Habilitation à Diriger des Recherches

Spécialité :

INFORMATIQUE

présentée par

Caroline FONTAINE

Université de Bretagne Occidentale, école doctorale SICMA

Sujet :

ASSURER LA SÉCURITÉ DES CONTENUS MULTIMÉDIA, DE LEUR CRÉATION À LEUR DIFFUSION

Présentée le 28 novembre 2011 devant le jury composé de :

`Rapporteurs / Referees :`

M. Philippe GABORIT	Université de Limoges, XLIM
M. Grigory KABATYANSKIY	Khalifa University of Science, Technology and Research, Abu Dhabi, on leave from Institute for Information Transmission Problems (IPIT), Russian Academy of Sciences, Moscow
M. Boris ŠKORIĆ	Technische Universiteit Eindhoven, Department of Mathematics and Computer Science

`Examinateurs / Examiners :`

M. Gilles BUREL	Université de Bretagne Occidentale, Lab-STICC
M. David GROSS-AMBLARD	Université Rennes 1, IRISA
M. Johannes HUISMAN	Université de Bretagne Occidentale, Laboratoire de Mathématiques

なぎさふりかえる

我が　足跡も　無く

尾崎　放哉

Remerciements

Le métier de chercheur est très prenant, exaltant quand une idée émerge, déconcertant et parfois décourageant lorsque l'on piétine. Dans ces moments d'exaltation comme de doute, il est précieux d'être bien entouré. Beaucoup de collègues et d'amis m'ont accompagnée de près ou de loin dans mon cheminement, et je les en remercie.

Mes premiers remerciements vont aux membres du jury, pour avoir accepté de relire ce manuscrit, pour l'intérêt qu'ils y ont porté, leurs questions, et leurs remarques. Je remercie tout particulièrement les rapporteurs pour le temps qu'ils ont consacré à ce travail, leur relecture attentive, et les nombreuses discussions que nous avons eues avant comme après la soutenance. Je commencerai par ceux dont le port d'attache est le plus éloigné.

Grigory Kabatyanskiy is a one of the most impressive researcher I know in coding theory. It is a great honour for me he accepted to review this dissertation. We worked independently on common topics, as Syndrome Coding and Anti-collusion codes. His wide and impressive knowledge of coding and information theory, and his external and rigorous look on my work have been really precious to me. I really appreciated our discussions.

Boris Škorić is "the" worldwide expert of q-ary Tardos codes, and I have been really happy and honoured he accepted to review this work. Even if we both work on Tardos codes and know our respective works, we never met before 2011. These last months we had the opportunity to meet and discuss, and it was really stimulating.

Philippe Gaborit et moi nous côtoyons depuis longtemps dans la communauté « codes/-crypto » française, sans pour autant avoir travaillé ensemble jusqu'à aujourd'hui. Depuis quelques temps, nous avons des thèmes de recherche communs, et je suis très heureuse qu'il ait accepté d'être rapporteur pour cette habilitation. Son parcours, différent du mien, donne un regard enrichissant sur mon travail.

David Gross-Amblard est l'un des rares chercheurs au monde à s'être penché sur le tatouage de bases de données. Car si la majeur partie des travaux en tatouage ou *fingerprinting* porte sur les documents multimédia, il y a bien d'autres types de documents à considérer, et qui n'ont pour l'instant reçu que peu d'attention. Nous nous suivons de loin depuis de longues années, et je suis vraiment heureuse qu'il ait accepté de faire partie de ce jury.

Je remercie également chaleureusement Gilles Burel et Johannes Huisman d'avoir accepté de faire partie de mon jury. Certaines de leurs thématiques de recherche, bien que différentes des miennes, traitent d'objets similaires comme les codes correcteurs, et j'ai été très sensible à l'intérêt qu'ils ont porté à mon travail, ainsi qu'aux questions qu'ils ont soulevées.

Les travaux présentés dans ce mémoire couvrent quinze ans de recherches, menées à l'INRIA Rocquencourt, au LRI à Orsay, au LIFL à Lille, à l'IRISA et à l'INRIA à Rennes, et maintenant au Lab-STICC et à Télécom Bretagne à Brest. Ils doivent beaucoup à mes co-auteurs, aux collègues et étudiants avec qui j'ai travaillé, à ceux que j'ai côtoyés dans les projets auxquels j'ai participés ou que j'ai coordonnés. Merci en particulier à Pascale Charpin, Anne Canteaut, Claude Carlet et Eric Filiol pour nos travaux sur les fonctions

booléennes ; à Daniel Augot pour nos travaux sur le tatouage, les protocoles cryptographiques, la sécurisation des réseaux ad hoc, et l'utilisation des codes en stégano ; à Eric Wegrzynowski, François Recher, et Vincent Bénony pour nos travaux sur les NLFSRs ; à Fabien Petitcolas, Frédéric Raynal, Teddy Furon, François Cayre, Fabien Galand, Claude Delpha, Ana Charpentier et Fuchun Xie pour nos travaux sur les techniques de tatouage et les codes anti-collusion ; à Patrick Bas pour nos discussions et le projet ESTIVALE dont j'ai vraiment gardé un excellent souvenir ; à Morgan Barbier pour nos travaux sur l'utilisation des codes en stéganographie ; à Guy Gogniat, Renaud Sirdey et Simon Fau pour les travaux en cours sur le chiffrement homomorphe.

Au-delà de mes co-auteurs, j'y ai aussi rencontré de nombreux chercheurs, enseignants-chercheurs, PRAG, ingénieurs, techniciens et personnels administratifs. J'ai partagé avec nombre d'entre eux d'excellents moments. Tous ne seront pas cités ici, mais ils se reconnaîtront et je les en remercie sincèrement. Je remercie également chacun de ces laboratoires et équipes pour la confiance qu'ils m'ont témoignée. Chacun à leur manière ils ont contribué au cheminement qui a abouti à ce travail.

```
#ifdef verbose
```

J'ai eu la chance de rencontrer pendant mes cours de DEA des chercheurs en cryptographie et codes si passionnés que j'ai eu envie de poursuivre dans ces disciplines. Je dois mes premiers pas en cryptographie à Gilles Brassard et Jean-Jacques Quisquater, dont les cours étaient particulièrement exaltants et ludiques. Ce fut une révélation.

Mes premiers pas en théorie des codes ont été guidés par Pascale Charpin et Daniel Augot, que j'ai retrouvés lors de mon stage, puis de ma thèse, au projet CODES (maintenant SECRET) de l'INRIA Rocquencourt. Cette équipe garde une place à part dans mon cœur pour l'exaltation, l'ouverture d'esprit et la rigueur qui y règnent, et pour l'émulation scientifique qui en découle. Principalement en codes, cryptographie, et bien sûr à leur interface, avec une curiosité pour d'autres technologies comme le tatouage où la stéganographie. C'est là que j'ai commencé à travailler pendant ma thèse à la frontière de plusieurs domaines. Un merci tout particulier à Pascale Charpin pour sa rigueur et son écoute, Anne Canteaut pour le soin qu'elle met en tout, et tous les bons moments que nous avons partagés, Daniel Augot pour sa spontanéité, son émerveillement, et son enthousiasme. Merci Daniel pour tous nos brainstorming et nos discussions, sur les codes, la crypto, les réseaux, la stégano ... nous avons travaillé ensemble sur beaucoup de sujets et toujours avec grand plaisir. Un grand merci également à Nicolas Sendrier pour sa patience et son optimisme, Jean-Pierre Tillich pour son exigence, Christelle Guiziou pour sa gentillesse et son efficacité. Merci aussi à Claude Carlet, Françoise Lévy-dit-Véhel, Gaétan Haché, Grégoire Bommier, Eric Filiol, Pierre Loidreau, Marion Videau, Marine Minier, Cédric Lauradoux, avec qui j'ai partagé d'excellents moment, et qui comme moi sont partis vers de nouveaux horizons. En écrivant ces lignes, j'ai également une pensée pour le projet ALGO, toujours accueillant, merci en particulier à Philippe Flajolet qui nous a malheureusement quittés cette année, ainsi qu'à Bruno Salvy, Frédéric Chyzak, François Morain, et Virginie Collette.

Mon court passage au LRI dans l'équipe ALGO m'a permis d'étendre mes connaissances en codes et théorie de l'information, et bien sûr algorithmique en général. J'en garde un souvenir ému d'une année de transition où je me suis un peu cherchée. Merci en particulier à Jean-Pierre Tillich, Jean-Paul Allouche, Frédéric Magniez, Sophie Laplante, Cristina Bazgan, Johanne Cohen, David Gross-Amblard, Julien Stern, pour m'avoir accompagnée dans cette transition entre la thèse et un poste permanent. C'était pour beaucoup d'entre eux aussi une période de transition et c'est toujours un plaisir de les recroiser quand l'occasion s'en présente.

Lors de cette année de transition j'ai également été accueillie en filigrane à Télécom ParisTech aux départements TSI et INFRES. Je remercie particulièrement Henri Maître pour l'accueil chaleureux qu'il m'a réservé alors même que nous ne nous connaissions pas, pour la confiance et le soutien qu'il m'a témoignés. Mes séjours à TSI m'ont permis de parfaire mes connaissances en tatouage et partager mes connaissances en cryptographie et codes, et je remercie toute l'équipe pour son accueil. Le "baby seminar" que nous y avons mis en place avec Gouenou Coatrieux et Séverine Baudry a également été un lieu d'échanges particulièrement stimulant.

Mon premier poste permanent m'a emmenée vers le Nord, à Lille. Je n'y connaissais personne, et je me souviens avec émotion de l'accueil chaleureux que mes nouveaux collègues du LIFL m'ont réservé. Je remercie avant tout Jean-Marc Geib, Sophie Tison, Vincent Cordonnier et David Simplot-Ryl pour la confiance qu'ils m'ont témoignée en tant que membres de la direction et chefs d'équipe. Merci aussi à Jean-Paul Delahaye pour m'avoir mis le pied à l'étrier, et pour sa passion des mathématiques et de la logique. Merci bien sûr à la bande de RD2P pour nos discussions OS, réseau, et sécurité en général, pour leur bonne humeur, les parties de jeu vidéo, les « japonaiseries ». Merci notamment à Gilles Grimaud, Sébastien Jean, Julien Iguchi-Cartigny, Alexandre Courbot, Damien Deville, Michaël Hauspie, Vincent Bénony, Kévin Marquet, Jean-Jacques Vandewalle, Jean Carle, Farid Naït-Abdesselam.

Durant ces six années passées à Lille, mon activité de recherche s'est ainsi ouverte à de nouvelles thématiques, comme les contraintes liées à l'embarqué et aux réseaux ad hoc. Le projet SERAC de l'ACISI que j'ai coordonné à cette époque a été pour moi une expérience extrêmement enrichissante par la diversité des compétences que nous avons tous réussi à partager. Un grand merci à tous les partenaires, cryptographes, experts des réseaux et des méthodes formelles de validation pour avoir eu la patience de partager vos compétences avec d'autres chercheurs aux profils si différents. En parallèle, mon activité s'est renforcée sur la cryptographie, les codes, et la dissimulation d'information, grâce aux différents projets et collaborations menés durant cette période, et grâce au groupe de travail CRYL que nous avons formé localement avec Eric Wegrzynowski, François Recher, Vincent Bénony et Julio Cesar Hernandez Castro. Ce groupe m'a laissé d'excellents souvenirs scientifiques et humains, dont l'organisation des journées C2 à Aussois en 2005 a constitué un point d'orgue. Quel souvenir, un beau travail d'équipe, et des journées vraiment réussies !

Un immense merci en particulier à Eric Wegrzynowski pour son amitié, sa confiance, sa curiosité jamais rassasiée, son enthousiasme, sa passion pour des codes, la crypto, et l'enseignement. J'ai beaucoup appris à ses côtés en travaillant sur le cours de codage en L1, et sur le cours de cryptographie que nous avons monté avec Michel Petitot en M1. Je lui dois beaucoup. Quels souvenirs que ces cours de crypto, et des challenges qui les ont animés. J'y ai retrouvé comme enseignante l'exaltation que j'avais connue en DEA comme étudiante.

```
UADNA ITLDH BXOUX BRAKB ANTEK XACYC LRDHX BOXKA UADNA HODIE AGFYP GRIDR UADNA
AXBTH XKJVX MBOUD XFPYM BGXIH FIUGF YUAOI KRCXD CHQXB OUXFA UADNA MKUTH OEIKG
UXIGR COUYD HPBIA THUTH OEIKC GFEXI CEUDI BTBIT UADNA AHXUT OGFIK GXDCO UBOBX
KAHED XUKHF YPCGA FDUBI HDXKX KHIKB UADNA XNQJK BHNDH CULIC FOPGY PCFUH XIWUM
IBNQI HERUG FITPY FDTXI UADNA KVTED
```

Au gré des enseignements, des séminaires, des commissions et des déménagements de bureaux, j'ai également apprécié travailler et discuter avec Pierre Boulet, Hélène Touzet, Jean-Marc Talbot, Laurence Duchien, ainsi qu'avec l'ensemble de l'équipe Calcul Formel et la joyeuse troupe du 3ème étage du M3.

Un merci tout particulier à mes amis du M3 et du M1 : Marie-Paule Quétu avec qui je partage tant de choses que je n'en reviens toujours pas, Laetitia Jourdan et Mathieu Vermeulen (on joue quand ?), Léopold Weinberg, Alexandre Sédoglavic, Olivier Perriquet, et

Gwenaëlle Castellan-Guérin (sais-tu Gwenaëlle que Les Mauvaises Langues continues d'accompagner chacun de mes déplacements en train ?), pour leur amitié, si précieuse durant ces années et encore aujourd'hui (à quand un Wanted-Breizh ?). En dehors du laboratoire, ma vie lilloise a été agrémentée de nombreuses activités, et c'est toujours avec une certaine nostalgie que je repense à l'atelier photo de l'USTL ainsi qu'aux cours de japonais de l'USTL et de l'association Japon et Culture. Ces années n'auraient pas été les mêmes sans Philippe Timmermann et Antoine Petitprez qui m'ont aidée à développer mon regard, Frédéric Lecœur, Sarah Guinand, et Marie-Paule Quétu avec qui j'ai partagé des expériences artistiques riches et variées, Elisabeth de Touchet et Pierre et Kaeko Régnier qui m'ont enseigné le japonais avec patience et passion, et Vanyda toujours de bonne humeur un crayon à la main.

J'ai ensuite rejoint la Bretagne pour raisons familiales, et je remercie Jean-Marc Geib et David Simplot-Ryl du LIFL, ainsi que Claude Labit et Christine Guillemot de l'IRISA à Rennes pour la compréhension dont ils ont fait preuve dans ce changement d'affectation.

Cette transition n'a pas toujours été simple à gérer, avec un recentrage thématique qui m'a conduite à interrompre certains travaux, notamment en cryptographie et en sécurité des réseaux ad hoc. Mais ce recentrage m'a aussi permis de me focaliser sur la thématique de la protection de contenus, et de la pousser plus loin qu'auparavant.

Je suis arrivée à l'IRISA dans l'équipe TEMICS, que je connaissais depuis 1999. J'avais en particulier collaboré avec Teddy Furon et François Cayre entre 2002 et 2005 sur la formalisation et l'étude de la sécurité des techniques de tatouage. Nos profils complémentaires nous avaient permis d'aborder la question avec un regard original, et cette collaboration m'a laissé le formidable souvenir d'avoir à nous trois apporté un regard nouveau sur un problème important. Un grand merci à tous les deux. Teddy Furon m'a toujours impressionné par son recul et son intuition sur les questions liées au tatouage, et cela a été très enthousiasmant de travailler à ses côtés. Nous y avons développé durant ces quatre années la thématique de la protection des contenus, avec des études consistantes, mêlant travaux théoriques mais aussi opérationnels. Merci à Fabien Galand pour son séjour post-doctoral riche en rebondissements (!) et nos séances de brainstorming mémorables sur ESTIVALE ; à Ana Charpentier, Mathieu Desoubeaux, Fuchun Xie et Çagatay Dikici pour leur travail et leur bonne humeur. Merci aussi à Luce Morin pour nos discussions personnelles, à Laurent Guillo avec qui j'ai été très heureuse de partager mon bureau (Mersi Bras Laurent), à Aline Roumy pour son amitié, à Cédric Herzet et Angélique Drémeaux pour les bons moments que nous avons partagés. J'ai découvert dans cette équipe des thématiques qui ne m'étaient pas ou peu familières, comme la reconstruction d'images 3D ou la compression d'images vidéos, que la préparation de la fête de la science en 2007 m'a permis d'appréhender plus en détail. J'ai particulièrement apprécié l'enthousiasme qui a accompagné cette préparation. Au-delà de l'équipe TEMICS, je souhaite adresser un merci spécial à Véronique Verdon, Myriam David, Pascale Sebillot, Emmanuelle Anceaume, Michel Hurffin, Hervé Jégou et Rémi Gribonval pour leur soutien et leur bonne humeur.

Rennes est une ville particulièrement stimulante sur le plan de la SSI. Merci aux habitués et surtout aux organisateurs des séminaires de crypto de l'IRMAR et des séminaires Diwall. Au-delà des collègues cotoyés à l'IRISA, je souhaite adresser un merci particulier aux collègues de l'IRMAR, de la DGA MI, de Technicolor, d'Orange Labs, de Supélec et de Télécom Bretagne pour nos nombreux échanges, alors et encore aujourd'hui, pour leur richesse et leur diversité. Une mention spéciale à Ludovic Mé et la clique de Supélec. Merci pour votre accueil toujours chaleureux, quel plaisir de passer chez vous ! Autre mention spéciale pour Mohamed Karroumi et Gwenaël Doërr avec qui je partage avec grand plaisir les cours de protection de contenus du Master Recherche.

La SSI à Rennes est aussi marquée par la conférence SSTIC, qui s'y déroule tous les ans et marque je crois tous ceux qui y sont allés, par son ambiance originale et stimulante (ainsi que par ses 450 participants qui réchauffent l'amphi Louis Antoine ...). Chanceux sont les inscrits, qui ont réussi à obtenir une place dans le rush des inscriptions ! J'ai eu la chance de côtoyer ses fondateurs, et d'avoir participé pendant plusieurs années à son organisation. Quelle expérience ! Organiser le SSTIC c'est toute une aventure, avec ses rebondissements, ses coups de tabac, et un résultat toujours surprenant. Beaucoup d'enthousiasme, d'idées, d'échanges, de remises en question permanentes au sein du CO, toujours avec une qualité d'écoute, ou chacun donne le meilleur de lui-même. Bref, une vraie communauté, au sens fort, exigente mais ouverte. Et au bout, un SSTIC du tonnerre ! Un grand merci au canal historique : Frédéric Raynal, Eric Filiol, Christophe Bidan, Thierry Martineau, Phil Biondi et les autres, pour avoir créé le SSTIC. Et un immense merci à Franck Veysset, Céline Entfellner, Benjamin Morin, Olivier Heen, Frédéric Tronel, Nicolas Bareil, Nicolas Prigent, Jean-Philippe Gaulier, avec qui j'ai partagé ces dernières années les affres du PRE et l'exaltation du POST SSTIC.

Dans un tout autre registre, un grand merci à Gilles Bourhis et aux camarades. Merci aussi à la Batouc' de l'IRISA pour m'avoir fait découvrir les joies de jouer de la musique en groupe et le rythme brésilien (dès que je peux je m'y remets !). Mille grands mercis enfin aux amis rennais pour leur amitié et pour tous les bons moments passés tous ensemble : Séverine Baudry et Philippe Nguyen pour nos week-end au grand air, nos discussions de plongée et bien d'autres choses, Aline Roumy et Bert Wiest pour leur éternelle bienveillance, Iryna Andriyanova et Serguei Bury pour leur inébranlable optimisme et leur extrême gentillesse, Axelle Amon et François Maucourant pour nos ch'ti moments (que vivent le home-brew, la musique et WoW !), Hélène Guérin et Arthur Charpentier pour leurs invitations salvatrices et leur sagesse, Agathe et Sébastien Gouézel pour leur éternelle bonne humeur (et hop le petit oiseau est parti !), Magali et Jean-Baptiste Bardet pour leur dynamisme et nos discussions de maman, Lise et Florent Malrieu pour leur amour des tous petits, Pierre et Delphine Loidreau pour nos discussions codes, voyages et petits bouts.

En 2009, j'ai de nouveau changé de laboratoire. Direction le grand ouest, sa pointe Finistère, ses tempêtes et ses paysages magnifiques. Je remercie Patrick Bouthémy, alors directeur de l'IRISA, pour la compréhension dont il a fait preuve dans ce changement d'affectation, et suis très reconnaissante à Alain Hillion pour l'accueil qu'il m'a réservé au Lab-STICC et la confiance qu'il m'a témoignée.

Un grand merci également aux collègues du Lab-STICC et de Télécom Bretagne, pour m'avoir accueillie, écoutée et intégrée si rapidement à leurs projets ainsi qu'aux enseignements de crypto. Merci en particulier à Claude Berrou, Gilles Coppin, Jean-Marc Le Caillec, Frédéric Cuppens, Basel Solaiman, Guy Gogniat, Christian Roux, Gouenou Coatrieux, Nora Cuppens, Sandrine Vaton, René Garello, Michel Jézéquel, Ramesh Pyndiah, Christophe Laot, et Emmanuel Boutillon. Merci à l'ensemble de l'équipe SFIIS du Lab-STICC et du département ITI de Télécom Bretagne pour les liens que nous avons tissés ces derniers mois, avec une mention spéciale pour Didier dont je partage le bureau (et qui profite de mes innombrables conf calls ...). Merci aussi aux collègues de SC, ELEC et INFO à Télécom, et aux collègues de l'UBO et de l'UBS que je croise régulièrement et avec qui je discute toujours avec plaisir.

Merci à Takako Salaun qui a vérifié la transcription du poème donné en début de ce mémoire. J'espère que nous aurons l'occasion de nous revoir régulièrement, et qui sait de parler un peu japonais. Merci aussi à Karine Langlet de nous avoir mises en contact. Un clin d'œil à Fabien Galand et Nathalie Berland pour m'avoir donné les extraits originaux des œuvres d'Hérodote.

Un immense merci à mes amis de Plouzané. Merci avant tout à Liyun et Serge He-Guelton. La vie nous réserve parfois de drôles de coïncidences. Merci à tous les deux pour votre sens du partage et de l'amitié. Merci aussi pour les cours de Tae Kwon Do, qui m'ont permis de me remettre avec bonheur aux arts martiaux après 30 ans d'interruption (!). Ces entrainements m'ont beaucoup apporté, en particulier ces derniers mois. Merci à Stéphanie Even et Ronan Keryell pour leur soutien et leur amitié. Merci pour nos discussions, pour votre maison et votre jardin si accueillants. Un clin d'œil à mes compagnons d'entrainement, Frédéric, Olivier, Séverine, Julie, Sébastien, et à Hyunseuk pour nous avoir entrainés à la coréenne. Autre clin d'œil aux compagnes des cours de zumba et de yoga (merci Maud !).

Au-delà de ces équipes et de mes pérégrinations, j'apprécie énormément de travailler au sein des communautés codes/crypto et signal/image, particulièrement actives et dynamiques en France. Un merci particulier aux collègues et amis de Paris, de Caen, de Limoges, de Nancy, de Toulon, de Bordeaux, de Grenoble, de Lyon, de Marseille, de Montpellier et d'ailleurs, que j'ai côtoyés tout au long de ces années. Merci pour nos discussions stimulantes lors des séminaires, des soutenances de thèses, et des nombreuses journées thématiques des GDRs IM/C2 et ISIS.

