) caused by the watermark in the WOA framework using the Hungarian optimisation method and the transportation theory. Then, we show that secure watermarking can be used in practical works with the example of still image watermarking.

In the second part, we are interested about the problem of active fingerprinting which allows to trace re-distributors of illegal copies of a numerical content. The codes we use here are the ones proposed by Gabor Tardos. These codes are resistant against collusion attacks e.g. a group of malicious users who forges a new content by mixing their copies. Since watermarking techniques allow the embedding of these codes in numerical contents, a new worst case attack taking into account the security level of the watermarking system is proposed to reduce the accusation rate of the coalition. We show that secure watermarking is more efficient that insecure one (with similar robustness) for fingerprinting application. Finally, traitor tracing codes are implemented on video sequences by using spread-spectrum techniques in order to demonstrate that the accusation of adversaries is practically possible.

Keywords: Watermarking; Spread-spectrum; Security; Optimal transport map; Hungarian method; Active Fingerprinting; Multimedia.

GIPSA-Lab
Dpt. Images et Signaux
ENSE3, Domaine Universitaire, BP 46
38402 St-Martin-d'Hères Cedex, France

Laboratoire TELE
Bâtiment Stévin
Place du Levant 2
1348 Louvain-la-Neuve, Belgium