En sus des codeurs et cryptographes déjà cités, je souhaite remercier chaleureusement Brigitte Vallée, Philippe Guillot, Pascal Véron, Ayoub Otmani, Fabien Laguillaumie, Jean-Marie Lebars, Gérard Cohen, Gilles Zémor, Thierry Berger, Damien Vergnaud, Javier Herranz, Pierrick Gaudry, Emmanuel Thomé, Robert Rolland, François Rodier, Thomas Sirvent et Sylvain Duquesne, pour nos échanges au cours de ces années, ainsi que Patrice Parraud et Christophe Guyeux pour nos discussions plus récentes.

J'ajoute un merci non dissimulé aux « tatoueurs » Benoît Macq, Henri Maître, Pierre Duhamel, Philippe Nguyen, Séverine Baudry, Gouenou Coatrieux, Fabien Petitcolas, Claude Delpha, Frédéric Lefèbvre, Franck Davoine, Stéphane Pateux, Jean-Marc Chassery, William Puech, Patrick Bas, Teddy Furon, François Cayre, Gaétan Le Guelvouit, Gwenaël Doërr, Johann Barbier, Andreas Westfeld, Ana Charpentier, Mathieu Désoubeaux, Marc Chaumont, Morgan Barbier, Rémi Cogranne, et Ingemar Cox, par ordre d'apparition à l'écran.

Mille mercis à mes amies de toujours Delphine Vaucouloux (longue vie à ta petite famille et à ton arche !), Lydie Legemble (vivement qu'on se croise à Paris), Claire Lauga (Mauruuru ! et j'espère à bientôt dans les îles), et Brigitte Fauvet (des projets plein la tête).

Merci à tous pour nos discussions scientifiques, nos échanges BD, nos lectures croisées, la musique, les home-brew, les trucs de « maman », les trucs de « pas maman », et tout le reste. Et bien sûr un immense merci aux miens. À mes parents et la mémoire de mes grands-parents pour leur amour et leur soutien, à Nicolas, Perrine, Prunelle, Gaspard et Albertine qui habitent trop loin, à ma taty, à Kiki, Gib et Agui que je ne sens jamais très loin malgré les kilomètres, à Juli et Eli un pied de chaque côté de l'océan, à Vanessa et Stéphane pour leurs gentils petits mots. À Christian et Monique pour tout, Florent et Nelly qui sont grands maintenant ;-), à Mémé Juliette qui voit toujours les choses du bon côté. À la mémoire d'Elisabeth partie bien trop tôt, à Roger et sa profonde humanité. À Andrée, Françoise et Claude que nous voyons trop peu.

```
#endif
```

Merci enfin, au-delà de mes collègues et nombreux amis aux deux amours de ma vie qui doivent se demander ce que font tous ces gens que je remercie avant eux. Sans vous, Yves, Philémon, que serais-je aujourd'hui ? Merci à tous les deux, pour votre amour, votre patience, votre enthousiasme, votre exigence aussi, pour tout ce que nous partageons tous les trois !

Merci à tous de m'avoir accompagnée jusque là.

Introduction générale

Mes travaux traitent de la protection de contenus, et se situent à la jonction de la cryptographie, des codes correcteurs, de la théorie de l'information, et du traitement du signal/image. Avant d'entrer plus avant dans les techniques mise en jeu, j'ai tenu à proposer au lecteur quelques repères scientifiques sur les objectifs et moyens relatifs aux questions qui seront abordées dans ce mémoire. J'ai tenu à présenter ensuite mon parcours, car il donne une assise et des repères utiles à l'appréhension de mon travail. Je propose enfin un résumé de mes contributions, dont certaines seront présentées plus en détail dans le corps du mémoire, et d'autres plus brièvement en annexe.

La protection des documents numériques comporte divers aspects. Les *codes correcteurs d'erreur* sont utilisés pour corriger les erreurs qui peuvent survenir lors du stockage ou de la transmission des documents. La *cryptographie*, elle, propose des solutions pour protéger la confidentialité des données (*chiffrement*), pour assurer leur intégrité (*fonctions de hachage*, *signature*), ou encore pour s'assurer de l'identité de la personne qui les envoie (*authentification*, *signature*). De telles primitives sont aujourd'hui intégrées dans de nombreux protocoles de communication standardisés, ainsi que dans les techniques de protection de droit d'auteur comme les DRM (*Digital Rights Management*).

Mais ces techniques ne peuvent, à elles seules, résoudre tous les problèmes liés à la protection de contenus. Avec l'avènement du numérique et la popularisation d'Internet est rapidement apparue la perspective de redistributions illicites massives de documents. Le contrôle d'accès et la mise en place de DRM, conçus pour limiter cette fraude, ont rapidement montré leurs limites. Car ces protections n'agissent que lors de la distribution du document. Mais une fois que celui-ci est joué en clair, sur un balladeur, un ordinateur, une télévision ou tout autre type de terminal, il ne contient plus aucune protection. C'est pourquoi des techniques de protection complémentaires, comme le *tatouage robuste*, ont été étudiées dès le début des années 90. Leur objectif est d'offrir une protection permanente, y compris lorsque le document est joué en clair, ou subit certaines transformations (compression, re-numérisation du signal analogique, etc).

Quelques repères scientifiques

Protéger les contenus contre qui, contre quoi, combien de temps et à quel coût ? Voilà les questions qu'on doit se poser systématiquement. Car il n'y a (presque) jamais de sécurité absolue, et la sécurité a un coût : financier, humain, algorithmique. Il est donc important, premièrement, de décider quelles sont les limites acceptables afin de concevoir une solution présentant un bon compromis complexité/coût/sécurité, et deuxièmement, de savoir analyser proprement la sécurité de la solution une fois celle-ci finalisée. Cette analyse est effectuée au regard de l'état de l'art, et on assiste à une course en avant entre concepteurs et attaquants, course qui repousse sans cesse les limites au gré de l'ingéniosité des uns et des autres. Je présente ci-après brièvement quelques rappels sur les domaines abordés dans ce document.

La Cryptologie remonte à l'antiquité. Jusqu'en 1976, son unique objet a été d'assurer la confidentialité des données, grâce à des systèmes de *chiffrement*. La *cryptologie* comporte deux volets : la *cryptographie* [1], qui traite de la conception des systèmes ; et la *cryptanalyse*, qui traite de leurs attaques. Parmi les systèmes historiques, on peut citer le code de César dans l'antiquité, le système de chiffrement de Vigenère au XVIIème siècle, ou encore le chiffrement de Vernam ou One-Time-Pad pendant la première guerre mondiale. Tous ces systèmes sont dits *à clé secrète*, ou *symétriques* ; ils sont caractérisés par le fait que les clés de chiffrement et de déchiffrement sont immédiatement déductibles l'une de l'autre (elles sont même le plus souvent égales). Ceci implique que l'émetteur et le récepteur du message doivent s'accorder à l'avance sur leur valeur, et que personne d'autre ne doit connaître les clés utilisées. Depuis 1976 [DH76] on dispose de systèmes de chiffrement *à clé publique*, ou *asymétriques*, qui ne nécessitent plus cet accord préalable, car la clé de chiffrement peut être publiquement diffusée sans mettre en danger la clé, elle privée, qui sert à déchiffrer. Par ailleurs, les fonctionnalités de sécurisation se sont étendues à la *signature numérique*, à la vérification de l'*intégrité* des données grâce aux *fonctions de hachage*, ou encore à l'*authentification* de personnes, entités ou données pour citer les plus classiques ; mais aussi à des fonctionnalités plus surprenantes : aux preuves *zero-knowledge* (ou comment prouver que l'on connaît une donnée sans rien révéler sur cette dernière), au *partage de secret*, aux calculs opérés sur données chiffrées (donc en aveugle), etc.

Les systèmes cryptographiques sont aujourd'hui présents partout, car la sécurité n'est plus seulement une affaire militaire comme elle l'a longtemps été. Elle touche aujourd'hui à la vie quotidienne des citoyens pour leurs moyens de paiement, leur dossier médical, et toutes les transactions numériques opérées par les entreprises, les banques, etc. La sécurité des systèmes et des primitives cryptographiques qui les compose est soit empirique, soit prouvée dans le cadre de modèles de preuve particuliers reposant sur des hypothèses plus ou moins réalistes (concernant par exemple la robustesse des fonctions de hachage, ou encore la réduction de cryptanalyses à des problèmes mathématiques comme la factorisation ou le calcul du logarithme discret). Le seul système dont la sécurité absolue a été prouvée pour des conditions

1. Terme fréquemment utilisé en lieu et place de cryptologie quand la distinction n'est pas nécessaire.

d'utilisations réelles est le chiffrement de Vernam, la preuve en ayant été établie par Shannon en 1949 [Sha49]. Cette preuve repose malheureusement sur des conditions d'utilisation extrêmement contraignantes, limitées à des communications très sensibles, comme le téléphone rouge. Ainsi, pour tous les autres systèmes, la sécurité est évaluée au regard de l'état de l'art, et en fonction des compromis à réaliser en termes de coûts, algorithmiques et éventuellement financiers.

La recherche en cryptologie est aujourd'hui très vaste, et repose sur des problèmes mathématiques très variés : théorie des nombres, algèbre sur corps finis, interaction avec la théorie des codes, théorie de l'information, preuves formelles, complexité, choix d'implémentation, etc. Elle est aussi très dynamique, car nous avons toujours besoin d'éprouver les systèmes en cours d'utilisation, et d'en concevoir de nouveaux, plus performants, et surtout de conserver une grande variété de systèmes, au cas ou certains s'avèreraient insuffisament sûrs. Aussi de nombreux appels ont été lancés ces dernières années pour stimuler la conception de nouveaux systèmes de chiffrement, de fonctions de hachage, etc.

Pour approfondir les multiples facettes de la cryptologie moderne, je recommande la lecture de [CLdV, MvOV97, vT05, Vau06].

La théorie des codes correcteurs et la théorie de l'information sont plus récentes et datent des travaux de Shannon [Sha48]. Avec l'avènement des communications et des ordinateurs, il est devenu indispensable de protéger les données transmises ou stockées contre les erreurs. Pour ce faire, on enrichit les données en « rajoutant » des symboles supplémentaires, introduisant ainsi de la *redondance*. Le *code* est caractérisé par la longueur et la nature de cette redondance, qui servira en cas de déterioration à retrouver les symboles altérés (erreurs) ou perdus (effacements), reconstituant le mot de code d'origine et par là même les données. Les contextes de stockage/transmission (satellite, internet filaire, acoustique sous-marine, wi-fi, téléphone cellulaire) ou de stockage (disque dur, CD-ROM, Blu-Ray) sont modélisés mathématiquement par des *canaux*, qui permettent de déterminer les meilleurs paramètres à utiliser et les codes les plus adaptés. Le principal problème est de déterminer les codes offrant non seulement une bonne capacité de correction théorique, mais aussi un bon algorithme de décodage/correction. De nombreuses familles de codes ont été étudiées, comme les codes de Hamming, les codes BCH, les codes de Reed-Müller, les codes de Reed-Solomon, et plus récemment les turbo codes, etc. Il reste cependant de nombreux problèmes pour certaines familles de codes : détermination de certains paramètres comme le *rayon de recouvrement* ou la *distance minimale*, recherche d'algorithmes de *décodage* toujours plus performants ; mais aussi conception de nouveaux codes, plus performants ou adaptés à de nouveaux contextes applicatifs.

Quelques éléments d'introduction à la théorie des codes linéaires en blocs sont donnés dans le chapitre 1. Pour le lecteur désireux d'en savoir plus, je recommande la lecture de [PHB98, HP03].

Stéganographie, tatouage et *fingerprinting*. La *stéganographie* est ancienne et on en trouve trace jusque dans la Grèce antique. Son objectif est de rendre furtive une communication, et donc de masquer l'existence-même de cette dernière. Crâne de messager masqué par des cheveux fraîchement repoussés, tablettes gravées recouverts d'argile fraîche, utilisation de synonymes bien choisis ou encres sympathiques en ont assuré la fonction pendant des siècles. La stéganographie a connu un regain d'intérêt depuis les années 1990 avec l'avènement du numérique et la possibilité de cacher des messages dans des documents multimédia, grâce à des techniques de traitement de signal/image. Deux techniques complémentaires se sont également développées : le *tatouage numérique*, appelé *digital watermarking* en anglais, et le *traçage de copies*, appelé *(active) fingerprinting* en anglais.

Afin d'éviter toute ambiguïté, je redonne ici la terminologie employée par la majeure partie de la communauté. On parle généralement de *dissimulation d'information*, ou *information hiding* en anglais, pour désigner toute technique visant à cacher de manière imperceptible des données dans un document support, quelle que soit sa nature. On parlera plus spécifiquement de *stéganographie* lorsque l'objectif est de rendre cette communication furtive. On parlera de *tatouage*, lorsque la robustesse des données cachées est déterminante, par exemple dans les contextes de protection de droit d'auteur, ou bien encore lorsque l'on cherche à détecter d'éventuelles modifications dans le document. On parlera enfin de *fingerprinting* lorsque l'on cherche à personnaliser les différentes copies d'un même document, afin de tracer les utilisateurs indélicats qui les redistribueraient de manière illégitime.

Ces trois techniques ont en commun leur souci de cacher de l'information de manière imperceptible, transparente pour l'utilisateur, en conservant le format et la taille d'origine du document. Cependant, la stéganographie impose une exigence d'imperceptibilité plus forte, recherchant une réelle furtivité face à des distingueurs conçus pour différencier les documents porteurs de messages de ceux qui n'en portent pas. Par ailleurs, le fingerprinting et le tatouage exigent, dans le cadre de l'identification d'un utilisateur ou d'un ayant droit, une robustesse de l'information qui identifie l'utilisateur malgré toutes sortes de transformations ou attaques.

Notons que certaines applications relèvent de dissimulation d'information sans pour autant être de la stéganographie, du tatouage ou du fingerprinting. Par exemple lorsqu'il s'agit de faire porter par le document hôte des méta-données non nécessaire à son utilisation, mais qui enrichissent son contenu. On a alors les mêmes exigences en imperceptibilité que pour le tatouage ou le fingerprinting, mais moins d'exigence en termes de robustesse. Nous ne reviendrons pas dans ce document sur ce type d'application.

Comme nous le verrons dans ce mémoire, il existe des liens assez forts entre le trio stéganographie-tatouage-*fingerprinting* et la cryptographie. Tout d'abord parce que les notions de sécurité développées en cryptographie ont aidé à la formalisation de la sécurité des techniques de stéganographie et de tatouage. Ensuite parce que le tatouage aborde des questions parfois proches de celles de la cryptographie, comme la vérification de l'identité et/ou de l'intégrité de l'auteur d'un document, ou d'un document lui-même. Ensuite parce que le tatouage a été conçu comme une technique de protection complémentaire aux primitives cryptographiques dans le cadre de la

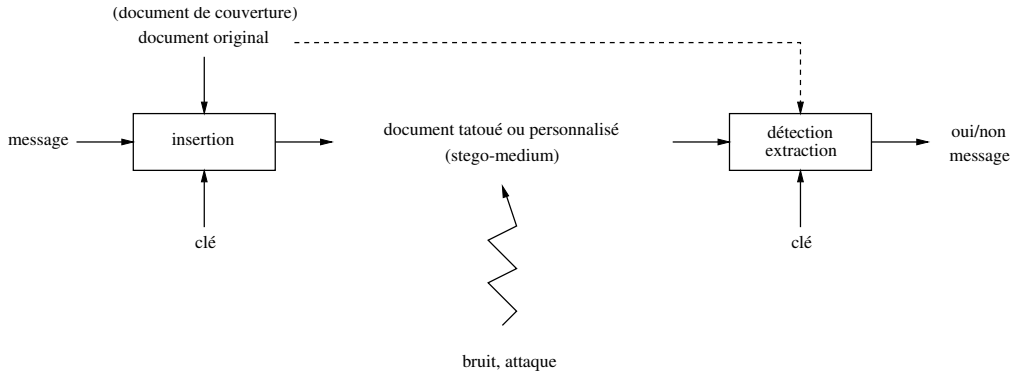


FIGURE 1: Modélisation de la stéganographie, du tatouage et du *fingerprinting* comme des problématiques de transmission. Le vocabulaire propre à la stéganographie est mis entre parenthèses. La flèche en pointillés indique que dans certains cas particuliers de tatouage ou *fingerprinting* le document original peut être disponible lors de la détection/extraction.

protection de contenus, comme cela a été évoqué au début de cette introduction. Ainsi, ces outils sont amenés à s'épauler au sein d'architectures parfois complexes.

Par ailleurs, comme le montre la figure 1, les problématiques de stéganographie-tatouage-*fingerprinting* peuvent être formalisées comme des problèmes de communication, bien que certaines contraintes leur soient spécifiques. Elles sont donc également fortement liées à la théorie de l'information et aux techniques de télécommunications. En ce qui concerne les documents multimédia, elles s'appuient fortement sur la théorie du signal/image. On retrouve aussi dans certaines publications des apports de la théorie des jeux.

Comme nous l'avons rapidement vu, la dissimulation d'information est riche en contributions, car elle se situe à la frontière de nombreux domaines. Bien qu'assez récente, le nombre de publications y a été colossal, comme le montre la figure 2, et elle atteint aujourd'hui une bonne maturité.

Les ouvrages de référence pour approfondir le sujet sont [KP99, CMB+08, Fri09b, Böh10].

Mon parcours

Mon activité de recherche a débuté en 1995 lors du stage de DEA que j'ai effectué au projet CODES (maintenant SECRET) de l'INRIA Rocquencourt sous la direction de Pascale Charpin. Le DEA que j'avais suivi m'avait particulièrement plu par la diversité des problématiques abordées, de la méthode B à la cryptographie, en passant par l'algorithmique, la complexité, les bases de Gröbner ou encore les codes correcteurs d'erreur. Lors de mon stage, puis de ma thèse [Fon98], j'ai exploré un corpus de translatés des codes de Reed-Müller d'ordre 1 dans le but d'étudier le *rayon de recouvrement* de ces codes et les *distributions des poids* associées à ces translatés. Ce rayon est un élément primordial pour connaître la capacité de correction du code, et

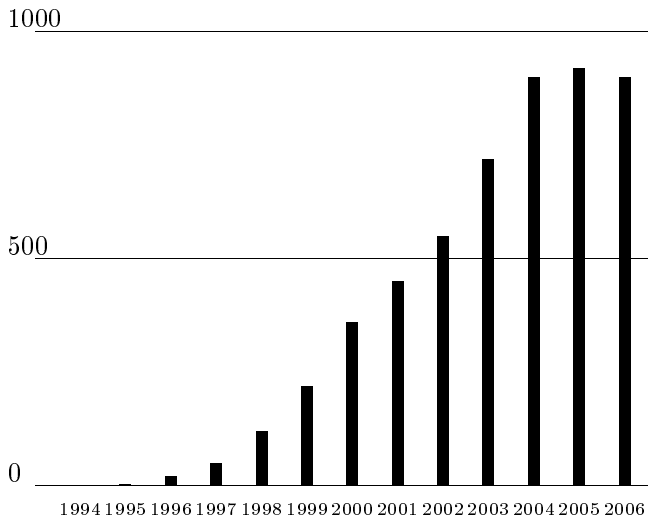


FIGURE 2: Nombre annuel de publications IEEE portant sur le tatouage et la stéganographie [CMB+08].

reste malheureusement inconnu pour toute une famille de paramètres. Par ailleurs, ce problème peut être envisagé sous un autre angle, celui de la cryptographie. Il correspond alors à la recherche de fonctions booléennes suffisamment robustes pour être utilisées comme primitives cryptographiques dans des systèmes de chiffrement symétriques. De fait, les cryptographes sont probablement les plus motivés pour le résoudre, mais les meilleurs outils proviennent de la théorie des codes. En parallèle à ce travail sur les fonctions booléennes, j'ai été impliquées pendant ma thèse dans un projet européen, Aquarelle, qui visait à proposer une architecture permettant aux musées européens de diffuser des images de leurs œuvres, tout en leur garantissant un certain niveau de sécurité, et notamment en limitant les risques de réutilisation frauduleuse de ces images. J'ai participé à la sécurisation de ces images, *via* une technologie alors émergente, le *tatouage de données numériques*, qui relevait à l'époque principalement du traitement du signal/image.

Après ma thèse, j'ai passé un an comme ATER au LRI à Orsay dans l'équipe ALGO, puis j'ai obtenu en 1999 un poste de Maître de Conférences au LIFL à Lille, dans l'équipe RD2P. En 2002, j'ai intégré le CNRS, toujours au LIFL. Pour raisons familiales, j'ai rejoint l'IRISA à Rennes en 2005, dans l'équipe TEMICS, puis le Lab-STICC et Télécom Bretagne à Brest en 2009, respectivement dans l'équipe SFIIS et le département ITI.

J'ai donc été membre de différents laboratoires, différentes équipes, et surtout côtoyé des collègues aux profils scientifiques variés et complémentaires : codes et cryptographie à Rocquencourt, algorithmique à Orsay, systèmes d'exploitation et réseaux ad-hoc à Lille, codage de source et dissimulation d'information à Rennes, traitement de l'image et de l'information, et sécurité, à Brest, si je me limite au strict périmètre des équipes.

Contributions

J'ai réellement apprécié être à la croisée de plusieurs domaines et communautés, travailler avec des chercheurs aux profils et aux méthodes si différents. C'est pourquoi mes contributions sont presque toutes à cheval entre deux ou trois domaines de recherche, comme l'illustre la figure 3.

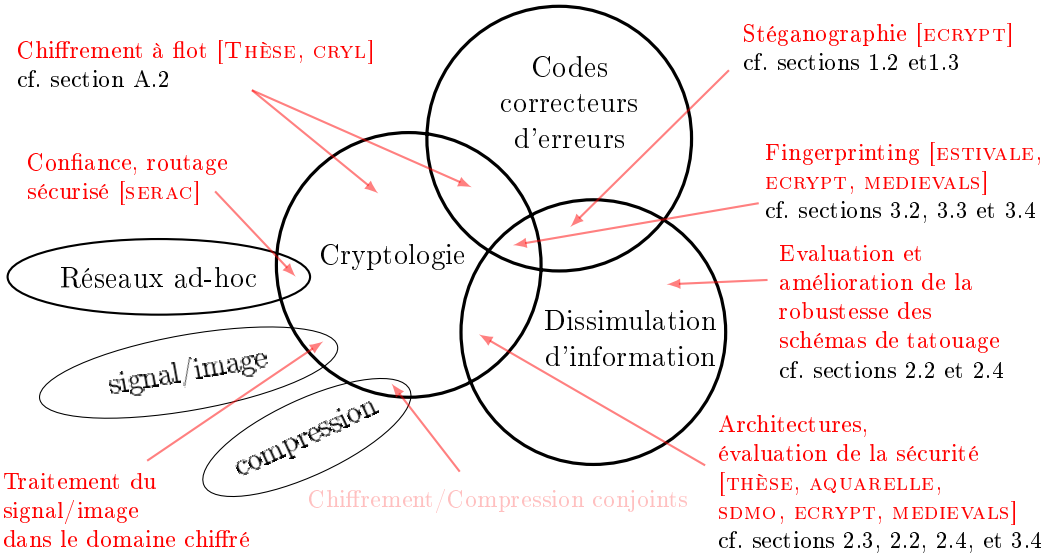


Figure 3: Thématiques abordées. Les noms entre crochets permettent de faire le lien avec la liste des projets donnée dans la section C.1 de l'annexe C.

Pour résumer en quelques lignes mes contributions, j'ai au cours de ces années :

- étudié des systèmes de chiffrement, utilisant pour cela des objets mathématiques venant de la théorie des codes correcteurs [FF98, Fon99, CCCF00b, CCCF01, FFJ04, BRWF05] ;
- élaboré des architectures dédiées mêlant outils cryptographiques et tatouage [AFD98, ABD+99, BTFF+06, FDD+08] ;
- étudié la sécurité des techniques de tatouage en m'appuyant sur des méthodologies issues de la cryptographie [FR02, CFF05b, CFF05e] ;
- travaillé à l'amélioration de la robustesse et de la sécurité de techniques de tatouage [CXFF09, XFF10a, XFF10b] ;
- développé des techniques de traçage de vidéo en faisant interagir des systèmes de tatouage et des codes anti-collusion [XFF08, CXFF09, CFF10, CFFC11] ;
- proposé de nouvelles techniques de stéganographie reposant sur les codes correcteurs, et dont le succès d'insertion est garanti [FG07b, FG09, ABF11].

J'ai par ailleurs exploré d'autres pistes comme l'utilisation de systèmes de chiffrement homomorphiques en signal [FG07a] ou encore l'élaboration de techniques de tatouage-compression conjoints [DGF+08], compression-chiffrement conjoints, ou de codage-chiffrement conjoints.

Organisation du manuscrit

J'ai choisi de placer au cœur de ce manuscrit la protection des documents multimédia par des techniques de dissimulation d'information. Le chapitre 1 aborde la conception de schémas de stéganographie *via* l'utilisation détournée de codes correcteurs d'erreur. Le chapitre 2 est dédié à la conception et à l'attaque de schémas de tatouage robuste. Le chapitre 3 discute de l'utilisation de codes *anti-collusion* pour la personnalisation et l'identification de copies. La cryptographie et les codes correcteurs apparaissent en filigrane dans ces trois chapitres.

Les introductions aux trois chapitres principaux reflètent ce que je présenterais à un étudiant qui voudrait travailler avec moi sur ces questions. C'est aussi pour cela que j'y présente un panorama de chacun des domaines que j'ai abordés, panorama bien entendu légèrement biaisé par mon profil scientifique, certains aspects étant plus détaillés que d'autres. Mais j'ai néanmoins tenu à n'omettre aucun aspect important de ces domaines, et en particulier à ne pas oublier les points sur lesquels je n'ai pas moi-même travaillé, mais qui sont cependant essentiels. En ce qui concerne la stéganographie, il m'a semblé utile de détailler la présentation du codage par syndrome qui est généralement présentée trop brièvement pour être accessible à ceux qui ne pratiquent pas la théorie algébrique des codes tous les jours. Pour les autres chapitres, j'ai été plus concise car la littérature me semble plus claire et demande moins d'efforts pour être maîtrisée.

Le résumé en anglais, qui se situe juste après cette introduction, présente des introductions beaucoup plus synthétiques.

Quelques annexes viennent compléter ce manuscrit. L'annexe A présente tout d'abord brièvement d'autres travaux que j'ai menés et qui n'ont pas été détaillés dans le corps du manuscrit, pour certains parce qu'ils me semblaient moins marquants, pour d'autres parce qu'ils s'éloignent de la thématique centrale retenue ici (c'est en particulier le cas pour mes contribution en cryptographie). Viennent ensuite dans l'annexe B les transparents utilisés lors de ma soutenance. L'annexe C est, elle, consacrée à la présentation de mon *curriculum vitæ*, ainsi que de ma liste de publications. Enfin, j'ai inclus dans l'annexe D cinq de mes publications dont la présentation dans le corps du manuscrit a été particulièrement synthétique, voire frustrante pour le lecteur qui souhaite s'y plonger.

Notations

Les notations utilisées dans ce document sont propres à chaque chapitre. En effet, les notations standard utilisées pour la stéganographie, le tatouage, les codes de Tardos, ou encore pour les protocoles cryptographiques, sont différentes. Aussi j'ai fait le choix de conserver pour chaque chapitre les notations les plus standard possibles, afin de ne pas égarer le lecteur familier du domaine. En contrepartie, le lecteur devra faire l'effort d'éventuellement changer de notations d'un chapitre à l'autre.

Overview

This manuscript presents a synthetic overview of the results obtained since my PhD defense in 1998. Its purpose is to provide some landmarks to locate my contributions in the wide area of content protection. This English part summarizes my research contributions, and provides a short *curriculum vitæ*. The French part summaries them as well, but also provides a more detailed and personal overview of the research areas. More details on the contributions can be found in the referred articles.

Before going deeper in the subject, I would like to recall some important events that marked my researcher's life, as they impacted my research interests and work. When I defended my PhD thesis, my scientific background was mainly oriented towards discrete mathematics and algebra, and towards computer science. I was studying how coding theory can help designing good symmetric encryption schemes, through the identification of cosets of the first order Reed-Müller code of high minimum weight with nonlinear Boolean functions used as cryptographic primitives. I was also interacting with researchers from signal and image processing, as I worked on the integration of digital watermarking techniques in a system dedicated to the copyright protection of European Museums' still images. As I liked working with people coming from different areas, and with different backgrounds, I continued to do so. During my PhD, I worked in Rocquencourt (at INRIA, 1995-1998, while teaching at Cergy Pontoise University) and Orsay (at LRI, 1998-1999, while teaching at Paris XI University) near Paris. Then I worked as an associate professor and then a CNRS full time researcher in several laboratories in Lille (at LIFL, 1999-2005, while teaching at Lille 1 University), in Rennes (at IRISA, 2005-2009), and finally in Brest (at Lab-STICC and Télécom Bretagne/ITI, since 2009). Different teams, but also different research profiles : cryptography and coding in Rocquencourt, algorithmics in Orsay, ad-hoc network and operating systems in Lille, source coding and information hiding in Rennes, image and information processing and security in Brest.

This manuscript summarizes my contributions in the area of content protection, focusing on results related to information hiding, which encompasses *steganography*, *watermarking* and *fingerprinting*. All of them embed a message in a document in an imperceptible way, but their objectives are different, leading to different constraints, and then to different techniques. Hiding a robust but invisible mark in delivered documents, *watermarking* and *fingerprinting* have been introduced in the 1990s to complement cryptographic primitives for copyright protection : their goal is to provide a protection even once the document has been (legally) decrypted by the end

user, and is stored, played, or transmitted as a plaintext. Hence, they offer a more perennial security, as long as the hidden signal remains detectable. During last decade, they have reached many other application scenarios, each of them leading to different constraints in terms of robustness, imperceptibility and capacity. But in any case, robustness is an essential criterion, that has to be correctly managed. *Watermarking* and *fingerprinting* differ in the sense that we usually refer to *watermarking* when all the end users receive the same version of the marked document, whereas we refer to *fingerprinting* when each end user receives a personalized version of the document, which can help tracing any of its unauthorized use. But in both cases, the adversary is aware that there is some particular information hidden in the document. The purpose of *steganography* is really different, as it aims at hiding the message transmission itself. In this context, the adversary tries to detect if there is something hidden or not. Hence, imperceptibility for the end user is still necessary, but no more sufficient. On the other hand, robustness is usually of no interest.

Chapter 1 – Stealth message transmissions : how error correcting codes can help designing efficient steganographic schemes.

Hiding messages in innocuous-looking *cover-media* in a *stealthy* way, steganography is the art of stealth communications. The sender and receiver may proceed by cover selection, cover synthesis, or cover modification to exchange messages. Here, we focus on the cover modification scenario, where the sender chooses some *cover-medium* in his library, and modifies it to carry the message he wants to send. Once the cover-medium is chosen, the sender extracts some of its components to construct a *cover-data* vector, here denoted by $\mathbf{x}$. Then, he modifies it to embed the message. This modified vector, $\mathbf{y}$, is called the *stego-data* ; it leads back to the *stego-medium* that is communicated to the recipient (see Fig. 1.1). In the case of digital images, the insertion may for example consist in modifying some of the images components, *e.g.* the luminance of the pixels or the values of some transform (*e.g.*, DCT or wavelet) coefficients. For a given transmitted document, only the sender and receiver have to be able to tell if it carries a hidden message or not [Sim84]. This means that the *stego-media*, which carry the messages, have to be statistically indistinguishable from original media [Cac98, Cac04]. But ensuring perfect indistinguishability is currently intractable, because of the difficulty to model properly the media [Böh10]. We then usually use heuristics to design steganographic schemes. As statistical detectability of most steganographic schemes increases with *embedding distortion* [KFP07], minimizing embedding distortion is one possible heuristic. Several distortion measures can be considered, but the most studied one is the number of embedding changes. Hence, one possible design criterion is to minimize the number of components of $\mathbf{x}$ that will be modified during embedding.

In 1998, Crandall proposed to model the embedding and extraction process with the use of linear error correcting codes. He proposed to use Hamming codes, which are covering codes [Cra98]. The key idea of this approach, called *syndrome coding*, or *matrix embedding*, is to modify the cover-data $\mathbf{x}$ to obtain a stego-data $\mathbf{y}$ that

lies in the *right coset* of the code, its *syndrome* being then precisely equal to the message to hide. Later on, it has been showed that designing steganographic schemes is precisely equivalent to designing covering codes [Bie01, GK03, GK09], meaning that this covering codes approach is not restrictive. Moreover, it has been shown to be really helpful and efficient to minimize the embedding distortion [Bie01, GK03, GK09, BF08]. It has also been made popular due to its use in the famous steganographic algorithm `F5` [Wes01]. For all these reasons, this approach is of interest.

The process that states which components of the cover-data can actually be modified is called the *selection channel* [AP98]. Since the message embedding should introduce as little distortion as possible, the selection channel is of utmost importance. The selection channel may be arbitrary, but a more efficient approach is to select it dynamically during the embedding step, accordingly to the cover-medium and the message. This leads to a better undetectability, and makes attacks on the system harder to run, but in this context the extraction of the hidden message is more difficult as the selection channel is only known to the sender, and not to the recipient. *Wet Paper Codes* were introduced to tackle this non-shared selection channel, through the notions of *dry* and *wet* components [FGLS05]. By analogy with a sheet of paper that has been exposed to rain, we can still write easily on dry spots whereas we cannot write on wet spots. The idea is, adaptively to the message and the cover-medium, to *lock* some components of the cover-data — the wet components — to prevent them being modified. The other components — the dry components — of the cover-data remain free to be modified to embed the message.

Algorithmically speaking, syndrome coding provides the recipient an easy way to access the message, through a simple syndrome computation. But to embed the message, the sender has to tackle a harder challenge, linked with unique decoding. It has been shown that if random codes may seem interesting for their asymptotic behavior, their use leads to solve really hard problems : syndrome complete decoding and covering radius computation, which are proved to be NP-complete and Π_2-complete respectively [Var97, McL84]. Moreover, no efficient decoding algorithm is known, even for a small non trivial family of codes. Hence, attention has been given on structured codes to design Wet Paper Codes : Hamming codes [Cra98, FGS05a, ZW06, FL07], Golay codes [ZW06, Mie06, FL07], Simplex codes [FS06], BCH codes [SW06, SW07, ZSK09, SKZ09, OMS10], Reed-Solomon codes [FG07b, FG09], perfect product codes [RPR09, RR10], low density generator matrix codes [FF07, ZZW08, ZZW10, FF09b], and convolutional codes [FJF10, FF10b, FJF11].

Embedding techniques efficiency is usually evaluated through their *relative payload* (number of message symbols per cover-data (modifiable) symbol) and *average embedding efficiency* (average number of message symbols per cover-data modification). Today, we can find in the literature quasi-optimal codes in terms of average embedding efficiency and payload [FF07, ZZW08, ZZW10, Fri09a, FF09b]. These last years, I have been interested in another criterion, which is usually not discussed : the probability for the embedding to fail. In fact, there are only two cases for which it never fails : (a) when using perfect codes if we do not need to lock any component of the cover-data ; or (b) when using Maximum Distance Separable codes if we need to lock some components of the cover-data. But very few codes are perfect (namely

the Hamming and Golay codes), and their average embedding efficiency is quite low. Moreover it is really important in practice to be able to lock some components of the cover-data. Hence, Case (a) is not of real interest. Case (b) is of interest, but there are not many MDS codes that can be used in a steganographic context. Hence, efficient practical schemes usually do not fit either (a) or (b), leading to a non-zero probability for the embedding to fail. And this probability increases with the number of locked components. More precisely, syndrome coding usually divides the whole message into fragments, that are separately inserted in different cover-data vectors (coming from one or several cover-medium). Inserting each fragment involves finding a low weight solution of a linear system which may not always have a solution for a given set of locked components. Consequently, the probability that the whole message can be embedded decreases exponentially with the number of fragments to hide and with the number of locked components [FGS05a].

Hence, we have to decide what to do when embedding fails. In the common scenario where the sender has to choose a cover-medium in a huge collection of documents, he can drop the cover-medium that leads to a failure and choose another one, iterating the process until finding a cover-medium that is adequate to embed the message. Another solution may be to cut the message into smaller pieces, in order to have shorter messages to embed, and a lower probability of failure. If none of these is possible, for example if the sender only has few pieces of content, he may unlock some locked components [FJF10] to make the probability of failure decrease. But, even performing such a modified embedding, which decreases the probability of failure, the sender will not be able to drop it to zero, except if he falls back to perfect codes without locked components.

This failure probability may be decreased, and even suppressed, if we leave the unique decoding approach and authorize the decoding process to provide a set of solutions which are even non-optimal, but sub-optimal, in terms of the number of modifications they imply on the cover-data. Moreover, enlarging the set of solutions will enable to choose the best one, not only for the minimization of the number of modifications, but also for other distortion measures. This approach has not been paid too much attention yet, even if promising, as in this case not only perfect or MDS codes may ensure embedding.

My first contribution has been to show with Fabien Galand that Reed-Solomon codes, which are Maximum Distance Separable, may be really helpful to manage as many locked components as possible, while ensuring embedding through a combination of Lagrange interpolation and list decoding. This work is summarized in Section 1.2.

My second contribution has been to propose with Daniel Augot and Morgan Barbier a new kind of syndrome coding scheme, that always ensures embedding. It is summarized in Section 1.3.

Section 1.2 – How Reed-Solomon codes can improve the management of locked components : F. Galand's postdoc (2006-2007). With Fabien Galand, we focused on Schönfeld *et al.* paper published in 2006, which suggests to use binary BCH codes to perform syndrome coding in the wet paper context [SW06]. The authors showed that binary BCH codes behave well when no component has to be locked, but also pointed out that choosing the most appropriate code among the BCH family is quite hard, as we do not know good complete syndrome decoding algorithms for BCH codes. In the wet paper context, they showed that there is a trade-off between the number of components that can be locked and the efficiency of the code.

Generalized Reed-Solomon codes are a sub-family of BCH codes, defined on the Galois field $GF(q)$. They are particularly interesting in steganography because the generally hard problems we have to solve is syndrome coding (efficient decoding, covering radius computation) are solved for Generalized Reed-Solomon codes : their minimum distance and covering radius are known, with interesting values, and we know efficient decoding algorithms, for the unique decoding paradigm, but also for the list decoding one ; moreover, they are Maximum Distance Separable and can then manage as many locked positions as possible without any failure. In this sense, they are optimal.

We first provided an efficient algorithm to perform embedding through Lagrange interpolation. This approach also has the good property to never fail, even if Generalized Reed-Solomon codes are not perfect codes, as we accept not to get the strict minimum number of modifications, but even a slightly higher number of modifications. To improve the embedding efficiency of this algorithm, we mixed the Lagrange interpolation with Guruswami-Sudan's list decoding. List decoding is performed when possible, leading to a set of solutions among with we can choose the best one in terms of distortion. This also improves embedding efficiency, as we can sometimes embed the same message while modifying 50% less components of the cover-data. If list decoding fails, we fall back to Lagrange interpolation, to ensure embedding. This leads to an adaptive trade-off, enabling the management of the whole set of locked positions, while keeping the best possible embedding efficiency and undetectability. To illustrate our results, Fig. 1.4 shows, for $q = 64$, some parameters for which the list decoding is really helpful. Sub-figures 1.4(a-d) shows a pessimistic estimation of the probability of success for the list decoding, according to the code parameters (length n and dimension k). The case $q = 32$ is provided to better see what is going on the edge of the triangle. Sub-figure 1.4(e) shows the gain it provides in terms of the number of modifications of the cover-data, for some of the interesting parameters.

These results have been published it the proceedings of the *International Workshop on Information Hiding, IH'07* [FG07b], and in a special issue of *EURASIP Journal on Information Security*, entitled *Secure Steganography in Multimedia Content* [FG09] [2]. They have been supported by the European Network Of Excellence ECRYPT.

2. This article can be found in Appendix D.

In order to use them in real applications, several issues still have to be addressed. First, we need to choose an appropriate measure to properly estimate the distortion induced at the medium level when modifying the symbols at the data level. Second, we need to use a non binary, and preferably large, alphabet. A straightforward way to deal with this would be to simply regroup bits to obtain symbols of our alphabet and consider that a symbol should be locked if it contains a bit that should be. Unfortunately, it would lead to a large number of locked symbols (e.g. 5% of locked bits leads to up to 20% of locked symbols if we use $GF(16)$). A better way would be to use grid coloring [FL07], keeping a 1-to-1 ratio. But, the price to this 1-to-1 ratio would be a cut in payload. We think a good solution has yet to be figured out. Nevertheless, in some settings, a large alphabet arises naturally : for example, in [ZZW08], a (binary) wet paper code is used on the syndromes of a $[2^k-1, 2^k-k-1]$ Hamming code, some of these syndromes being locked ; here, since whole syndromes are locked, we can view syndromes as elements of the larger field $GF(2^k)$ and use our proposal. Third, no efficient implementation of the Guruswami-Sudan's list decoding algorithm is available ; and as the involved mathematical problems are really tricky, only a specialist can perform a real efficient one. Today, these issues remain open.

Section 1.3 – Ensuring embedding : M. Barbier's PhD thesis (2008-2011). With Daniel Augot and Morgan Barbier, we focused on the "worst case" scenario, where the sender does not have too much cover documents to hide his message in, and then absolutely needs embedding to succeed. This scenario is not the most studied one, and concerns very constrained situations.

Our contribution is to propose a new generic embedding scheme that will never fail, whatever the code used, and does not relax the management of locked components of his cover-data to make embedding succeed. It is, to our knowledge, the first bounded syndrome coding scheme that manages locked components while guaranteeing the complete embedding of the message for any code, be it perfect or not.

To do so, we modify the classical syndrome coding approach, where embedding and extraction steps are usually written as

$$\begin{aligned} \mathrm{Emb}(\mathbf{x}, \mathbf{m}) &= \mathbf{v} + \mathrm{D}^*_{\mathcal{W}}(\mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t) \\ \mathrm{Ext}(\mathbf{y}) &= \mathbf{y} \cdot \mathbf{H}^t \end{aligned}$$

with $\mathrm{D}^*_{\mathcal{W}}()$ a function returning a word (e) of the given syndrome, such that $\mathbf{e}_i = 0 \ \forall i \in \mathcal{W}$ (wet components must not be modified) and $d_H(\mathbf{e}, 0)$ (the number of modifications) has not to be greater than a given upper-bound. Our idea is to use some part of the syndrome for randomization. More precisely, the syndrome is now composed of two parts : the message m we want to embed, and a free part R that could take any value, as in Courtois *et al.*'s signature scheme [CFS01]. Embedding and extraction are now written as

$$\begin{aligned} \mathrm{Emb}(\mathbf{x}, \mathbf{m}) &= \mathbf{v} + \mathrm{D}^*_{\mathcal{W}}((\mathbf{m}||\mathbf{R}) - \mathbf{x} \cdot \mathbf{H}^t) \\ \mathrm{Ext}(\mathbf{y}) &= \mathbf{y} \cdot \mathbf{H}^t \end{aligned}$$

As the syndrome's length is still $n-k$, the embedded message length is now $n-k-r$, where r denotes the length of the random part R. Then, there is a loss in embedding

efficiency, compared with the original syndrome coding approach. This is the cost to pay to ensure embedding.

We estimated the relative loss in embedding efficiency for a given number of locked components $\ell_{\mathcal{W}}$:

$$\frac{e - e'}{e} = \frac{r}{n-k},$$

where e denotes the embedding efficiency of the traditional approach, while e' denotes the efficiency of our randomized one. Optimizing the parameter r is crucial, to ensure that our reformulated problem always has a solution, while preserving the best possible embedding efficiency. We derived sufficient conditions on r to ensure a successful embedding, in the case of perfect linear codes.

We showed that for the binary $[23, 12, 7]$ Golay code, a sufficient condition for the embedding to be always successful is

$$r \geq \log_2 \left(1 + \frac{796}{3}\ell_{\mathcal{W}} - \frac{23}{2}\ell_{\mathcal{W}}^2 + \frac{1}{6}\ell_{\mathcal{W}}^3\right).$$

For the ternary $[11, 6, 5]$ Golay code, a sufficient condition is

$$r \geq \log_3 \left(1 + 44\ell_{\mathcal{W}} - 2\ell_{\mathcal{W}}^2\right).$$

Fig. 1.5 shows that for both the number $n - k - r$ of symbols that can be used to embed the message decreases very fast when the number $\ell_{\mathcal{W}}$ of locked components increases.

In the case of q-ary Hamming codes, of parameters $[(q^p - 1)/(q - 1), n - p, 3]$, we showed that we can embed at least one message symbol if $\ell_{\mathcal{W}} \leq n/q$. This is of course best for $q = 2$. Moreover, a sufficient condition to ensure embedding is

$$r \geq \lceil \log_q((q-1)\ell_{\mathcal{W}} + 1) \rceil.$$

Asymptotically, the embedding efficiency relative loss is then

$$\frac{\lceil \log_q((q-1)\ell_{\mathcal{W}} + 1) \rceil}{p},$$

which is satisfactory.

This construction is then really interesting. But to be used in practice, the recipient must know the parameter r, which is chosen during embedding. We then proposed to encapsulate our scheme in a wider one, based on ZZW construction [ZZW08], which enables the transmission of parameter r to the recipient.

These results have been presented during the French workshop on *Codes and Steganography* in January 2011. They will be published in the proceedings of the international *IMA Conference on Cryptography and Coding 2011* that will hold in December 2011 [ABF11] [3].

This first study showed the interest of such a construction. It also showed that Hamming codes behave quite well in this context. However, it would be interesting to study the behaviour of other codes, and to study how to refine the sufficient conditions that ensure the embedding.

3. This article can be found in Appendix D.

Chapter 2 – Copyright protection : what about robustness and security in the design and use of digital watermarking schemes ?

Digital *watermarking* has been introduced in the 90's as a new and complementary tool to address copyright ownership issues. Steganography and watermarking share some fundamental similarities, as both hide a message or signal in a piece of content in an *imperceptible* way for the end-user. But they address very different applications, and then have to tackle really different constraints. In steganography, the message has no relationship with the cover document that hides it, whereas in *watermarking* the message contains important information on the document, such as labels identifying its copyright owner, or its user (*fingerprinting*), or information to check the document integrity. In steganography, attacks aim at deciding if a given document contains an embedded message, whereas in digital watermarking the opponent knows that the document is protected through the embedding of some data. This leads to different imperceptibility requirements.

Here, we focus on what is called *robust* watermarking, where the embedded message or signal has to remain detectable if the watermarked piece of content is subject to some processes as lossy compression, scaling, etc. This is the case when addressing copyright ownership or *fingerprinting* issues. The principal interest of robust watermarking, when compared with traditional protection provided by cryptography or access control, is that the piece of content remains protected, even when it is used, in a clear format, by the end-user. Hence, each technology provides its own kind of security, and the best is to use them together.

Beyond imperceptibility and robustness, a third criterion has been usually considered : the *capacity*, that is, the number of symbols that can be extracted at the receiving end.

Since the first publications in the 90's, these three criteria have always been considered, and it appeared that there are some tradeoffs to address as the three of them cannot be optimized at the same time. Studies focusing on copyright protection and fingerprinting have always been driven by the improvement of *robustness*, as it is necessary in such a context, and because it was not so easy to address. Most of articles of this field deal with this criterion, presenting more and more impressive experimental assessments. Some key events in this quest were the use of spread spectrum [CMB01], the invention of re-synchronization schemes [OP98, PP99], the discovery of side information channel [CMM99, CW01], and the formulation of the opponent actions as a game [Mou01].

On the contrary, *security* received little attention in the watermarking community. The first difficulty is that security and robustness are neighboring concepts, which are hardly perceived as different. The intention behind the attack is not enough to make a clear cut between these two concepts. An image compression is clearly an attack related to robustness, but it might happen intentionally, *i.e.* with the purpose of removing the watermark, or not. *Robust* watermarking is defined in [Kal01] as a communication channel multiplexed into original content in a non-perceptible way,

and whose "*capacity degrades as a smooth function of the degradation of the marked content*". We add that the degradation is due to a classical content processing (compression, low-pass filtering, noise addition, geometric attack ...). The attacker has three known strategies to defeat watermark robustness : to remove enough watermark signal energy, to jam the hidden communication channel, or to desynchronize the watermarked content. Kalker then defines watermarking *security* as "*the inability by unauthorized users to access* [*i.e.* to remove, to read, or to write the hidden message] *the communication channel*" established by a robust watermarking. Security deals with intentional attacks whose aims are not only the removal of the watermark signal, excluding those already encompassed in the robustness category since the watermarking technique is assumed to be robust.

Some seminal works have already warned the watermarking community that digital watermarking may not be a secure primitive (*i.e.*, a tool providing information security) despite its robustness. However, they only deal with dedicated attacks relevant to particular applications. The deadlock attack concerns copyright protection and illustrates the impossibility to prevent somebody to watermark content with his own technique and key (by embedding a watermark signal or by creating a fake original) [CNBM97]. This ruins the identification of the owner because two watermarking channels interfere in the same piece of content. The collusion attack (*i.e.*, the mixing of several watermarked versions of the same content) is related to the fingerprinting application. Multiple problems in the field of copyright protection and authentication stems from the copy attack, where the attacker first copies a watermark and then pastes it in a different piece of content [KSA00]. The oracle attack is a threat whenever the opponent has access to a watermarking detector as in copy protection for consumer electronics devices [CJ98]. The attacker first estimates the secret key, testing the detection process on different pieces of content [LvD98] ; this disclosure then helps him forging pirated content. Note that in this last case, the number of detection tries is of utmost importance.

Articles proposing a complete analysis of robust watermarking security were extremely rare. It seems that the one first analysis was the pioneer work [Mit99], where two digital modulation schemes achieve perfect secrecy, and more recent works discussing the links between watermarking and cryptography [FR02], and sketching a general framework for security analysis [FD03,BBF03]. In the steps of these works, we published with Teddy Furon and François Cayre the first security analysis following such a framework. The main idea of this analysis is to adapt Shannon's definition of cryptography security to watermarking. At the beginning of the game, the watermarker selects a watermarking technique and picks up randomly a private key. According to the Kerckhoffs's principle, the opponent knows the selected algorithm but not the private key. Then, the watermarker starts producing some marked pieces of content. The opponent has access to some observations and his aim is to estimate the private key. The main idea of Shannon's theory is that information about the private key might leak from the observations [Sha49]. Hence, the *a posteriori* uncertainty of the opponent decreases as he makes more and more observations. We first studied the security in a theoretical framework, to derive the necessary number of observations needed to recover the secret key of the scheme. This was done

with the help of Shannon's information theory, but also with the help of Fisher's theory. We completed it with a practical study, setting efficient algorithms to recover the key. We focused on additive spread spectrum based watermarking schemes, and substitutive ones [CFF05b, CFF05e], as summarized in Section 2.2. Our methodology has then served the following publications dealing with the security of watermarking techniques [CPG05, BH05, PFCPG05, PFPGFC06, PFCTPPG06, PFPG07, BD07, PFPG08, PFPG09, BW09]. Algorithmically speaking, these studies rely on the use of PCA (Principal Component Analysis) and ICA (Independent Component Analysis) to recover a basis of the secret subspace where the embedding is processed [CFF05e,DD04b,BW09], clustering and classification techniques [DD04b,BW09], MLE (Maximum Likelyhood Estimation) techniques [CFF05e], and EM (Expectation Maximization) techniques [DD04b].

To stimulate research in this area, two international contests have been organized. The first one, BOWS-1 (Break Our Watermarking System), hold in 2005-2006 [PB07] [4]. The second one, BOWS-2, hold in 2007-2008 [5]. BOWS-2 was structured in three independent challenges, called *Episodes* respectively dedicated to robustness attacks, sensitivity attacks, and finally attacks based on the knowledge of the watermarking technique and the observation of a huge number pieces of content that have been watermarked with the same key. It put in to the test a very robust zero-bit watermarking technique named `Broken Arrows` [FB08], which has been design specially for the contest. These contests really motivated people to design new attacks, and showed how much attacked based on the knowledge of the watermarking technique may be dangerous, and cannot be omitted when designed a new scheme.

Hence, security is now well defined, and more often tackled as in the past. One of today's most important challenge is to derive heuristics on how to prevent such attacks during the scheme design. A few works tackled this challenge [BC06a,BC06b, CB08, MCB07, MBCM09a, MBCPG08, MBCM09b, XFF10b, MBCM10b, MBCM10a], and it appeared that we have to manage a tradeoff between security and robustness. In this context, we proposed with Teddy Furon and Fuchun Xie some studies on the robustness and security improvement of `Broken Arrows`. This work is presented in Section 2.4.

Section 2.2 – Watermarking schemes' cryptanalysis (2002-2005). My first thoughts about the links between digital watermarking and security [FR02] motivated me to study the gain an attacker could derive from the knowledge of the watermarking technique, to estimate the secret key that has been used. As mentioned above, other publications had the same intuition that such an approach would be of interest, and help us define security levels for digital watermarking [Kal01, BBF03]. Hence, at this point the door was open, and it was time to enter. With Teddy Furon and François Cayre, we decided to follow the emerging security framework and to study the security of popular spread spectrum based watermarking schemes, and substitutive ones. This study took us about two years. We followed Shannon's me-

4. `http://lci.det.unifi.it/BOWS/`
5. `http://bows2.ec-lille.fr/`

thodology [Sha49], considering the attacker may observe N_o watermarked pieces of content, all watermarked with the same secret key. He tries, from these observations, to derive some information on the secret key. We transposed usual cryptographic contexts of attacks : 1) WOA (Watermark Only Attack) when the attacker only has access to the watermarked pieces of content ; 2) KOA (Know Message Attack) when he has access to couples of corresponding original and watermarked pieces of content ; 3) KMA (Known Message Attack) when he has access to couples of corresponding messages and watermarked pieces of content. We first addressed the problem with a theoretical point of view, to set the security level of the studied watermarking schemes. Shannon's information theory concepts were sufficient to study the security levels of substitutive schemes. But we had to use Fisher's theory concepts to study the spread spectrum schemes case.

We then proved that substitutive schemes can achieve *perfect covering*, which is the highest security level (no information on the secret key is leaking from the observations), but only in the WOA context. In the KOA and KMA contexts, information is leaking. More precisely, the attacker can in theory recover the secret with $\log_2 N$ observations in the KOA case, with N the size of the message and of the key ; in the KMA case, $\log_2 n$ observations are sufficient, with n the size of the vector $\mathbf{x}$ which is extracted from the piece of content to perform embedding. Fig. 2.5 illustrates these results. Concerning additive spread spectrum based techniques, there is always information leaking about the secret key, and perfect covering is never achieved. The attacker has sufficient information to recover the key with $\mathcal{O}(N\sigma_{\mathbf{x}}^2/\gamma^2)$ observations in the WOA and KMA contexts, with N the number of secret carriers, $\sigma_{\mathbf{x}}^2$ the host signal variance, and γ the embedding strength. In the KOA context, $\mathcal{O}(N)$ observations are sufficient. We also stated that for both kinds of techniques, the key can only be exactly recovered in the KMA context. In the WOA and KOA contexts the recovery may only be achieved up to a permutation of the components of the key, and up to sign in the spread spectrum case.

We then proposed efficient algorithms to perform such attacks in practice. The substitutive case was easy to tackle, but the spread spectrum one was harder. The easiest case is KMA, for which an MLE (Maximum Likelihood Estimator) approach works very well. For KOA and WOA cases, we leaned on blind separation source techniques, namely PCA (Principal Component Analysis) and ICA (Independent Component Analysis). For the KOA case, ICA was sufficient. But for the WOA case, it could not be used as is, because of the huge size of the data ; we had in this case to design an iterative process combining ICA with MLE. These attacks were experimented on synthetic signals and real images. We tested different watermarking techniques, some of them being side informed. Our results are depicted in Figs. 2.6 and 2.7. They show the efficiency of the attack, compared with traditional robustness attacks.

This study shows that such attacks are really more powerful than traditional robustness attacks. This does not mean that the studied watermarking techniques are too weak, but this points out that keys must be changed often. Cryptographers are familiar with this fact, but watermarkers were not. One other advantage of the study is that the methodology we proposed is not dependent of the nature of the me-

dium, nor the watermarking technique. Hence, it can be applied to other techniques, as showed in [CPG05, BH05, PFCPG05, PFPGFC06, PFCTPPG06, PFPG07, BD07, PFPG08, PFPG09, BW09].

This work has been partly supported by the European Network Of Excellence ECRYPT. It has been presented at *IWDW : International Workshop on Digital Watermarking* in 2004 [CFF05b] – where it has been awarded as the *Best Paper* – and *IS&T/SPIE International Symposium on Electronic Imaging'2005* [CFF05c, CFF05d] in 2005 ; the theoretical part of the analysis has been presented at the *IEEE International Symposium on Information Theory'2005* [CFF05a]. The whole work has been published in *IEEE Transactions on Signal Processing*, in a special issue entitled *Supplement on Secure Media* [CFF05e] [6]. A book chapter has also been published [CFF07].

Section 2.3 – Combining audio watermarking and DRMs to secure music delivery on 3G mobile phones (2003-2006). In the French RNRT SDMO project, we defined an architecture dedicated to a secured diffusion of music on 3G mobile phones. This architecture leans on a proper use of the SIM card to tackle DRM's licences (Digital Right Management) issues, secured transactions (confidentiality and authentication being ensured by dedicated protocols which are compliant with the OMA – Open Mobile Alliance – specifications) between the SIM card and the final audio device that plays the music (the sound card system). Technically speaking, we used cryptographic primitives to tackle DRM's licences issues and secure the transactions, and digital watermarking. Two watermarking techniques were used : the first one is used to embed the identification number specified in the associated licence, and then establishes a permanent link between the file and the licence ; the second one embeds only one bit, just to tell that the file has been protected and has to be further checked. In charge of the security work package, I mainly worked on the articulation of all these primitives.

A prototype platform has been presented to the Ministry of Industry in May 2006. This work has been presented at some industrial meetings and a patent has been filed [GFF+06]. It has also been presented during academic conferences [ABTD+06, BTFF+06]. A journal article has also been published in *ISAST Transactions on Communication and Networking* [FDD+08].

Section 2.4 – Improving `Broken Arrows`' robustness and security : F. Xie's PhD thesis (2007-2010). As mentioned above, several articles studied the security of some watermarking techniques, whereas only few articles tried to consider this criterion during the design of the scheme itself. These works were boosted by the organization of BOWS-1 and BOWS-2 contests.

With Fuchun Xie and Teddy Furon, we focused on the zero-bit watermarking technique called `Broken Arrows`. Having being designed for BOWS-2, it provides a particularly robust embedding. BOWS-2 was composed of 3 independent challenges, called "Episodes". A. Westfeld won the first one, dedicated to robustness

6. This article can be found in Appendix D.

attacks [Wes08]. He also won the third one, which was dedicated to security attacks leaning on the observation of a huge set of watermarked pieces of content, all watermarked with the same key. The attack he used for Episode 3 is an improvement of the one he used for Episode 1. At the same time, P. Bas, one of the `Broken Arrows` inventors, thought about another security attack. They published there results together in 2009 [BW09].

With Fuchun Xie and Teddy Furon, we first improved `Broken Arrows` to make it robust to Westfeld's first attack, which acts as a denoising process. Then we improved its security to make it robust to both Westfeld's and Bas' attacks.

Let us first focus on the first improvement. To understand it, we need to explain some parts of the embedding process of `Broken Arrows`. The embedding is performed in the wavelet domain. It leans on a modulation, and a projection in a secret MCB plane. Both are parametered by the secret key. The modulation implies N_v secret carriers, and the final secret subspace is composed of N_c hypercones. Watermarked pieces of content lie in the hypercones, whereas original ones lie outside the hypercones. More information on the algorithm may be found in [FB08], and an implementation can be found on the web site of BOWS-2 contest [7]. The watermarked vector $\mathbf{y}$ is obtained by addition of the original vector $\mathbf{x}$ with a watermark $\mathbf{w}$: $\mathbf{y} = \mathbf{x}+\mathbf{w}$. The watermark is of the form $\mathbf{w} = \text{mask}\ \mathbf{s_w}$, with $\mathbf{s_w}$ the signal generated in the wavelet domain according to the secret key, and mask the perceptual mask which ensures visual imperceptibility. In the original setup, $\text{mask}_{\texttt{BA}} = |\mathbf{x}|$, where $|\mathbf{x}|$ denotes the absolute value of the wavelet coefficients of the host vector $\mathbf{x}$, chosen in all the sub-bands except the LL sub-band. To improve `Broken Arrows`'s robustness, we took into account the dependency between the neighbors coefficients during embedding. We proposed to replace each coefficient (except in the LL sub-band) by an averaging of five coefficients : itself $\mathbf{x}[k,\ell]$, and four neighbors $\mathbf{x}[k-1,\ell]$, $\mathbf{x}[k,\ell-1]$, $\mathbf{x}[k+1,\ell]$, et $\mathbf{x}[k,\ell+1]$. This gives a new perceptual mask :

$$\text{mask}_{\texttt{AWC}}[k,\ell] = \frac{1}{5}\left|\sum_{s=k-1}^{k+1}\sum_{t=\ell-1}^{\ell+1}\mathbf{x}[s,t]\right|.$$

Collecting all the local masks $\text{mask}_{\texttt{AWC}}[m,n]$, we obtain the global mask $\text{mask}_{\texttt{AWC}}$ of the new proportional embedding, that we named `BA-AWC`. Intuitively, this reinforce the dependency between the neighbors coefficients of the watermarked signal. For a given coefficient, the mask is larger than at least one of these five coefficients. Hence, the embedding of the watermark signal may modify the signs of the host coefficients, and the watermark is now not only present in the amplitudes of the coefficients, but also in their signs, and this is sufficient to counter the attack.

To estimate the efficiency of this counter-measure, we performed the same robustness experiments (same $2,000$ images, same tests) as it was performed on the original `Broken Arrows` scheme in [FB08]. Our `BA-AWC` variant is a little less robust than the original `Broken Arrows` for some of the attacks, and preserve an equivalent robustness for the other ones. In any case, the robustness of `BA-AWC` is almost excellent.

7. `http://bows2.ec-lille.fr/`

In his attack, Westfeld used the 10,000 images of BOWS-2 contest (including the 2,000 mentioned above) [Wes08]. Nevertheless, we performed our tests only on the set of 2,000 images used to evaluate the robustness of the scheme. This is the reason why our results concerning the performance of Westfeld's attack on `Broken Arrows` differ a little from Westfeld's ones. But this difference is really not significant, and the behavior of the attack is the same. The PSNR (Peak Signal to Noise Ratio) of the attacked images varies from 19.9 to 46.2 dB (it varies from 19.7 to 45.0 dB in [Wes08]). Fig. 2.8 shows the drop of the percentage of images that have been attacked with success, according to their PSNR. For the original `Broken Arrows`, Westfeld's attack is really terrible, with a 100% success for attacked images whose PSNR is lower than 30 dB. Moreover, even if its efficiency decreases when the PSNR increases, it remains extremely efficient for about 40% of the images when the PSNR is around 35 dB. Our variant `BA-AWC` resists really well to the attack, since the percentage of successfully attacked images remains near 0, whatever the PSNR.

Hence, `BA-AWC` is a good plug-in as it offers a general robustness which is almost the same than the original `Broken Arrows`, whereas it is really more robust to Westfeld's attack.

These results have been presented at *IS&T/SPIE International Symposium on Electronic Imaging'2009* [CXFF09] and during the French GRETSI 2009 [XFF09]. They have been supported by the French ANR-RIAM ESTIVALE and MEDIEVALS projects.

Let us now discuss the other improvements we proposed to reinforce `Broken Arrows` security. Westfeld's second attack was based on his first attack, which is here completed by a clustering step. As our variant `BA-AWC` counters efficiently Westfeld's first attack, it should also counter this second attack. Our experiments confirm this assertion. To measure the clustering efficiency, we used the AMI (Ajusted Mutual Information). The higher the AMI, the better the attack. For the original `Broken Arrows`, this AMI is equal to 0.7; this means that the estimated clusters almost coincide with the real clusters, and that the attack performs well. But this classifier does not behave well for our variant `BA-AWC`, as the AMI is then lower than 0.05. Our variant `BA-AWC` is then sufficient to counter Westfeld's attack.

Bas' subspace estimation attack is really different. It leans on the following fact : `Broken Arrows` embedding is considerably modifying the power distribution of the signal in the secret subspace. Using PCA techniques, Bas showed how he is able to recover the secret subspace. Hence, to counter it, we tried to find a way to limit this modification, in order to get a more uniform distribution. We proposed to use a normalized Square Chordal Distance to measure the distance between the secret subspace and the estimated subspace. The smaller the distance, the better the attack : the distance equals 0 if both subspaces coincide (the attack is successful), and equals 1 if they are orthogonal to each other (the attack fails). We then tuned the parameters of `Broken Arrows`, to make the distance increase. We chose $N_v = 1024$ and $N_c = 256$, whereas in the original scheme the parameters where $N_v = 256$ and $N_c = 30$. Our experimental results, depicted in Fig. 2.9, showed that for the original `Broken Arrows` with original parameters SCD_{norm} decreases quite

fast when the number of observations increases. This confirms the results of [BW09], and the choice of SCD_{norm} as a measure. However if we tune the parameters to be $N_v = 1024$ and $N_c = 256$, the distance decreases very slowly : even after 3.10^4 observations, SCD_{norm} is still greater than 0.9, meaning that the attack does not work anymore. This is true when tuning the parameters of the original `Broken Arrows`, but also if we tune the parameters of the `BA-AWC` variant. This latter leads to a new variant which is called in the sequel `BA-AWC+`. We then checked the generic robustness of these new variants, with the same benchmark as in [FB08] (and as before when we estimated the generic robustness of `BA-AWC`). This showed that the generic robustness is not so good as with the original parameters, but still remains sufficient. Hence, tuning the parameters globally improves the scheme's security, without sacrificing the robustness. However, we must say that the time needed for embedding is multiplied by a factor of 4.

These results were satisfactory. However, we pushed the study further, to prevent future attacks. We focused on attacks that could be based on second order statistics, and proposed a way to force the corresponding distributions to be as uniform as possible. Without describing our study in detail, it encompasses : 1) the introduction of a security criterion ; 2) an embedding process which aims at maximizing the robustness, under constraints on imperceptibility and security ; 3) a watermark detection based on a *a contrario* decision criterion. We then formalized `Broken Arrows` security relatively to such attacks, and derived a new variant, called here `BA-AWC++`, which is more secure than the previous ones. However, we noticed that this security improvement leads to a real loss in robustness, and to a much higher false alarm probability. But, despite this loss in robustness and this high false alarm probability, this scheme has some serious potential in some watermarking applications, especially in multimedia *fingerprinting*, as it is discussed in Section 3.2. In such a scenario, all the distributed pieces of content are watermarked, hence the probability of false alarm is no longer important. The question is more about the symbols likely to be hidden in the pirated copy. As far as we know, any watermark detector outputs binary decision about the presence or absence of embedded symbols. The *a contrario* decision test can indeed provides a probability of the presence of a given symbol, that is, a soft output, leading to a really interesting information for the accusation process.

These results have been presented at *IS&T/SPIE International Symposium on Electronic Imaging'2010* [XFF10a] and *ACM Multimedia & Security, MM&SEC'10* [XFF10b]. They have been supported by the French ANR-RIAM MEDIEVALS project.

This work leads to some remarks. First, taking into account this particular security constraint in the embedding guarantees that the scheme is only secure against second order statistics analysis tools. However, some high order statistics might leak information on the secret space. Therefore, the issues now turn to be how many contents and computing power a high order analysis requires to work accurately. We believe that it is significantly more demanding. Second, despite its poor trade-off probability of false alarm vs. robustness, we believe that this scheme has some serious potential in the images (or video) watermarking applications, especially in

multimedia *fingerprinting* (see next chapter), since the contents are all watermarked in this scenario the probability of false alarm is no longer a problem. The question is more about the symbols likely to be hidden in the pirated copy. As far as we know, any watermark detector outputs binary decision about the presence or absence of the watermarks (this includes potential multiple detections). The *a contrario* decision test can indeed provide a probability of the presence of a given symbol, as a soft output, bringing more information for the anti-collusion code accusation step.

Chapter 3 – Document traceability : how to trace fraudulent usage of documents with the help of anti-collusion codes ?

We are interested here with the ability for a content provider to trace illegal redistribution of the pieces of content he is in charge of. This may be called (active) *fingerprinting*, or sometimes traitor tracing, copy serialization, transactional watermarking, or forensics. The addressed problem is the following : a content provider distributes personal copies of the same content to n different users. Some c dishonest users, called *colluders* in the sequel, mix their copies to forge a pirated content they will illegally redistribute. An accusation process aims at tracing back the colluders' identity by analyzing this pirated content. A *fingerprinting code*, or *anti-collusion code* is a set of n different sequences $\{\mathbf{X}_j\}_{j=1}^{n}$ of length m. Each sequence, identifying a specific user, has to be hidden in his/her personal copy with a watermarking technique. When a pirated copy is found, the analysis of the sequence Y it contains should lead to at least one of the collusion members.

Since the seminal papers published by Wagner [Wag83] and Blakley *et al.* [BMP86], a lot of publications tackled the issue of designing appropriate anti-collusion codes. Several attack models and design strategies were explored. The most studied attack model comes from the pioneering work of Boneh and Shaw [BS95, BS98], and an extension of [GP00]. These models, which may at a first glance look more theoretical than practical, can be effective if we proceed like this : we cut the piece of content into blocks, and hide in each block one component/symbol of the user's fingerprint $\mathbf{X}_j$. Personalization can then be organized as a quick process that dynamically "combines" master copies of the content, each of them hiding the same symbol in each block, as depicted in Fig. 3.3. The embedding step is then processed off-line, the on-line step only consisting of the combination of the blocks that correspond to the good fingerprint. When the colluders compare their copies of the content, they will proceed block by block. They will see that some positions their blocks are identical, and we suppose here that in this case they will keep the block as is : this assumption is called the *marking assumption*. When the blocks differ, they will understand that some personal information is hidden in these blocks, and they will construct a pirated block by mixing their blocks. Let us denote by $\mathbf{Y}$ the vector that is extracted from the pirated content. Such a scenario precisely corresponds to the original model of Boneh and Shaw denoted here as (a)+(b), and its extension here (a)+(b′) : (a) the marking assumption is respected (if all the colluders' blocks have the same symbol embedded, the symbol in the pirated block is the same as in the colluders' blocks) ;

(b) the colluders put one of their blocks as is (the symbol in the pirated block is equal to the symbol of a colluder's block) ; (b′) the colluders mix their blocks to produce a new block, which has to be perceptively similar, but is different from any of their blocks (the symbol in the pirated block may be any symbol, then eventually leading to errors, but may also be erased). Fig. 3.2 illustrates the models at the fingerprints level, as considered by Boneh and Shaw. Figs. 3.5, 3.6 and 3.7 show what is going on at the content level, and the consequences at the fingerprints level. Model (b) corresponds to copy/paste attacks, or block exchange attacks, and is depicted in Fig. 3.5. Model (b′) is depicted in Fig. 3.6 corresponds to fusion attacks, as for example the averaging attack (each pixel of the pirated image is obtained by averaging the pixels of the colluders' images). Fig.3.7 depicts usual signal processing on the blocks, that is, usual robustness attacks that can be performed by any user : lossy compression, scaling, etc. Of course, the more robust the watermarking technique, the more reliable the sequence $\mathbf{Y}$, and then the more efficient the accusation process.

It has been shown that a *strong* traceability, which never accuses an innocent user, cannot be used in practice, as it leads to anti-collusion codes whose words cannot be embedded by today's watermarking techniques ; they are too long, and defined over too large alphabets [HvLLT98, BCE+01, SSW01]. Moreover it is limited to Model (a)+(b), which is not the most realistic one. Hence in practice we have to deal with a *weak* traceability, which may sometimes accuse innocent users. Such traceability is achievable with anti-collusion codes that are compatible with actual embedding techniques, and is achievable in the realistic (a)+(b′) model. In 2003, Tardos [Tar03, Tar08] and Peikert *et al.* [PSS03] published independently a lower bound on the length m of anti-collusion codes achieving weak traceability, for a given probability ε to accuse an innocent user, and facing a collusion of a given size c. Moreover, Tardos – who was unknown to the fingerprinting community – also provided in his article a construction of a binary probabilistic code, whose length achieves this lower bound. Hence, its length order is optimal. Moreover, it is easy to understand and implement, and really powerful to trace users in the (a)+(b′) model, for practical collusion sizes. This breakthrough has appealed a numerous amount of research works, which are not detailed here : among them, some refined the bound on the code length, and some improved the accusation process, *e.g.* [HHI06, NHWI07, NFH+07, ŠVCT08, ŠKC08, FGC08, CXFF09, FPFGC09, Nui10, MF11a].

Without going too deep into the details, Tardos' idea [Tar03, Tar08] was to use as fingerprints some binary codewords of length m of the form $Ac^2\lceil \ln(1/\varepsilon_1)\rceil$ drawn at random according to a particular distribution $\mathbf{p}$, which is half-sparse half-dense. This distribution is kept secret by the content provider. If a pirated content is caught, the content provider computes for each suspected user a score $S_j = \sum_{i=1}^{m} g(Y_i, X_{ji}, \mathbf{p}_i)$ that reflects the probability for User j to be a colluder. If the score is greater than a given threshold Z of the form $Bc\lceil \ln(1/\varepsilon_1)\rceil$, then User j is considered as guilty ; if not, he is considered as innocent. Tardos' construction ensures that the probability for a given innocent user to be accused is less than a given parameter ε_1, whereas the probability that the scheme cannot accuse any user is less than a given $\varepsilon_2 >> \varepsilon_1$. Škorić *et al.* showed in [ŠKC08] how to refine the lower bound on the code length by slightly modifying the $g()$ function, to force it to have a symmetric behavior (Tardos'

one did not). Moreover, they derived a construction in the q-ary case. In all the following papers, this is this "symmetric" version which is implicitly considered. Precise formulas are given Page 70 for both binary and q-ary cases, with this "symmetric" scores, and the binary case is also sketched in Fig. 3.8. Škorić *et al.* also showed in [ŠKC08] that with this setup colludes' scores and innocent users' scores follow Gaussian distributions, as illustrated by Fig. 3.9.

Section 3.2 – Improving Tardos codes accusation process to combine them with `Broken Arrows` watermarking scheme : F. Xie's PhD thesis (2007-2010). We saw in the introduction that a fingerprinting scheme is based on a watermarking technique and an anti-collusion code. So far, the designs of these two technologies have often been made separately. Fingerprinting codes have been mostly proposed by the cryptography and coding community, whereas watermarking techniques have been mainly studied by people in the image or signal processing community. But the choice of the watermarking technique has an important impact on the reliability of the sequence $\mathbf{Y}$, which determines the behavior of the accusation process of the anti-collusion code. With Fuchun Xie and Teddy Furon, we designed a complete scheme based on the combination of the very robust watermarking technique called `Broken Arrows` (see Chapter 2 above), and the famous Tardos anti-collusion code.

Our ambition was to tackle fusion attacks, *e.g.* averaging attacks, which are easy to perform and usually critical for the anti-collusion code. To do so, we used a q-ary Tardos code, and adapted `Broken Arrows` to enabled the embedding of q-ary symbols. It appeared that `Broken Arrows` is so robust it can detects several symbols at the same time in case of an averaging attack. Hence, we modified Tardos' scores computation to use all the additional information provided by this multiple detections.

Let us just recall that in [ŠKC08] the score is computed like this :

$$\begin{aligned} S_j &= \textstyle\sum_{i=1}^{m} g(Y_i, X_{ji}, \mathbf{p}_i) \\ \text{taking } g(Y_i, X_{ji}, \mathbf{p}_i) &= \delta_{Y_i = X_{ji}} g_1(p_i^{Y_i}) + (1 - \delta_{Y_i \neq X_{ji}}) g_0(p_i^{Y_i}) \\ \text{with } g_1(p) &= \sqrt{(1-p)/p} \text{ and } g_0(p) = -\sqrt{p/(1-p)}. \end{aligned}$$

Let us denote by $\mathcal{Y}_i = \{Y_i^1, \ldots, Y_i^{L_i}\}$ the set of extracted symbols in the i-th block. Our first proposition was to keep function g and to set :

$$S_j = \sum_{i=1}^{m} \sum_{\ell=1}^{L_i} g(Y_i^\ell, X_{ji}, \mathbf{p}_i) \tag{3.1}$$

It is almost like if the code length had been increased from m to $m\bar{L} = \sum_{i=1}^{m} L_i$.

Our second proposition was to keep the classical score computation sum, but to set :

$$g(\mathcal{Y}_i, X_{ji}, \mathbf{p}_i) = \delta_{X_{ji} \in \mathcal{Y}_i} g_1(p_i^{\mathcal{Y}_i}) + (1 - \delta_{X_{ji} \notin \mathcal{Y}_i}) g_0(p_i^{\mathcal{Y}_i}) \tag{3.2}$$

with $p_i^{\mathcal{Y}_i} = \sum_{\ell=1}^{L_i} p_i^{Y_i^\ell}$. In this case, the variance of the colluders scores is decreased : whatever their symbol $X_{ji} \in \mathcal{Y}_i$, they get the same accusation term $g_1(p_{\mathcal{Y}_i})$.

Both propositions were experimented on real images. Parameters used were $m = 300$, $q = 4$ and $c = 20$. Parameter κ, involved in the shape of the secret distribution $\mathbf{p}$, was taken between 0.1 and 0.5. Our statistical results are based on the scores of $32,000$ innocent users, and $8,000$ guilty users. We compared the efficiency of our setup when facing a fusion (averaging) attack, with the efficiency of the [ŠKC08] setup when facing a copy/paste attack. Results showed that the expectations μ_I of innocent users' scores are null in the three cases, as attended. The expectations μ_C of the colluders' scores are quite similar for our both propositions, and greater than the one obtained with [ŠKC08]'s setup. Moreover, our variants provide smaller variances for the innocent users' scores as well as for the colluders' scores, when compared to [ŠKC08]'s setup. Hence, our setups seem really more reliable than the classical approach. Measuring the Kullback-Leibler distance between the innocent users' scores' distribution and the colluders' scores distribution confirmed this assertion, as illustrated by Fig. 3.10. Our setups lead to a better distinction between innocent users and colluders. This means that with our setup, if the colluders perform an averaging attack, thinking it will be more efficient than a simple copy/paste attack, they misbehave : doing this, they will be caught more efficiently than for a simple copy/paste attack (which is already really well tackled).

These results have been presented at *ACM Multimedia & Security, MM&SEC'08* [XFF08], and during the French GRETSI 2009 [XFF09]. They have been supported by the European Network Of Excellence ECRYPT, and the French ANR-RIAM ESTIVALE project.

This work showed, in some way, that Boneh and Shaw's extended model has to be improved to fit practical issues as multiple detection. Following our steps, Škorić *et al.* pushed this idea further, and finally proposed a formalization of such a model, called the *Combined Digit Model* [ŠKSC11]. They also showed that within this realistic and more complete model, our improvements of Tardos accusation process behave really well.

These results have been integrated in the `FANTOMAS` platform[8], developed by Mathieu Desoubeaux between 2007 and 2009. `Broken Arrows`, which was originally designed to watermark still images, has been adapted to video processing. `FANTOMAS` enables the personalization of entire movies. It also includes an accusation process to track colluders. We also integrated in the platform results from [CFG08], which provide a precise experimental estimation, for each user, of the probability to falsely accuse him/her. These real probabilities are far below the ε_1 upper bound for the three highest scores (we get a false alarm probability near 10^{-25} for the highest score, for $\varepsilon_1 = 10^{-3}$). Fig. 3.11 provides a screen shot of the accusation results for an averaging attack.

In parallel, we proceeded with Teddy Furon and Fuchun Xie to the integration in our experiments of the improved variants of `Broken Arrows` that have been presented

8. `http://www.irisa.fr/temics/demos/BrokenArrows/index.php`

in Section 2.4 and published in [CXFF09,XFF10a,XFF10b]. These synthesis has not been published, but is detailed in the PhD manuscript of Fuchun Xie [Xie10]. Fig. 3.12 illustrates these results. We see that Eq. (3.2) offers a more reliable accusation in practice than Eq. (3.1). We also see that both provide better results than the classical scores computation. The `BA-AWC+` embedding technique (which is `BA-AWC` tuned with parameters $N_v = 1024$ and $N_c = 256$) is the one that provides the best accusation, while resisting any known attack (including Westfeld's and Bas' attacks [BW09]). When using `BA-AWC++` (which is more secure but less robust), the accusation is still more efficient with our new accusation scores than with the classical ones, but the gain is not so good as with `BA-AWC+`. Nevertheless, as we discussed in Section 2.4, the *a contrario* decision test can in this case provide a probability of the presence of a given symbol, as a soft output, bringing more information for the anti-collusion code accusation step.

Section 3.3 – Dynamical optimization of Tardos codes accusation process : A. Charpentier's PhD thesis (2008-2011). In [FGC08], Furon *et al.* investigated Tardos choices in the binary case, to understand why he chose such functions f (to shape the secret distribution $\mathbf{p}$), $g()$ (to compute the scores), etc. Optimizing the functions to get the highest expectation for the colluders' scores, under Tardos' assumptions, they obtained the same functions as in [Tar03, Tar08, ŠKC08]. Hence, whereas Tardos did not explain his choices, they are now well understood. In the same paper, Furon *et al.* discussed some of the assumptions. One assumption in particular was discussed. Tardos code was designed to face equally any attack satisfying the Marking Assumption (a). Under this assumption, [Tar03, Tar08, ŠKC08]'s accusation function $g()$ is the best one. But what if we relax this assumption ? Furon *et al.* showed than when the collusion size and the colluding strategy are known to the content provider, the function $g()$ can be optimized to provide a more reliable accusation.

With Ana Charpentier and Teddy Furon, we also focused on the binary case and assumed that the colluders apply the same attack strategy to any block/component of the medium. Under this assumption, we proposed an optimization of the accusation function $g()$ which is more efficient than the one of [FGC08]. Moreover, we proposed a way to dynamically estimate the colluders' size and strategy, which is necessary to use such an optimization in practice. We then proposed an iterative accusation process : the estimation of the colluding strategy leads at each iteration to an optimization of the accusation function, which leads to a new estimation of the collusion, with the help of an Expectation-Maximization technique ; then, the estimation of the collusion leads to a new estimation of the colluding strategy, etc, as depicted in Fig. 3.3.

1. Initialization : we computed the scores with [ŠKC08]'s accusation function.
2. Expectation-Maximization, based on these scores and the expectations and variances of the innocent users' scores and colluders' scores, provides for each score S_j a probability $\hat{T}_j$ for it to correspond to a colluder.
3. Estimation of the collusion's size and strategy. The collusion size was estimated as $\hat{c} = \lceil \Sigma_{j=1}^{n} \hat{T}_j \rceil$. We then considered that the colluders correspond to the $\hat{c}$

highest $\hat{T}_j$. Based on this estimation of the collusion, we estimated the colluders' strategy. We modelled the colluders' strategy through the set of probabilities $\theta = \{\mathbb{P}(Y_i = 1|\Sigma_i = \sigma_i), \sigma_i = 0..c\}_{i=1..m}$, the random variable $\Sigma_i = \sum_{j \in \mathcal{C}} X_{ji}$ corresponding to the number of colluders' fingerprints whose i-th component is equal to 1.

4. The new accusation functions are obtained through an optimization under constraints, for a given estimated collusion $\hat{\theta}$. We denote by μ_I and σ_I^2 (resp. μ_C and σ_I^2) the expectation and variance of the distribution of the innocent users' scores (resp. colluders' scores). By construction of the code, the fingerprints' components are i.i.d from index to another. Hence, according to the score computation formula, the statistics of the scores are linear with m : $\mu = m\tilde{\mu}$, $\sigma^2 = m\tilde{\sigma}^2$. Accusation function optimization's goal is to maximize, through a Lagrangian, the Kullback-Leibler distance $D_{KL}(\mathcal{N}_C, \mathcal{N}_I)$ between the colluders' scores and the innocent users' scores distributions, with $\mathcal{N}_C = \mathcal{N}(\mu_C, \sigma_C^2) = \mathcal{N}(m\tilde{\mu}_C, m\tilde{\sigma}_C^2)$ and $\mathcal{N}_I = \mathcal{N}(\mu_I, \sigma_I^2) = \mathcal{N}(m\tilde{\mu}_I, m\tilde{\sigma}_I^2)$. The optimization constraints are : $\tilde{\mu}I = 0$, $\tilde{\sigma}_I = 1$, and any two innocent users have independent scores. We can write

$$D_{KL}\left(\mathcal{N}_C, \mathcal{N}_I\right) = \frac{1}{2}\left(m\tilde{\mu}_C^2 - \log(\tilde{\sigma}_C^2) + \tilde{\sigma}_C^2 - 1\right).$$

These are the same constraints as in [FGC08], except that no consideration is paid to the variance of the colluders' scores (which was set to $\tilde{\sigma}_C^2 = 1$ in [FGC08]). Moreover, we relaxed the assumption on the independence between the variance of the colluders' scores and the colluders' strategy. We obtained explicit formulas to set the optimal accusation function $g()$, and to express the colluders' scores expectation, as shown in Fig. 3.14.

5. Based on these optimized accusation function, new scores were computed, which led to a new iteration of the process, at Step 2.

To compare our optimization with the one of [FGC08], we proceeded to the same experiments. Our experiments hence addressed only the anti-collusion code, not a complete embedding scheme. We considered the same copy/paste strategies as in [FGC08] : *Uniform, Majority, Minority, All1, All0.* To estimate the accusation function optimization efficiency, we computed the ratio $cm\tilde{\mu}_C/\tilde{\sigma}_I$. We compared the values obtained with our optimized function with the results of [FGC08]. This comparison is summarized in Table 3.1. Our accusation function is, as attended, more efficient when the colluders' strategy match the estimated strategy (*i.e.* the one we considered in the optimization). Results also show that our accusation function is always better than [FGC08]'s one, even when the real and estimated strategies do not match.

To estimate the efficiency of our iterative process, as well as the efficiency of our estimation of the collusion's size and strategy, we considered a code length which is too small for the usual accusation process to succeed. Fig. 3.15 summarizes our results for $c = 8$. It shows on Subfigure (a) the accusation efficiency of the classical setup [ŠKC08], which is not really efficient here because of the too small code length.

Subfigures (b), (c) and (d) illustrate the efficiency of our iterative accusation process : (b) exact values of c and $\hat{\theta}$ are provided to the optimization, (c) the exact value of c is provided, but $\hat{\theta}$ is estimated, and finally (d) which corresponds to our complete iterative process, which automatically estimates the parameters. We see that the iterative process is efficient, and leads to a better accusation than the classical setup, and compensates the short length of the code. For some strategies, we are able to accuse all the colluders, but for some others we are not able to do so. We did not understand why such a difference between the strategies. We also noted that our estimation of c through Expectation-Maximization is not accurate, and is often much larger than the real number of colluders. Then, it is safer to accuse the highest score than to accuse all the c highest ones.

These results have been presented at *IS&T/SPIE International Symposium on Electronic Imaging'2009* [CXFF09], and at the French GRETSI 2009 [CFF09] and *Journées C2 du GDR IM* in 2009. They have also been published in the French journal *Traitement du Signal* in 2010 [CFF10]. They have been supported by the French ANR-RIAM ESTIVALE et MEDIEVALS projects.

Note that Furon and Pérez-Freire proposed later a better estimation of the colluders' strategy [FPF09a].

Section 3.4 – Designing an asymmetric fingerprinting protocol based on Tardos codes : A. Charpentier's PhD thesis (2008-2011). With Ana Charpentier, Teddy Furon and Ingemar Cox, we worked during the last two years on the design of an *asymmetric fingerprinting protocol* dedicated to Tardos codes.

While some researcher were trying to design efficient anti-collusion codes, some other ones considered the scenario where the Provider is untrustworthy, and proposed protocols to address this issue. Given knowledge of a Buyer's fingerprint, the Provider creates a pirated copy of a piece of content, implicating the innocent Buyer. To prevent this, Pfitzmann *et al.* [PS96] first introduced the concept of *asymmetric fingerprinting protocol*[9], in which the Provider does not need to know the Buyer's fingerprint. The Buyer first commits to a secret (the fingerprint) that only he/she knows. The Buyer and Provider then follow a protocol which results in the Buyer receiving a copy of the Work with his/her secret fingerprint (and some additional information coming from the Provider) embedded within it. The Provider does not learn the Buyer's secret, and cannot therefore create a forgery. Unfortunately, the early implementations of this concept were not practical due to the very long length of the collusion resistant codes.

The advent of Tardos codes has reduced the length of the collusion resistant codes to a practical size. However, generation of these codes depends on a secret distribution $\mathbf{p}$, that is only known to the Provider. This implies that traditional asymmetric fingerprinting protocols cannot be used as is for Tardos codes. The first reason is that the Buyer has to generate a fingerprint coherent with this distribution $\mathbf{p}$ his does not know. Hence, designing such a protocol we have to tackle this particular

9. Be careful, here "asymmetric" is used in a very different way as in cryptography or watermarking, and has no link with the "symmetric" accusation Škorić *et al.* proposed for Tardos codes.

issue. Moreover, as we showed with Ana Charpentier, Teddy Furon and Ingemar Cox, the Provider may cheat during the accusation process by providing a biased $\mathbf{p}$, different from the one used during the fingerprints generation, in order to falsely accuse an innocent Buyer. Fig. 3.16 shows that using a different distribution during the accusation process makes all the scores increase, putting some innocent users' scores beyond the threshold. It then illustrates the power of such an attack. We then have to address, in the design of a protocol based on Tardos codes, the integrity of the secret distribution $\mathbf{p}$ from the fingerprints generation to the accusation.

Hence, we designed a protocol which is dedicated to binary Tardos codes, and addresses all the critical issues we thought of. These issues are summarized below.

1. The Buyer generates his/her fingerprint, with the help of the Provider. Here, there are two issues. First, the Buyer has to follow some recommendations of the Provider, which sometimes encompass some secret parameters of the Provider. Provider must be able to verify that the Buyer did not cheated, without knowing the whole generated fingerprint.
2. The Provider must deliver to the Buyer the correct piece of content, hiding the right fingerprint. But the Provider must never access the whole fingerprint of the Buyer, and the Buyer must never access the original piece of content (without the fingerprint).
3. When the Provider suspects a forgery, he run the accusation process. We must check at this point that the Provider did not cheat during this crucial step.

Our protocol is too technical to be presented here in details. Here is a sketch of the solutions we proposed to tackle these issues.

The most difficult issue was the fingerprint generation (Point 1.), because the distribution $\mathbf{p}$ must not be revealed to the Buyer, but is needed to help him/her generate his/her own fingerprint. More precisely, the Buyer has, for each of his/her fingerprint component, to pick a bit, with $\mathbb{P}(X_{ji} = 1) = 1 - \mathbb{P}(X_{ji} = 0) = p_i$, but without knowing p_i. We solved this problem with the help of an *Oblivious Transfer protocol*. An Oblivious Transfer protocol is a cryptographic primitive [Rab81] which helps Bob to pick at random k elements in a list of N elements possessed by Alice, in such a way that : 1) Bob can only get elements that really belong to Alice's list ; 2) Bob gets no information on the elements he did not pick ; 3) Alice does not know which elements were picked by Bob. We transformed each real number p_i in a list of N bits which contains L_i '1', p_i being as close as possible to L_i/N. This list is the one used in the Oblivious Transfer protocol, the Provider acting as Alice, while the Buyer is acting as Bob. The list is permuted in a different way for each new transaction/Buyer. Picking an element at random in the list provides a bit, with a probability for this bit to be 1 really close to p_i. This protocol is repeated independently for each component of the fingerprint to generate. We showed that for security reasons N must be at least equal to 100. We also showed that the larger the collusion, the smaller the impact of the quantification of p_i on the scores. We discussed the appropriateness of the Oblivious Transfer protocols provided in the literature, and showed when needed how to adapt them so that they perfectly fit our purpose. Most of them come from the cryptographic community [NP99, CT05, GH07] (and ensure a

more formal security), whereas a few others come from very different communities, and rely on different approaches, as the *Commutative Encryption Scheme* [BDF01, HC05, WZW03]. This latter has been less studied, and its security has been less formalized, but it fits our constraints better. For these reasons, we detailed both solutions, each of them having interesting advantages. And I plan to work in the future of the enhancement of the security of Commutative Encryption Schemes, to make them achieve a semantic security level, which is essential in our case and is not achieved by the schemes presented in the literature.

Once the Buyer generated his/her fingerprint, the Provider would like to check that he/she did not cheat, and that the fingerprint's components follow the secret distribution $\mathbf{p}$. But the Provider must never access the whole fingerprint. However, he must be able to run a pre-accusation process on a partial fingerprint, in order to decide if the corresponding Buyer has to be considered as a suspect. If such, the Provider may forward to the Judge for final verification with the whole fingerprint. For these reasons, we use Oblivious Transfer at this step to make the Buyer reveal some components of his/her fingerprints. Hence, the Buyer cannot decide, and does not know, which components will be revealed to the Provider. We pay attention to the fact that both parties only know what they have to. This closes Point 1.

Point 2. concerns the fingerprint embedding in the piece of content, with the help of a watermarking technique. But the original piece of content is only known to the Provider, whereas the fingerprint is only known to the Buyer. To tackle this issue, we lean on state of the art watermarking schemes, which rely on homomorphic encryption [Kur10, DBPP09]. With such schemes, the Provider is able to embed an encrypted fingerprint, and the Buyer gets at the end, after decryption, a piece of content with his/her fingerprint embedded in clear.

Point 3. is tackled with the help of a *WORM* (Write Once Read Many) memory. We force the Provider to proceed to a commitment of an encrypted version of the lists used to quantify $\mathbf{p}$'s components. These commitments enable the verification of the integrity of the secret distribution at any time, without revealing its value. This commitment is stored in the WORM memory.

All the critical issues we thought of were addressed in our protocol. And the protocol has been as described as precisely as possible.

These results have been presented last Spring at *International Workshop on Information Hiding, IH'11* [CFFC11] [10], and we are currently working on a journal version of this article. This work has been supported by the French ANR-RIAM MEDIEVALS project.

I found this work particularly challenging and exciting. Several points can be improved in future work : Commutative Encryption Schemes security should be enhanced ; the global protocol security should be formalized and properly proved ; one should study how to modify it to provide an *anonymous*, and evenly *private* fingerprinting protocol, which is a sub family that provides anonymity and evenly privacy on the transactions.

10. This article can be found in Appendix D.

Conclusion and further work.

This manuscript summarizes my contributions in the field of multimedia content protection.

Problems related to the design and analysis of steganographic schemes have been presented in Chapter 1. Focusing on the use of error correcting codes in the design of such schemes, we proposed two ways to ensure embedding success, while preserving as stealthiness and embedding efficiency as possible.

Hence, we discussed the security of robust watermarking schemes in Chapter 2. We showed that robustness is not sufficient to face really malicious attackers, and that security should be studied in its own way. We then introduced a methodology which greatly helps to formalize and analyze security. We used it to analyze properly the security of substitutive schemes, and of additive spread spectrum based schemes. We then studied how to enhance both robustness and security of a particular watermarking scheme, `Broken Arrows`.

Then, with the help of this very robust embedding techniques, and its variations, we designed in Chapter 3 a complete (active) fingerprinting scheme. This scheme also involves Tardos famous anti-collusion codes, and we showed how this `Broken Arrows`-Tardos association can counter a very easy but dramatic attack : the averaging attack. For this purpose, we sightly modified the accusation process of Tardos codes, to take into account multiple detections. In this chapter, we also showed how the accusation process of Tardos codes can be optimized when the colluders' forgery strategy is known. This led to the design of an automatic optimization process, that automatically estimate the collusion' size and strategy, and then optimizes the scores' computation. In a third contribution, we showed how Tardos codes can be encapsulated in an asymmetric fingerprinting protocol, which ensures that neither the Buyer nor the Provider may cheat the other.

These results lead to really interesting further works, which have already been presented in the chapters' conclusions. The following ones are those I find the most interesting and motivating.

On one hand, the use of error correcting codes in steganography lead to many open questions. For example, q-ary codes and non-linear codes have not been sufficiently considered, and may lead to interesting schemes. But the problem I would like to tackle first is to deeply study the new construction we proposed with Daniel Augot and Morgan Barbier. We already have many ideas related to it that should be properly studied.

On the other hand, Tardos codes have been studied a lot these last years, especially in the binary case. Nevertheless, some open questions remain. While designing our asymmetric fingerprinting protocol based on Tardos code, we saw that some non-usual Oblivious Transfer protocols have been proposed outside of the cryptographic community, under the name Commutative Encryption Schemes. This alternative construction is of particular interest in our context. Unfortunately, such published schemes' security is not sufficient to fit our needs. However, it is certain that new

Commutative Encryption Schemes based on public-key cryptographic encryption can be proposed, to achieve higher semantic security levels. Another extension of this work is to propose an anonymous, and evenly private, fingerprinting protocol based on Tardos codes. A third extension is to prove its security according to formal security models used to formally analyze such protocols.

Appendix A – Other contributions.

This appendix summarizes my other contributions.

Watermarking : a public automated platform for watermarking algorithms benchmarking (1999-2001). In November 1997, Fabien Petitcolas published the first version of `StirMark`. It was a generic tool for simple robustness testing of image watermarking algorithms. It introduced random bilinear geometric distortions to desynchronize watermarking algorithms. This tool was really popular in the watermarking community, and several versions followed, improving the original attack but also introducing a longer lists of tests. In January 1999, with Fabien Petitcolas and Frédéric Raynal, we discussed the urgent need for fair evaluation procedures for watermarking systems and a first benchmark was made possible with the release of `StirMark 3.1`.

The natural extension to this work was an automated independent public service with extended evaluation profiles to evaluate quickly watermarking libraries. This was the goal of the `StirMark Benchmark` Service project. Other researchers then joint the project : Nazim Fates (Microsoft Research Lab, United Kingdom), Jana Dittmann and Martin Steinebach (German National Research Center for Information Technology, Germany). We finally proposed a quite large set of tests, distributed as open source libraries [11]. They address different kinds of media : still images, video and audio files.

Note that other projects were developed in parallel with the same goal, as the European project `Certimark` [12], or `Checkmark` [13]. These three projects are complementary as they do not provide exactly the same tests, concern different media types, and are implemented in different languages.

This work has been presented at *IS&T/SPIE International Symposium on Electronic Imaging'2001* [PSR+01] and *International Conference on Information Technology : Coding and Computing, ITCC 2001* [SPR+01]. It has been published in the French journal *Traitement du signal* [RPF01], and in a book chapter [PF04].

11. `http://www.petitcolas.net/fabien/watermarking/stirmark/` and `http://wwwiti.cs.uni-magdeburg.de/~alang/smba.php`.
12. `http://www.certimark.org/`
13. `http://cvml.unige.ch/ResearchProjects/Watermarking/Checkmark/`

Watermarking and Compression : dirty paper coding with partial state information : a joint watermarking-compression approach (2006-2008). Let us consider a message which is embedded in a piece of content with the help of a robust watermarking technique. This message must be extracted from the watermarked piece of content, even if this latter has been subject to some noise or attacks. As documents are often compressed, for storage or transmission purpose, the minimum requirement is for the watermark to resist to a fair lossy compression. Hence, as both watermarking and compression can be modelled as transmissions, it is of interest to try to design joint schemes, which tackle both issues at the same time.

With Çağatay Dikici, Christine Guillemot, Khalid Idrissi and Attila Baskurt, we worked in this direction, to study a scheme that is based on both a watermarking technique and a Wyner-Ziv like encoder. We considered a robust watermarking scheme with side information at the sender side, modelled as a transmission process like in [GP80], and depicted in Fig. 2.4. Costa showed in 1983 in this case that if the noise is additive Gaussian, then the channel capacity is the same as if the channel noise would have been known at the sender but also at the receiver side [Cos83]. Codes that achieve this capacity are called Dirty Paper Codes. Moulin *et al.* focused in [MW07] on the coding problem suggested by Gel'fand-Pinsker in the case of a discrete source alphabet. They computed the channel capacity, and the parameters for an effective construction of the codes achieving the capacity, when the sender, the attacker and/or the receiver have access to some side information.

We focused on Dirty Paper Coding in the case where the receiver has access to a partial (*i.e.* noisy) side information (on the original document). Distortions introduced by the watermarking process, and then the compression process, are constraint to remain lower than some upper bounds. The compression module does not have access to any information on the original document, and uses only the watermarked document. This compression step can be identified with the coding problem of Wyner-Ziv [WZ76], that is, a lossy compression with side information at the decoder side.

We first determined the rate-distortion function of the Wyner-Ziv module. The particularity of our work was to take into account the constraints due to the watermarking module, in terms of signal power and distortion. We first modelled the compression and decompression processes as a transmission channel, to derive the channel capacity of the watermarking module for Gaussian signals. We were then able to analyze the gain in terms of rate and capacity, when the receiver can access this partial side information. We showed that, for a given compression rate, the knowledge of this partial side information during the Wyner-Ziv decoding can lead to a smaller compression distortion. Undirectly, this enables us to increase the global channel capacity.

This theoretical study has been followed by an implementation and experiments. Many codes have been proposed in the literature. We based our implementation on the superposition codes of [BBCS06]. Our watermarking module is based on a TCQ (Trellis Coded Quantization), and an LDPC (Low Density Parity Check) code. The compression module is based on a scalar quantizer, followed by LDPC codes. Our simulations showed that our scheme almost achieves the bounds derived for the

rate-distortion function, as well as for the capacity

This work has been presented at the *IEEE International Symposium on Image/Video Communications over fixed and mobile networks, ISIVC 2008* [DGF+08].

Cryptography : on the design and cryptanalysis of symmetric stream ciphers. During by PhD thesis, I worked on the design of well suited Boolean functions for the design of strong symmetric stream and block ciphers. In order to prevent cryptanalysis, these functions have to satisfy several incompatible criteria. In the steps of my PhD work, I worked with Anne Canteaut, Pascale Charpin, and Claude Carlet on the formalization of the links and compromises between a high nonlinearity and a high propagation criterion, which are particularly important in the design of block ciphers.

Our results have been presented at *EUROCRYPT'00* [CCCF00b] and during the *IEEE International Symposium on Information Theory'2000* [CCCF00a]. They also have been published in *IEEE Transactions on Information Theory* [CCCF01] [14].

When I got a permanent position in Lille in 1999, I created a working group on cryptography, where I supervised my first PhD student, Vincent Bénony. During this period, I worked on the design and attack of symmetric stream ciphers. Such ciphers are usually based on linear feedback shift registers (LFSRs), combined or filtered by well chosen Boolean functions. Finding appropriate Boolean functions is a difficult task, because of the growing set of cryptanalysis attacks on such ciphers, e.g., Berlekamp-Massey reconstruction [Mas69], (fast) correlation attacks [Sie85, MS88, CT00, JJ02], and algebraic attacks [Cou02, CM03b, Cou03, Cou05, HR04, AFI+04]. Some few alternatives to LFSRs have then been studied to bypass these attacks, through the use of non-linear feedback shift registers (NLFSRs), Feedback Carry Shift Registers (FCSRs), or really different primitives as T-functions [KS02, KS04a, KS04b].

There, I published with Eric Filiol a new cipher, named `COS`, where we used the strong Boolean functions I constructed during my PhD thesis. This ciphers, based on the use of NLFSRs, provides very good encryption and decryption speeds. It proposes two operating modes : Mode I for processing structured data, which first breaks its structure before encrypting it ; and Mode II for processing non-structured data, which directly encrypts it. Several preprints discussing its security have been posted on the web, but the only published attack is [WB02], which exhibits a weakness in the primitive that breaks data structure in Mode I, before encryption. The encryption core itself has not been attacked yet. We corrected the weakness and provided a stronger version of the scheme, named `COSvd` in 2004.

The original `COS` has been presented at *IEEE International Symposium on Information Theory'2001* [FFV01] and *IMA Conference on Cryptography and Coding 2001* [FF01]. The improved version, `COSvd`, has been presented at the workshop *SASC : the State of the Art of Stream Ciphers* [FFJ04].

14. This article can be found in Appendix D.

Then, I developed a more personal work with Vincent Bénony, and two colleagues Eric Wegrzynowski and François Recher. We worked on the cryptanalysis on a stream cipher construction based on the use of T-functions. We also worked on the design of NLFSRs generating de Bruijn sequences. We obtained interesting and promising results on such constructions, presented in the manuscript of Vincent Bénony's PhD thesis. More effort should have been made to provide a complete stream cipher, which could have been published. Unfortunately, as Vincent Bénony left the academic world after his thesis, and as my new laboratory did not want me to continue working on such a topic, this work has not been finished properly.

Our cryptanalysis of Klimov and Shamir's stream cipher has been presented during the conference *Sequences and their applications* [BRWF05]. Vincent's de Bruijn sequences generator is described in his PhD thesis manuscript [Bén06].

Appendix B – Defense's slides.

My defense's slides (in English) can be found in Appendix B.

Appendix C – Curriculum Vitæ and publications.

This English version is less detailed than the French one, which can be found on Page 121. My complete list of publications can be found on Page 130.

I have officially supervised or co-supervised four PhD students : Vincent Bénony (2002-2006), Fuchun Xie (2007-2010), Ana Charpentier (2008-2011), and Morgan Barbier (2008-2011). I also supervised two postdocs : Julio Cesar Hernandez-Castro (2004-2005), and Fabien Galand (2006-2007).

I have been asked to be a member of PhD defense committees, from 2002 to 2011. For two of them I have been asked to act as a reviewer.

I am currently a member of the editorial board of Springer's Journal in Computer Virology.

I have been a member of the Program Committees of the international conferences Indocrypt (2004, 2005), WCC (2007, 2009), Wacha (2007-chair) and IEEE/ACM ICDIM'07 (2007), and the national events and conferences *Journées Codes et Stéganographie* (2011), *Journées Codage et Cryptographie* (2005-chair), SSTIC (2004, 2006-2011), C&ESAR (2009), SAR (2004), SAR-SSI (2009), et CORESA (2007, 2009, 2010).

I have been a member of the Organization Committees of the international conferences WCC (1999, 2003, 2007) and Wacha (2007-chair), and the national events and conferences "journées Codage et Cryptographie" (2005-chair), CORESA (2004), SSTIC (2007-2011).

I have been involved in two European founded projects, six National founded projects (two projects as a coordinator), one Regional founded project, and one expertise contract.

I have always taught, as I really like to share and exchange with students. From 2007, I have been co-responsible of the security track of the Master Research in Computer Science of Brittany. In this Research Master, I am in charge of the course dedicated to Content Protection. When I was in Lille, I participated actively to the creation of a course on cryptography. I currently teach this topic now in Telecom Bretagne.

Appendix D – Some selected publications.

I attached in Appendix D some selected publications, for which I did not provide too much details in the dissertation.

Chapitre 1

Assurer la furtivité des communications grâce à la stéganographie

Ὁ γὰρ Ἰστιαῖος βουλόμενος τῷ Ἀρισταγόρῃ σημῆναι ἀποστῆναι ἄλλως μὲν οὐδαμῶς εἶχε ἀσφαλέως σημῆναι ὥστε φυλασσομένων τῶν ὁδῶν, ὁ δὲ τῶν δούλων τὸν πιστότατον ἀποξυρώσας τήν κεφαλὴν ἔστιξε καὶ ἀνέμεινε ἀναφῦναι τὰς τρίχας· ὡς δὲ ἀνέφυσαν τάχιστα, ἀπέπεμπε ἐς Μίλητον ἐντειλάμενος αὐτῷ ἄλλο μὲν οὐδέν, ἐπεὰν δὲ ἀπίκηται ἐς Μίλητον, κελεύειν Ἀρισταγόρην ξυρώσανστά μιν τὰς τρίχας κατιδέοθαι ἐς τὴν κεφαλήν·

Hérodote, Histoires, V, 35

Ce chapitre dresse tout d'abord dans la section 1.1 un panorama des notions et problèmes liés à la stéganographie. Après une présentation des notions de base comme la sécurité ou la capacité des schémas stéganographiques, j'aborderai ensuite plus en détail la problématique du codage par syndrome, sur laquelle j'ai travaillé. Je présenterai enfin dans la section 1.2 les résultats que nous avons obtenus avec Fabien Galand en 2006-2007 lors de son séjour post-doctoral, et dans la section 1.3 ceux que nous avons obtenus avec Daniel Augot et Morgan Barbier en 2010-2011 dans le cadre de la thèse de Morgan.

1.1 Introduction à la stéganographie

La *stéganographie* a pour objet de transmettre un message sans que l'existence même de cette transmission soit détectée. L'Histoire nous a fourni de nombreux exemples de communications secrètes, militaires ou civiles, et de procédés de stéganographie ingénieux, de l'écriture sur les supports des tablettes de cire ou les crânes

des messagers dans l'Antiquité jusqu'aux lettres de Georges Sand, en passant par les encres sympatiques. L'avènement des communications numériques nous pousse aujourd'hui à utiliser comme support des flux de données numériques. Mais l'objectif reste le même : seul le destinataire du message doit pouvoir dire si le flux qu'il reçoit contient ou non un message caché [Sim84]. La *stéganalyse*, symbolisant l'adversaire, s'attache à déceler si un flux donné contient ou non un tel message.

1.1.1 Principes généraux, distorsion, furtivité et sécurité

L'article fondateur de la stéganographie moderne [Sim84] illustre son paradigme par le cas de deux prisonniers, Alice et Bob, retenus dans des cellules différentes, et qui tentent de communiquer malgré la surveillance d'un gardien pour préparer leur évasion. Leur seul moyen de communication est de confier des messages au gardien en lui demandant de les transmettre. Bien sûr, si le gardien suspecte que les messages contiennent des propos dangeureux, il ne les transmettra pas (tels quels). L'objectif d'Alice et Bob est donc de parvenir à s'échanger des informations importantes sur leur évasion en n'échangeant que des données qui semblent annodines au gardien.

Trois stratégies peuvent s'offrir à Alice. La première, nommée en anglais *steganography by selection*, consiste à choisir dans une collection de documents un document particulier et à l'envoyer tel quel à Bob ; c'est alors le choix-même du document qui a une signification ; cette stratégie nécessite *a priori* qu'Alice et Bob se soient mis d'accord au préalable sur la signification de ce choix. La deuxième stratégie, connue sous le nom de *steganography by synthesis*, est de créer le flux ou document qui va cacher le message, en fonction de celui-ci. La dernière, *steganography by modification*, s'appuie sur la modification de documents ou flux supports préexistants. Dans la suite de ce chapitre, nous nous en tiendrons à cette stratégie, qui est la plus commune et la plus étudiée.

Le message à cacher est découpé en blocs, chaque bloc étant inséré séparément dans un support. Comme illustré par la figure 1.1, on distingue quatre étapes dans le traitement d'un bloc de message donné.

Étape 1 : l'émetteur extrait du *document de couverture* $\mathbf{X}$, par des techniques de traitement du signal, un vecteur $\mathbf{x}$ de longueur n représentant les *données de couverture* (en anglais respectivement *cover-medium* et *cover-data*) [1].

Étape 2 : le vecteur $\mathbf{x}$ est modifié par la fonction d'insertion Emb, en fonction du bloc de message $\mathbf{m}$ de longueur m à cacher, pour donner le vecteur de *données-stégo* $\mathbf{y}$, de longueur n (en anglais *stego-data*).

Étape 3 : ces modifications sont ensuite répercutées sur le document (on repasse du vecteur au signal) pour donner le *document-stégo* (en anglais *stego-medium*) $\mathbf{Y}$, qui est envoyé au destinataire.

1. Les différents vecteurs supports utilisés pour les différents blocs de messages peuvent être obtenus à partir de documents de couverture différents ou non. Ils peuvent par exemple être issus de différents blocs de pixels ou coefficients DCT d'une même image. Cet aspect n'est pas discuté plus avant ici.

Étape 4 : le destinataire extrait le vecteur $\mathbf{y}$ des données-stégo du document-stégo $\mathbf{Y}$, et en extrait le bloc de message du document reçu grâce à la fonction d'extraction Ext.

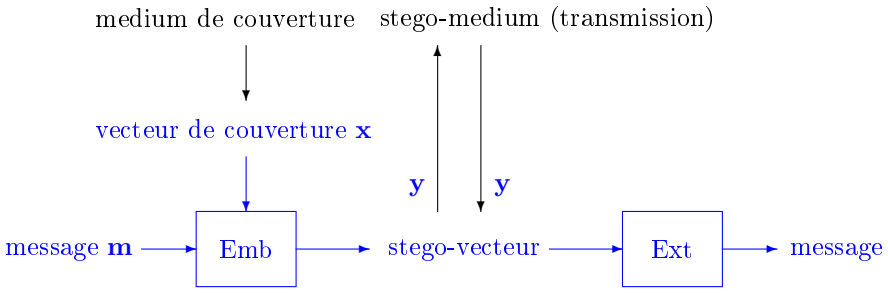


FIGURE 1.1: Stéganographie : principe général et notations.

Quelques considérations générales

Principe de Kerckhoffs et clés. La stéganographie a hérité de la cryptographie quelques principes de sécurité. Le premier est le *principe de Kerckhoffs*, énoncé par Kerckhoffs en 1883 [Ker83a,Ker83b], qui stipule que la sécurité d'un système ne doit pas reposer sur la tenue secrète de la technique utilisée (algorithme). L'histoire nous a en effet montré qu'il est fort probable que les techniques propriétaires — autrement dit conservées secrètes — se trouvent dévoilées un jour, et que le non respect de ce principe peut mener à la catastrophe. Les procédés d'insertion et d'extraction du message sont donc considérés comme publics. Ils sont en revanche paramétrés par des clés, qui vont notamment déterminer quels sous-ensembles des données de couverture vont être modifiés pour porter le message, et dans quel ordre ils vont être traités. La manière dont la clé agit concrètement n'est pas abordée ici, car elle ne constitue pas l'enjeu principal de la sécurité en stéganographie. C'est pourquoi elle n'apparaît pas dans les notations. Comme dans le cas du tatouage, la plupart des schémas sont symétriques, ou à clé secrète, et utilisent la « même » clé pour insérer et extraire le message. Seuls de très rares publications ont étudié la possibilité de construire des schémas à clé publique [GFD02, vAH04, BC05].

Attaquant passif, actif ou malicieux. L'objectif principal de l'attaquant est de déterminer l'existence ou l'absence de communication. Dans un deuxième temps, il peut tenter de retrouver le message caché, ou de tendre des pièges aux protagonistes, comme dans des attaques de type « Homme du milieu ».

La terminologie anglo-saxone se réfère pour l'attaquant au cas du gardien des prisonniers, en termes de *Passive Warden*, *Active Warden* et *Malicious Warden*. Un *gardien passif* se contente de transmettre ou jeter le document qu'Alice veut envoyer à Bob. Un *gardien actif* peut, en sus, altérer le document avant de le transmettre, par exemple en lui appliquant une compression avec perte. Enfin, un *gardien malicieux*

mène des actions spécifiques au schéma stéganographique utilisé, pour piéger les protagonistes en faussant leurs communications, par exemple en se faisant passer pour Alice auprès de Bob, ou en modifiant sciemment le message caché.

La plupart des études de sécurité s'appuie sur le modèle du gardien passif, qui cherche à décider si le document qu'il a entre les mains contient ou non un message caché et s'apparente alors à une communication sur un canal non bruité. C'est ce modèle que nous prenons en considération dans ce chapitre.

Confidentialité du message. Bien entendu, si l'attaquant arrive à déceler la présence d'un message, il essaiera d'en connaître le contenu. Il est donc conseillé à l'émetteur de chiffrer le message avant de le cacher. Par ailleurs, l'émetteur a intérêt à compresser son message pour le rendre le plus court possible avant de le cacher. Ainsi, que le message soit chiffré ou non, on peut sans perte de généralité partir de l'hypothèse qu'il n'a pas de structure particulière.

Choix du document de couverture. En stéganographie, le document de couverture n'est qu'un support, c'est le message qui porte l'information que l'on souhaite protéger. On peut donc la plupart du temps se permettre de choisir parmi un ensemble de documents celui qui se prétera le mieux à une bonne insertion. Cet ensemble de documents dépend du contexte de communication : pour ne pas éveiller de soupçons, Alice ne postera sur un forum de discussions sur les animaux que des photos d'animaux par exemple. Néanmoins, dans certains cas critiques l'émetteur peut ne pas avoir le choix du document de couverture, par exemple parce qu'il est en milieu hostile et n'en a que très peu à sa disposition.

Niveaux de sécurité, et capacité

La stéganographie visant avant tout à transmettre un message de manière furtive, il convient de définir plus précisément cette notion.

Sécurité pragmatique : la stéganalyse. La stéganalyse vise à décider si un document donné est porteur ou non d'un message, et se formalise naturellement comme un problème de détection. Les techniques de stéganalyse s'appuient tout d'abord sur l'extraction de caractéristiques du document (par exemple un histogramme de niveaux de gris pour une image) et sur une modélisation typique de ces caractéristiques. Grâce à des tests d'hypothèses (de type Neyman-Pearson, ou bayésiens par exemple), on décide alors qu'un document donné contient ou non un message, selon que ses caractéristiques semblent vérifier l'hypothèse $\mathcal{H}_0$ ou l'hyphothèse $\mathcal{H}_1$. Pour que la stéganalyse fonctionne, il faut déterminer quelles caractéristiques vont s'avérer pertinentes, et choisir adroitement les hypothèses que l'on va tester. On distingue deux catégories de stéganalyses : on parle de *stéganalyse ciblée* lorsque le détecteur est conçu pour une technique de stéganographie précise (*e.g.* [WP99, FGH02, FGHS03, FGD01a]), et de stéganalyse aveugle lorsque le détecteur est générique, non spécialisé (*e.g.* [LF02, XSG+05, LF06, GFH06, WM07, Ker07b, PF07, KF11, KP11]). La

stéganalyse ciblée s'appuie la plupart du temps sur la mesure d'une seule caractéristique, pertinente pour la technique d'insertion visée. La stéganalyse aveugle nécessite, elle, d'utiliser plusieurs caractéristiques. On utilise ensuite généralement des outils de classification, comme les réseaux de neurones, les algorithmes de *clustering*, ou encore les *support vector machines* (SVM). Ce sont ces derniers qui sont aujourd'hui les plus utilisés. L'utilisation de tels outils demande une phase d'apprentissage, pour calibrer la classification.

La stéganalyse statistique peut également donner lieu à une stéganalyse dite *quantitative* lorsqu'elle vise à estimer le nombre de modifications qui ont été opérées dans le document lors d'une éventuelle insertion de message (*e.g.* [ZP03, PFK09, Pev11, KF10]). Cette dernière approche tente finalement en quelque sorte d'estimer si le document que l'on observe est un document naturel ou s'il a été modifié par une opération d'insertion. La recherche en stéganalyse effective a été stimulée récemment par le concours BOSS[2] qui s'est déroulé en 2010 [BFP11]. La technique mise à l'épreuve, nommée `Hugo`, a été conçue pour l'occasion [PFB10], et les attaques les plus performantes ont été présentées dans [GK11, FKGH11, FGKH11].

Sécurité formelle : comment définir la sécurité parfaite. Assez naturellement se sont développés en parallèle des travaux visant à formaliser la sécurité des systèmes stéganographiques. Zöllner *et al.* [ZFK+98] et Mittelholzer [Mit99] ont proposé une définition de la sécurité calquée sur la notion de *perfect secrecy* de Shannon. En parallèle aux travaux de Zöllner *et al.*, Cachin a proposé une approche légèrement différente, qui fait aujourd'hui référence. Il a défini un système de stéganographie comme *parfaitement sûr* si la distance[3] de Kullbach-Leibler entre les distributions des données de couverture et des données-stégo est nulle [Cac98,Cac04] : $D_{KL}(P_c||P_s) = \sum_{\mathbf{x}} P_c(\mathbf{x}) \log(P_c(\mathbf{x})/P_s(\mathbf{x})) = 0$. Il pose également les limites de cette approche. Tout d'abord, cette caractérisation dépend de la modélisation probabiliste des données de couverture, et n'a de sens que si l'utilisateur et l'attaquant partagent le même modèle. Cette question est très délicate, car contrairement aux cas de la cryptographie ou de la compression où les espaces des cryptogrammes ou des données compressées sont déterminés sans ambiguïté (par exemple toutes les séquences binaires de longueur n, toutes étant équiprobables), ici la situation est bien plus ambiguë car dans la majorité des cas les documents de couverture sont *empiriques*, *i.e.* issus de la numérisation d'objets analogiques « réels » [Böh09]. Aussi, toutes les études menées suivant cette approche sont freinées par la difficulté de caractériser la distribution des données de couverture pour les documents empiriques[4]. Néanmoins, cette notion de distance entre les distributions des données de couverture et des données stégo est maintenant présente dans les esprits des concepteurs de systèmes, et a donné lieu à une amélioration significative des techniques d'insertion. Utilisant les mêmes critères, on peut bien sûr relâcher un peu la condition de nullité de la distance

2. Break Our Steganographic Scheme, `http://www.agents.cz/boss/BOSSFinal/`

3. Cette notion est parfois appelée divergence, car elle n'est pas une distance au sens mathématique.

4. Cette approche donne de meilleurs résultats concrets dans les contextes de *steganography by selection* ou de *steganography by synthesis* pour lesquels les documents de couverture sont maîtrisés.

de Kullbach-Leibler et parler de systèmes ε-sûrs lorsque la distance est bornée par ε. On peut également s'intéresser à la convergence asymptotique de cette distance lorsqu'elle est normalisée par la taille n des données de couverture pour voir si elle tend alors vers 0 lorsque n tend vers l'infini. Moulin *et al.* ont suivi la voie ouverte par Cachin. Ils montrent dans [WM08] les liens forts entre la notion de stéganographie parfaitement sûre au sens de Cachin et celle de tatouage public telles qu'étudiée dans [MO03, SBM03, SBM04, MW07] et proposent une construction de schéma s'appuyant sur des *stacked-binning codes* utilisés pour le codage de canal avec information adjacente [MW07]. Ils étudient par ailleurs la nature des codes qui sont compatibles avec cette notion de sécurité parfaite, et le compromis éventuel à réaliser entre sécurité et capacité. D'autres contributions ont poursuivi dans cette voie. Filler *et al.* proposent une avancée vers la caractérisation des systèmes parfaitement sûrs reposant sur des opérations d'insertion mutuellement indépendantes [FF09a]. Pour la plupart des systèmes relevant de cette catégorie — par exemple LSB ± 1, `F5`, quantification perturbée, `MMx`, modulation stochastique [KFP07] — cette étude caratérise pour une opération d'insertion donnée la nature géométrique de l'espace des sources qui leur assurent une sécurité parfaite. La question est donc en quelque sorte retournée pour caractériser non pas les schémas qui seraient parfaitement sûrs pour un type de documents de couverture donné, mais pour caractériser à schéma fixé quels documents de couverture lui permettront d'atteindre ce niveau de sécurité. Les auteurs introduisent par ailleurs une autre manière de mesurer la sécurité parfaite, équivalente à celle de Cachin, mais reposant sur la théorie de Fisher, et cette nouvelle formulation devrait déboucher sur de nouveaux résultats en stéganalyse quantitative. On commence donc aujourd'hui à entrevoir comment cette notion théorique de sécurité parfaite au sens de Cachin peut être reliée à la conception d'outils de stéganalyse opérationnels. De leur côté, Bas et Cayre ont apporté une réflexion générale à la notion de sécurité formelle en dissimulation d'information, proposant une vision unifiée englobant la stéganographie et le tatouage [CB08].

Katzenbeisser *et al.* et Hopper *et al.* ont adopté une approche très différente, partant de la notion d'indistinguabilité calculatoire au sens de la théorie de la complexité [KP02, HLvA02, HLvA09]. Cette approche permet une gestion plus aisée du cas des documents de couverture empiriques, pour peu que l'on dispose d'un *oracle d'échantillonnage* [HLW06].

Comment combler le fossé entre les deux. Afin de combler le fossé entre les notions formelles très fortes de Cachin et les techniques très pragmatiques de stéganalyse, plusieurs travaux ont proposé de formaliser des niveaux de sécurité intermédiaires.

Moulin *et al.* ont proposé des modèles statistiques pour caractériser les données de couverture et les données-stégo afin de pouvoir calculer la distance de Kullbach-Leibler en fonction de la longueur du message [WM04, MK05]. Ils ont montré qu'à taux d'insertion fixé la probabilité de faux négatifs tend exponentiellement vers zéro, et que la sécurité du schéma peut se mesurer avec l'exposant en question [WM08]. Malheureusement, les modèles utilisés pour caractériser les données de couvertures restent pour l'heure inappropriés en pratique [Böh09, Böh10].

Pour combler le fossé entre la notion théorique de sécurité introduite par Cachin et la résistance pratiques aux techniques expérimentales de stéganalyse, Ker est parti de la stéganalyse, pour en proposer une approche plus formelle, qui s'appuie sur une mesure de distance de Kullbach-Leibler. Cette nouvelle approche vise à donner un cadre d'étude formel pour les schémas qui ne sont pas parfaitement sûrs au sens de Cachin. Comme dans les travaux de Moulin, les travaux de Ker lient fortement le niveau de sécurité et la capacité, même si le terme capacité prend parfois un sens légèrement différent. Ker a ainsi prouvé en 2006 que la quantité d'information que l'on peut cacher de manière « sûre » avec un système non parfaitement sûr au sens de Cachin n'évolue pas du tout linéairement en la taille n des données de couverture comme on le pensait jusqu'alors (et comme c'est le cas pour les schémas parfaitement sûrs), mais doit rester en-deça [5] de $\sqrt{n}$. Ce résultat est connu sous le nom de *square root law*. Ker a principalement développé son approche dans [Ker07c,Ker07a,Ker08b, Ker08a, Ker09a, Ker09b, Ker10b, Ker10a], suivi par Filler *et al.* dans [FKF09]. Les articles récents traitant de conception de schémas stéganographiques se doivent donc, comme [FF09b, FF11], de prendre cette loi en compte et ne pas envisager d'insérer des messages trop long par rapport aux données-stégo. Et les résultats antérieurs sont à modérer, car ceux qui s'attachaient à insérer des message extrêmement longs sont finalement à proscrire.

En parallèle aux travaux menés par Ker *et al.*, Barbier *et al.* [BA08, BM08] se sont appuyés sur les notions d'*indistinguabilité* et de malléabilité, utilisées en cryptographie pour mesurer la sécurité des systèmes de chiffrement, pour formaliser la sécurité des schémas de stéganographie.

Capacité. Comme aperçu dans les lignes qui précèdent, la capacité des schémas stéganographiques est étudiée conjointement avec la sécurité. On distingue deux approches dans la littérature.

La première s'appuie sur la perception du système global comme moyen de transmission, et définit la capacité comme la capacité du canal modélisant la transmission en question. Cet type d'étude ne fixe pas le schéma d'insertion, et définit la capacité comme la plus grande longueur relative de message que l'on peut transmettre, tous schémas confondus, lorsque la taille des données de couverture n tend vers l'infini. Des études ont été menées dans cet esprit, pour estimer la capacité dans le cas des schémas sûrs et des schémas ε-sûrs. Fortement étudiée [Ett98, CM03a, MO03, CPG07a, MW04, WM08]), cette notion de capacité est reliée à la définition de sécurité de Cachin, et souffre donc comme elle pour l'instant de la difficulté à modéliser précisément les documents.

La deuxième approche est différente, puisqu'elle raisonne à schéma fixé, pour se demander quelle longueur de message ce schéma peut cacher sans donner prise à la stéganalyse. Cette capacité-là, souvent qualifiée de *secure payload*, est donc très différente de la précédente. Les travaux de Ker [Ker07c, Ker07a, Ker08b, Ker08a,

5. Anderson avait suspecté qu'elle devait évoluer de manière sous-linéaire dès 1996, mais il n'avait été conjecturé qu'en 2005 qu'elle ne pouvait excéder $\sqrt{n}$.

Ker09a, Ker09b, Ker10b, Ker10a] et Filler *et al.* [FKF09, FF09b, FF11] sur la *square root law* relèvent de cette définition.

Modélisation des documents. On l'aura compris, un des points délicats tant dans les estimations de sécurité que de capacité est la difficulté à modéliser les données support [Böh09, Böh10]. Cette question difficile rejoint le problème de la caractérisation de données *naturelles*, par rapport à des données modifiées ou trafiquées. Si l'on se concentre sur le cas des images, cela revient par exemple à décider si une image donnée peut être considérée comme émanant directement d'un capteur d'appareil photo, auquel cas elle peut être considérée comme originale et non modifiée. Sur ce point, la stéganalyse rejoint donc la question de l'investigation numérique, aussi appelée *digital forensics*. Depuis quelques années, la recherche en *forensics* s'est intéressée à l'identification de la source de l'image, *i.e.* à l'identification de l'appareil photo précis qui a créé l'image. En effet, chaque capteur introduit dans l'image qu'il génère un bruit qui lui est propre, et une analyse scrupuleuse de l'image peut permettre de remonter au capteur en question, donc à l'appareil, et de là au photographe (l'appareil photo est donc identifiable comme on le fait depuis longtemps pour les armes à feu). En stéganalyse on ne cherche pas à identifier un appareil précis, mais plutôt à dire si l'image est directement issue d'un appareil, ou si elle a donné lieu à l'insertion d'un message. Récemment, Cogranne *et al.* ont proposé dans ce sens une modélisation des images issues de capteurs d'appareils photos dans le but d'améliorer les performances de la stéganalyse [CZF+11, ZCR+11].

Stratégies de conception

La conception d'un schéma de stéganographie est délicate, et parmi la multitude de schémas proposés depuis une vingtaine d'années peu peuvent prétendre résister à la stéganalyse aujourd'hui.

Le concepteur doit avant tout effectuer deux choix cruciaux, à savoir quelle fonction utiliser pour passer de l'espace des documents à l'espace des vecteurs ($\mathbf{x}$ et $\mathbf{y}$), et quelle stratégie appliquer pour modifier le vecteur de couverture $\mathbf{x}$.

Comment constituer le vecteur $\mathbf{x}$? Un des moyens les plus populaires pour constituer le vecteur $\mathbf{x}$ à partir du document de couverture est de considérer les bits de poids faible de la luminance des pixels, ou des indices de palettes de couleur (format GIF), ou encore de coefficients transformés (format JPEG avec les coefficients DCT). Mais attention, il ne faut pas directement cacher le message tel quel dans les bits de « niveau 1 » (poids le plus faible), sous peine de donner lieu à des biais statistiques débouchant sur des techniques de stéganalyse comme l'attaque dite de l'histogramme [WP99] ou l'attaque RS [FGD01a]. Mieux vaut considérer des plans de bits d'ordre supérieur, et choisir soigneusement comment le message va les modifier.

Comment modifier le vecteur $\mathbf{x}$? Une fois le vecteur $\mathbf{x}$ ainsi constitué, on peut s'appuyer sur quatre « philosophies » pour choisir quel type de modification lui apporter.

(a) Assurer la sécurité. La première se raccroche à la notion de sécurité définie par Cachin : une fois défini le modèle des documents de couverture (*via* des caractérisations d'histogrammes, ou de statistiques d'ordre supérieur), on impose pour assurer la sécurité que le stégo-document reste bien dans ce même modèle (*i.e.* préserve ces statistiques) [TBHK03, Sal05, SSM+06, KF08]. Compte tenu de la difficulté de modéliser les images dans leur ensemble, cette approche s'avère limitée en pratique, car elle ne s'applique en général que sur des modèles simplifiés, ou sur de petites parties de l'image. De plus, dans certains cas cette limitation peut même avoir comme effet de bord de faciliter certaines stéganalyses [BW04,SCC06]. Devant la grande difficulté d'appliquer la notion de sécurité de Cachin en pratique lors de la conception, on se rabat en général sur des « philosophies » plus heuristiques. Notons néanmoins que les avancées récentes en termes de modélisation de documents [CZF+11], et de conception de techniques assurant la préservation des distributions des documents [BCG12] permettront peut-être dans un avenir assez proche d'obtenir des niveaux de sécurité réels prouvés.

(b) Rester naturel. La deuxième « philosophie » vise à insérer le message en effectuant des modifications qui peuvent sembler naturelles, comme par exemple introduire un bruit similaire à celui du capteur d'un appareil photo [FG03, FS05]. Malheureusement, si cette nature est suffisamment structurée pour être modélisée et maîtrisée lors de l'insertion, elle peut donner lieu à une stéganalyse aveugle par classification (SVM).

(c) Contrer la stéganalyse. La troisième « philosophie » s'attache à contrer des techniques de stéganalyse. `Outguess` [Pro01] et `F5` [Wes01] ont par exemple été conçus pour contrer les attaques par histogramme. Cette approche est limitée par le fait que le schéma peut quand même être sensible à d'autres techniques de stéganalyse à venir. Par exemple, `Outguess` et `F5` ne résistent pas aujourd'hui à la stéganalyse. Par ailleurs, il est parfois nécessaire, comme c'est le cas pour `Outguess` pour préserver la structure des données (*e.g.* histogramme) qui peuvent être utilisées lors de l'attaque considérée, de modifier beaucoup plus de symboles que ce qui est strictement nécessaire pour insérer le message lui-même.

(d) Minimiser la distorsion. La quatrième « philosophie » vise à minimiser les conséquences de l'insertion par la minimisation de fonctions de coûts reflétant la *distorsion* introduite. Cette minimisation ne garantit pas à elle seule la non-détectabilité de l'insertion [FGD01b, Fra02, BW04, Ker05], mais donne lieu en pratique à de bons schémas. C'est dans cette catégorie que l'on retrouve les techniques d'insertion de type codage par syndrome, comme `F5` [Wes01] et sa variante `nsF5` [KFP07], ou encore la famille de schémas dite `MMx` [KDR06], ainsi que certaines techniques de type *wet paper* comme par exemple la quantification perturbée [FG04,FGS05b,FF07,KFP07].

Mais la formalisation de la notion de *distorsion* n'est pas simple. Il faudrait en effet idéalement trouver une mesure qui soit à la fois réaliste du point de vue de ce que l'utilisateur va percevoir et de ce que les techniques de stéganalyse vont détecter, et qui soit à la fois facile à mesurer directement au niveau du vecteur $\mathbf{x}$ pour être minimisée lors de l'insertion. On ne connaît pas de telle mesure aujourd'hui. On est donc contraints de se rabattre sur des mesures imparfaites. La plupart des publications minimisent le nombre moyen de modifications R_a apportées au vecteur de

couverture $\mathbf{x}$, partant du principe que la distorsion évolue comme le carré du nombre de modifications [Ker08b]. Elles s'appuient alors sur l'*efficacité d'insertion moyenne* (*average embedding efficiency*) $e = m/R_a$ pour estimer la qualité de la discrétion de l'insertion. Cette mesure de distorsion n'est pas la meilleure, car elle reste uniquement locale et ne prend pas en compte la dépendance entre les pixels de l'image, mais offre l'avantage d'être facile à estimer et contrôler, et donc minimiser. De plus, il est clair que plus on modifie de composantes du vecteur $\mathbf{x}$, plus cela risque de donner lieu à une détection [Wes01, KDR06, KFP07]. Minimiser le nombre de modifications est donc nécessaire, même si cela n'est pas suffisant. Récemment, Filler et Fridrich se sont attachés à étudier plus précisément les liens qui relient le modèle de distorsion au modèle des données de couverture. Ils ont tout d'abord cherché à mesurer l'impact du modèle de distorsion additif (la distorsion globale étant calculée comme la somme des distorsions locales) sur le modèle des données de couverture, prenant en compte la dépendance qui existe entre chaque pixel de l'image et ses voisins [FF11]. En parallèle, ils ont exploré l'autre direction, étudiant l'impact du modèle des données de couverture sur le modèle de distorsion. Reprenant les critères de Cachin, cet impact est estimé au travers de la distance de Kullbach-Leibler entre la distribution des données de couverture et la distribution des données stégo [FF10a]. Toute avancée dans la modélisation des données devrait ainsi pouvoir donner lieu à la définition d'une meilleure mesure de distorsion à minimiser lors de l'insertion.

Sélection (non partagée) des composantes à modifier en fonction du document de couverture et du message. *Writing on wet paper.* L'approche dite des *wet paper codes* a été fortement étudiée dans le cadre du codage par syndrome décrit ci-après, mais n'y est pas spécifique.

Fridrich *et al.* ont introduit et discuté dans [FGLS05] une approche de l'insertion qui vise à diminuer le nombre moyen de modifications lors de l'insertion d'un message donné (suivant les préceptes de la quatrième « philosophie » de conception présentée au paragraphe précédent). Elle prend en compte le fait que lors de la constitution du vecteur $\mathbf{x}$ à l'étape 1. On peut d'ores et déjà décider qu'étant donnés le document-support et le bloc de message que l'on souhaite y cacher, certaines composantes du vecteur $\mathbf{x}$ ne devront absolument pas être modifiées, sous peine de perdre la furtivité. Ces composantes sont alors considérées comme verrouillées lors de l'insertion. Mais le destinataire, lui, ne les connaît pas, *a priori*, lorsqu'il procède à l'extraction du message. Dans la littérature, on parle de *non-shared selection channel*, les *dry components* désignant les composantes que l'on peut modifier (on peut écrire sur du papier sec), et les *wet components* désignant les composantes verrouillées (on ne peut pas écrire sur du papier mouillé). Lorsque le destinataire reçoit le stégo-médium, c'est comme si le papier avait séché et il ne sait pas faire la différence entre les composantes qui étaient sèches (et portent de l'information) et celles qui étaient mouillées (et ne portent pas d'information). Notons que la gestion du verrouillage de certaines composantes est très pertinente en stéganographie, mais impose des contraintes supplémentaires qui rendent le choix de la technique d'insertion plus délicate, et font par ailleurs augmenter la probabilité d'échec de l'insertion.

Comme souligné dans [FGLS05], cette notion est à rapprocher des travaux menés sur les mémoires à cellules défectueuses (*stuck*). Dans ce contexte, on peut montrer, en s'appuyant sur le théorème de Gel'fand-Pinsker [GP80] que si $\ell_{\mathcal{W}}$ composantes parmi n sont verrouillées/mouillées, alors asymptotiquement lorsque n tend vers l'infini mais que le ratio des composantes modifiables/sèches $(n - \ell_{\mathcal{W}})/n$ est fixé, on peut transmettre $n - \ell_{\mathcal{W}}$ symboles de message. Autrement dit on s'en sort aussi bien que si le destinataire connaissait l'emplacement des $n - \ell_{\mathcal{W}}$ composantes modifiables [FGLS05]. La construction menant à ce résultat s'appuie sur la technique dite de *random binning* : pour $\varepsilon > 0$ fixé, on associe les mots q-aires de n symboles à $q^{(1-\varepsilon)n-\ell_{\mathcal{W}}}$ dictionnaires disjoints numérotés ; ces dictionnaires sont partagés entre l'émetteur et le récepteur ; lorsque l'émetteur souhaite transmettre/stocker un message de $(1 - \varepsilon)n - \ell_{\mathcal{W}}$ symboles en utilisant comme support le mot $\mathbf{x}$, il sélectionne le dictionnaire dont l'indice est précisément égal au message ; dans ce dictionnaire, il choisit un mot $\mathbf{y}$ tel que $x_i = y_i$ pour toutes les composantes verrouillées ; c'est ce mot $\mathbf{y}$ qui est envoyé au destinataire. Dans chaque dictionnaire, il y a $q^n/q^{(1-\varepsilon)n-\ell_{\mathcal{W}}} = q^{\varepsilon n+\ell_{\mathcal{W}}}$ mots, et la probabilité qu'aucun des $q^{n-\ell_{\mathcal{W}}}$ mots compatibles avec $\mathbf{x}$ ne soit dans le dictionnaire sélectionné (et donc que la transmission ne puisse avoir lieu) est $((1 - 1/q^{\ell_{\mathcal{W}}})^{q^{\ell_{\mathcal{W}}}})^{q^{\varepsilon n}}$, qui tend vers 0 lorsque n tend vers l'infini et que $(n - \ell_{\mathcal{W}})/n$ est constant. Mais cette construction, bien qu'asympotiquement optimale, n'est absolument pas intéressante en pratique puisqu'elle demande aux protagonistes d'échanger au préalable de très gros dictionnaires. Par ailleurs elle ne garantit pas de borne sur le nombre de modifications effectuées (les bornes ne sont que probabilistes). Le codage par syndrome offre des solutions plus opérationnelles que l'utilisation de *random binning*, tout en limitant en sus la distorsion introduite, ce qui est essentiel en stéganographie.

Parmi les techniques d'insertion s'appuyant sur un canal de sélection non partagé, on peut citer l'insertion par quantification perturbée (les composantes modifiables/sèches pouvant être quantifiées non pas par la valeur la plus proche, mais par la seconde plus proche) [FG04, FGS05b, FF07, KFP07], la famille de schémas dite `MMx` qui combine le principe de quantification perturbée et de codage par syndrome [KDR06], ou encore `nsF5` [KFP07] variante améliorée de `F5` qui, comme elle, utilise du codage par syndrome. Toutes trois offrent aujourd'hui une bonne résistance à la stéganalyse universelle par classification.

1.1.2 Le codage par syndrome

Comme évoqué dans les paragraphes précédents, la modélisation des étapes d'insertion et d'extraction du message à l'aide de codes correcteurs d'erreurs est particulièrement pertinente lorsque l'on souhaite minimiser la distorsion [Cra98, Bie01, GK03, GK09, BF08]. Ce modèle, appelé *codage par syndrome* (en anglais *syndrome coding* ou encore *matrix embedding*), a initialement été développé pour la conception de schémas de stéganographie procédant par insertion du message dans des images. Il a notamment été rendu populaire par son utilisation dans la conception de l'algorithme `F5` [Wes01]. Néanmoins, cette approche n'est pas liée à la nature-même du document ou flux, et pourrait très bien *a priori* être appliquée à d'autres types

de support. Outre sa pertinence, c'est son lien avec les codes correcteurs qui m'a particulièrement intéressée.

Considérons que l'on travaille dans un alphabet q-aire, muni de la structure de corps[6] fini $\mathbb{F}_q$ et de la distance de Hamming d_H, avec un bloc de message $\mathbf{m}$ de longueur m et un vecteur $\mathbf{x}$ de longueur n, et que l'on autorise le changement d'au plus T symboles de $\mathbf{x}$. Le schéma, constitué des opérations d'insertion Emb et d'extraction Ext, doit alors naturellement satisfaire les conditions suivantes :

$$\forall(\mathbf{x},\mathbf{m}) \in \mathbb{F}_q^n \times \mathbb{F}_q^m, \quad \mathrm{Ext}(\mathrm{Emb}(\mathbf{x},\mathbf{m})) = \mathbf{m} \tag{1.1}$$

$$\forall(\mathbf{x},\mathbf{m}) \in \mathbb{F}_q^n \times \mathbb{F}_q^m, \quad d_H(\mathbf{x},\mathrm{Emb}(\mathbf{x},\mathbf{m})) \leq T \tag{1.2}$$

Le lien entre la conception d'un schéma de stéganographie satisfaisant ces conditions et les codes correcteurs est apparu dans les forums de discussions et lors d'échanges privés [Cra98,Bie01], avant d'être formellement établi et publié dans [GK03, GK09]. Galand et Kabatiansky y ont montré que les deux problèmes suivants sont formellent équivalents :

- construire un schéma satsisfaisant les équations (1.1) et (1.2) ;
- trouver q^m *codes de recouvrement*[7] disjoints de l'espace $\mathbb{F}_q^n$, recouvrant l'espace avec des boules de rayon T.

L'idée du codage par syndrome est d'utiliser comme partitionnement de l'espace ambiant les translatés d'un *code correcteur linéaire* $\mathcal{C}$, sous-espace vectoriel de dimension k de $\mathbb{F}_q^n$. Le cas des codes non-linéaires n'a été que très peu étudié pour l'instant [BF08, ZL08, MB, Bar11] et n'est pas présenté ici.

Comme tout code linéaire, $\mathcal{C}$ est caractérisé par une *matrice de parité* $(n-k)\times n$ (non unique) $\mathbf{H}$: $\mathbf{c} \in \mathcal{C} \Leftrightarrow \mathbf{c}\cdot\mathbf{H}^t = \mathbf{0}$. Les q^{n-k} *translatés* de $\mathcal{C}$ constituent une partition de l'espace ambiant, et sont en bijection avec les q^{n-k} *syndromes* $\mathbf{y}\cdot\mathbf{H}^t$ des mots de l'espace : il y a autant de translatés que de syndromes différents, et tous les vecteurs d'un même translaté ont le même syndrome. Les *mots de code* ont un syndrome nul. Extraire le bloc de message $\mathbf{m}$ de longueur $m = n-k$ du vecteur des données-stégo $\mathbf{y}$ revient à calculer le syndrome de $\mathbf{y}$, opération simple puisqu'il s'agit simplement d'un produit matrice-vecteur. Insérer $\mathbf{m}$ dans le vecteur des données de couverture $\mathbf{x}$ revient à ajouter à $\mathbf{x}$ un vecteur $\mathbf{e}$ qui nous permet de nous placer dans le bon translaté, et donc d'avoir le bon syndrome. La figure 1.2 illustre ce procédé. Ce vecteur $\mathbf{e}$ doit avoir un poids de Hamming inférieur ou égal à T. Il est important de noter dès maintenant que trouver ce vecteur $\mathbf{e}$ est un problème calculatoire difficile en général.

Dans le cas d'un canal de sélection non partagé (*wet paper*), on souhaite de surcroît forcer certaines composantes de $\mathbf{x}$ à rester intactes. Soit $\mathcal{W}$ l'ensemble des indices de ces composantes que l'on souhaite verrouiller (*wet components*). On a alors

6. La plupart des schémas travaillant sur un corps fini, je me focaliserai sur ce cas ici. Notons néanmoins que quelques rares publications traitent du codage par syndrome utilisant des codes définis sur $\mathbb{Z}_4$ ou $\mathbb{Z}_2\mathbb{Z}_4$.

7. En théorie des codes, on appelle *code de recouvrement* de l'espace $\mathbb{F}_q^n$ de rayon ϱ un *code* $\mathcal{C}$ (autrement dit un sous-ensemble de $\mathbb{F}_q^n$) tel que tout mot $\mathbf{v} \in \mathbb{F}_q^n$ de l'espace ambiant est à une distance au plus ϱ de $\mathcal{C}$.

une contrainte supplémentaires :

$$\forall i \in \mathcal{W},\ \mathbf{y}_i = \mathbf{x}_i \tag{1.3}$$

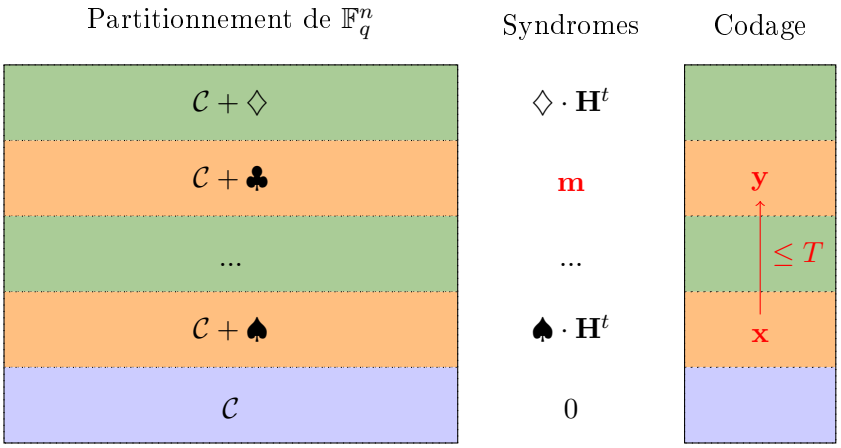


FIGURE 1.2: Principe du codage par syndrome avec un code linéaire $\mathcal{C}$ de longueur n et dimension k. Les vecteurs $\mathbf{x}$ et $\mathbf{y}$ sont de longueur n, le message $\mathbf{m}$ est de longueur $m = n-k$. On autorise au plus T modifications du vecteur $\mathbf{x}$ pour réaliser l'insertion.

Formalisation des problèmes à résoudre et analyse

Regardons maintenant plus formellement ce qui se passe, afin de mieux comprendre les enjeux liés au choix du code $\mathcal{C}$.

Concernant le cas simple où on ne se préoccupe pas du verrouillage de certaines composantes de $\mathbf{x}$, nos seules contraintes sont (1.1) et (1.2). Le problème à résoudre est donc le suivant.

Problème 1.1 (codage par syndrome) *Soient $\mathcal{C}$ un code linéaire de paramètres $[n,k]$, $\mathbf{H}$ une matrice de parité de $\mathcal{C}$, $\mathbf{x} \in \mathbb{F}_q^n$ le vecteur de couverture, $\mathbf{m} \in \mathbb{F}_q^{n-k}$ le message à insérer dans $\mathbf{x}$, et $T \in \mathbb{N}^*$ le nombre maximal de modifications autorisées dans $\mathbf{x}$. Le* problème du codage par syndrome *consiste à trouver $\mathbf{y} \in \mathbb{F}_q^n$ tel que $\mathbf{y} \cdot \mathbf{H}^t = \mathbf{m}$ et $d_H(\mathbf{x},\mathbf{y}) \leq T$.*

Le problème 1.1 nous amène assez naturellement à la construction suivante :

$$\mathbf{y} = \mathrm{Emb}(\mathbf{x},\mathbf{m}) = \mathbf{x} + \mathrm{D}(\mathbf{m} - \mathbf{x}\cdot\mathbf{H}^t) \tag{1.4}$$

$$\mathrm{Ext}(\mathbf{y}) = \mathbf{y}\cdot\mathbf{H}^t \tag{1.5}$$

avec D() une fonction capable de retourner un mot $\mathbf{e}$ de syndrome donné dont le poids de Hamming est inférieur ou égal à T. Par linéarité, on obtiendra alors un

vecteur $\mathbf{y}$ de bon syndrome, à distance au plus T de $\mathbf{x}$. Toute la question est de choisir $\mathcal{C}$ de sorte que la recherche de ce vecteur $\mathbf{e}$, satisfaisant le système

$$\mathbf{e} \cdot \mathbf{H}^t = \mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t \quad (1.6)$$

$$w_H(\mathbf{e}) \leq T \quad (1.7)$$

soit opérationnelle (calculatoirement réalisable), pour permettre l'insertion.

Focalisons-nous tout d'abord sur la résolution de l'équation (1.6). On a vu que les translatés du code couvrent l'espace, et que tous les syndromes y sont représentés. On est donc assurés de l'existence d'une solution, tout élément du translaté correspondant au syndrome recherché étant une solution possible. Bien entendu, l'ajout de la contrainte (1.7) au système complique sa résolution, et peut même la rendre impossible. Cette résolution est fortement liée aux *fonctions de décodage* de $\mathcal{C}$, comme nous allons le voir. Rappelons qu'*une fonction de décodage de $\mathcal{C}$ de rayon T'* est définie par : $f_{\mathcal{C},T'} : \mathbb{F}_q^n \longrightarrow \mathcal{P}(\mathcal{C}) \cup \{?\}$, avec pour tout $\mathbf{y} \in \mathbb{F}_q^n$, soit $f_{\mathcal{C},T'}(\mathbf{y}) =?$ (on n'a pas réussi à décoder $\mathbf{y}$), soit $f_{\mathcal{C},T'}(\mathbf{y}) \subseteq \mathcal{C}$ avec $d_H(\mathbf{y},\mathbf{c}) \leq T'$ (ou de manière équivalente $\mathbf{y} \in \mathcal{B}(\mathbf{c},T')$, avec $\mathcal{B}(\mathbf{c},T')$ la boule centrée en $\mathbf{c}$ et de rayon T') pour tout $c \in f_{\mathcal{C},T'}(\mathbf{y})$. Pour appréhender ce lien, et ses conséquences sur le choix de $\mathcal{C}$, il convient de revenir sur un certain nombre de notions de théorie des codes.

(Quelques rappels sur les codes linéaires)

Considérons comme espace ambiant l'espace vectoriel $\mathbb{F}_q^n$, muni de la distance de Hamming $d_H(\mathbf{x},\mathbf{y}) = |\{1 \leq i \leq n, \mathbf{x}_i \neq \mathbf{y}_i\}|$. La distance à un ensemble est définie comme la plus petite distance possible aux éléments de cet ensemble. On définit également le poids de Hamming d'un vecteur comme le nombre de composantes non nulles, $w_H(\mathbf{y}) = d_H(\mathbf{y},\mathbf{0})$.

Un *code correcteur linéaire* $\mathcal{C}$ de paramètres $[n,k]$ est un sous-espace vectoriel de dimension k de $\mathbb{F}_q^n$. Il est caractérisé par une *matrice de parité* (non unique) $\mathbf{H}$ de taille $(n-k) \times n$, telle que $\mathbf{c} \in \mathcal{C} \Leftrightarrow \mathbf{c} \cdot \mathbf{H}^t = \mathbf{0}$.

Les codes correcteurs d'erreurs ont principalement comme objectif de détecter et corriger les erreurs qui peuvent apparaître lors du stockage ou de la transmission des données. L'idée sous-jacente est de rajouter $n-k$ symboles de redondance au vecteur d'information de longueur k que l'on souhaite protéger, pour obtenir un mot de code $\mathbf{c}$ de longueur n. C'est ce mot de code $\mathbf{c} \in \mathbb{F}_q^n$ qui est finalement stocké ou transmis. Le code $\mathcal{C} \subseteq \mathbb{F}_q^n$ est constitué des mots de code $\mathbf{c}$ correspondant à tous les q^k vecteurs d'information possibles. À la réception, on regarde si le mot récupéré $\mathbf{v} \in \mathbb{F}_q^n$ est un mot de code. Si oui alors on considère [8] qu'il n'y a pas eu d'erreurs. Si non, alors on considère [8] souvent [9] (a) dans le cas d'un *décodage unique* que c'est le mot de code le plus proche de $\mathbf{v}$ qui était stocké ou transmis, ou (b) dans le cas d'un *décodage en liste* que c'est l'un des mots de code les plus proches de $\mathbf{v}$ qui était stocké ou transmis.

8. Même si c'est peut-être à tort.
9. On ne traitera pas ici des autres stratégies de décodage.

Décodage. Une *fonction de décodage de $\mathcal{C}$ de rayon T'* est définie par : $f_{\mathcal{C},T'} : \mathbb{F}_q^n \longrightarrow \mathcal{P}(\mathcal{C}) \cup \{?\}$, avec pour tout $\mathbf{y} \in \mathbb{F}_q^n$, soit $f_{\mathcal{C},T'}(\mathbf{y}) =?$ (on n'a pas réussi à décoder $\mathbf{y}$), soit $f_{\mathcal{C},T'}(\mathbf{y}) \subseteq \mathcal{C}$ avec $d_H(\mathbf{y},\mathbf{c}) \leq T'$ (ou de manière équivalente $\mathbf{y} \in \mathcal{B}(\mathbf{c},T')$) pour tout $c \in f_{\mathcal{C},T'}(\mathbf{y})$.

Afin de mesurer l'aptitude, au moins théorique, du code à corriger les erreurs, on regarde comment peut se comporter une telle fonction selon les valeurs du rayon T'. Supposons dans un premier temps que dès qu'il existe une boule $\mathcal{B}(\mathbf{c},T')$ qui contient le mot à décoder alors le décodage est réussi, autrement dit on fait abstraction de la difficulté algorithmique du décodage pour se focaliser sur des considérations uniquement topologiques. L'ensemble des mots décodables seraient alors précisément égal à l'union des boules de rayon T' centrées en les mots du code $\mathcal{U}_{\mathcal{C},T'} = \cup_{\mathbf{c}\in\mathcal{C}}\mathcal{B}(\mathbf{c},T')$.

(a) Décodage unique. Ce type de décodage vise à retourner l'unique mot de code le plus proche du mot à décoder. Pour s'assurer qu'il n'y aura pas d'ambiguïté lors du décodage, il faut que les boules de rayon T' centrées en les mots du code ne s'intersectent pas. Autrement dit, il faut prendre $T' \leq (d-1)/2$, avec d la *distance minimale* de $\mathcal{C}$, c'est-à-dire la plus petite distance de Hamming entre deux mots de code de $\mathcal{C}$ distincts. La plus grande valeur de T' qui assure cette unicité est appelée la *capacité de correction* de $\mathcal{C}$: elle est donc égale à $\lfloor (d-1)/2 \rfloor$. Ainsi, tant que $T' \leq \lfloor (d-1)/2 \rfloor$, tout mot de $\mathcal{U}_{\mathcal{C},T'}$ est théoriquement décodable de manière unique. Les mots qui ne sont pas couverts par cette union sont non décodables.

Assez naturellement, on cherche donc à caractériser en parallèle l'aptitude d'un code à pouvoir décoder tout mot de l'espace. On définit pour cela le *rayon de recouvrement* ϱ du code, qui mesure son aptitude à couvrir l'espace ambiant par des boules de rayon ϱ : $\varrho = \min\{r \in \mathbb{N}^* \ : \mathcal{U}_{\mathcal{C},r} = \mathbb{F}_q^n\}$. Il est également égal au plus grand poids possible pour un translaté du code (le poids d'un translaté étant défini comme le poids d'un *coset leader*, soit le poids minimum pour un mot du translaté). Il est clair que si on veut que tous les mots de l'espace ambiant puissent être décodables de manière unique, il faut et il suffit de considérer une fonction de décodage de rayon $T' = \varrho$. On parle alors de *décodage complet*.

Un code $\mathcal{C}$ est dit *parfait* si son rayon de recouvrement est égal à sa capacité de correction. Autrement dit, un code parfait est capable de corriger de manière unique toutes les erreurs, quelque soit leur poids : tout mot de l'espace ambiant $\mathbb{F}_q^n$ se situe à une distance inférieure ou égale à ϱ, ici précisément égal à $\lfloor (d-1)/2 \rfloor$, du code parfait $\mathcal{C}$. Tous les codes parfaits linéaires sont connus [Tie73, ZL73], et il y en a peu. Il s'agit des codes de Hamming, du code de Golay binaire de paramètres $[n = 23, k = 12, d = 7]$ et du code de Golay ternaire de paramètres $[n = 11, k = 6, d = 5]$.

(b) Décodage en liste. Dans un décodage en liste, on autorise la non-unicité du décodage. On demande alors à la fonction de décodage de retourner une liste des mots de code considérés comme candidats au décodage du mot $\mathbf{v}$. Ceci permet dans tous les cas de retourner l'ensemble des mots à plus faible distance. Par ailleurs, pousser le paramètre T' loin au-delà de $\lfloor (d-1)/2 \rfloor$ permet de décoder plus de mots, et donne par ailleurs une liste de mots de code candidats plus vaste.

Bien entendu, au-delà de ces considérations topologiques, il reste à déterminer pour un code donné si on dispose d'un algorithme polynomial efficace permettant de mener à bien le décodage. Il existe des techniques de décodage génériques, comme le décodage par syndrome dans le cas du décodage unique, et des techniques spécifiques à certaines familles de codes. Nous ne mentionnerons pas ici toutes ces techniques, et nous limiterons aux remarques pertinentes pour l'utilisation de certains codes en stéganographie.

Quelques résultats sur les translatés du code et la fonction syndrome. On appelle *translaté* de $\mathcal{C}$ engendré par $\mathbf{y}$ l'ensemble des q^k mots $\mathbf{y}+\mathcal{C} = \{\mathbf{y}+\mathbf{c}, \mathbf{c} \in \mathcal{C}\}$. Les q^{n-k} *translatés* de $\mathcal{C}$ constituent une partition de l'espace ambiant, et sont en bijection avec les q^{n-k} *syndromes* $\mathbf{y} \cdot \mathbf{H}^t$ des mots de l'espace : il y a autant de translatés que de syndromes différents, et tous les vecteurs d'un même translaté ont le même syndrome. Par définition de $\mathbf{H}$, les mots de code ont un syndrome nul.

On appelle *coset leader* d'un translaté $\mathbf{y}+\mathcal{C}$ un mot du translaté de poids de Hamming minimum, c'est-à-dire dont le poids est égal à $\min_{\mathbf{v}\in\mathbf{y}+\mathcal{C}} w_H(\mathbf{v})$. Ce poids minimum correspond précisément à la distance qui sépare le translaté du code $\mathcal{C}$.

Soit $\mathcal{C}$ un code linéaire de longueur n, dimension k, distance minimale d, et rayon de recouvrement ϱ, noté en résumé $[n,k,d]\varrho$. Soient $\mathbf{H}$ une matrice de parité de $\mathcal{C}$ et $\mathbf{y} \to \mathbf{y} \cdot \mathbf{H}^t$ la fonction syndrome associée. Alors, pour tout mot $\mathbf{v} \in \mathbb{F}_q^n$ de l'espace ambiant :

- la fonction syndrome restreinte à la boule $\mathcal{B}(\mathbf{v}, \lfloor (d-1)/2 \rfloor)$ est injective *(tous les mots de $\mathcal{B}(\mathbf{v}, \lfloor (d-1)/2 \rfloor)$ ont un syndrome différent)* ;
- la fonction syndrome restreinte à la boule $\mathcal{B}(\mathbf{v}, \varrho)$ est surjective *(les mots de $\mathcal{B}(\mathbf{v}, \varrho)$ permettent d'obtenir tous les syndromes possibles)* ;
- si $\mathcal{C}$ est parfait, alors pour tout mot $\mathbf{v} \in \mathbb{F}_q^n$ de l'espace ambiant, la fonction syndrome restreinte à la boule $\mathcal{B}(\mathbf{v}, \varrho = \lfloor (d-1)/2 \rfloor)$ est bijective.

(Atteindre une efficacité d'insertion optimale)

Il est également nécessaire, avant de reprendre notre analyse du codage par syndrome, de dire quelques mots sur l'impact du code utilisé sur l'efficacité d'insertion. Le codage par syndrome appartient à la catégorie des techniques d'insertion visant à minimiser la distorsion. Comme évoqué précédemment, cette minimisation est souvent abordée sous l'angle de la minimisation du nombre de modifications apportées au vecteur $\mathbf{x}$. Elle est alors évaluée au travers de l'efficacité d'insertion moyenne (*average embedding efficiency*) e du schéma, définie comme le quotient du nombre de symboles de message cachés par le nombre moyen R_a de composantes de $\mathbf{x}$ que l'on a modifiées : $e = m/R_a$.

Dans le cas général, q-aire, le nombre moyen de modifications R_a apportées lors de l'insertion est égal à la distance moyenne entre le code $\mathcal{C}$ et ses translatés, ce qui revient à $R_a = \frac{1}{q^n} \sum_{\mathbf{v}\in\mathbb{F}_q^n} d_H(\mathbf{v}, \mathcal{C})$, avec $d_H(\mathbf{v}, \mathcal{C}) = \min_{\mathbf{c}\in\mathcal{C}} d_H(\mathbf{v}, \mathbf{c})$. L'efficacité d'insertion moyenne est donnée par $e = m/R_a$ symboles par modification, que l'on peut considérer comme égale à $e = (m/R_a) \log_2 q$ bits par modification si on souhaite

se raccrocher à une représentation binaire [10]. Si on considère des codes de longueur n capables d'insérer des messages de longueur m, on parlera de la *longueur relative* de message $\alpha = m/n$. Pour n arbitraire et α fixé, on a alors la borne [FLS06]

$$e \leq \frac{\alpha}{H_q^\star(\alpha)} \tag{1.8}$$

avec $H_q^\star(x)$ la fonction inverse de l'entropie q-aire $H_q(x) = -x\log_2(x) - (1-x)\log_2(1-x) + x\log_2(q-1)$ sur $[0, 1-1/q]$. Un schéma présentant une efficacité d'insertion optimale peut donc insérer un message de longueur relative α en effectuant en moyenne $H_q^\star(\alpha)$ modifications. Le taux moyen de symboles modifiés est donné par $\beta = R_a/n$, et on a $e \leq H_q(\beta)/\beta$. Cette borne (1.8) a été prouvée dans le cas binaire par [Bie98], et dans le cas ternaire par [WvD05]. Dans le cas binaire, il a également été prouvé qu'à α fixé on peut asymptotiquement atteindre cette borne [Bie98] lorsque $n \to +\infty$, et même que cette borne est asymptotiquement saturée pour presque tous les codes linéaires [GK03] lorsque $n \to +\infty$. Mais ceci ne signifie pas pour autant que l'on peut prendre n'importe quel code linéaire, car on se heurte lorsque $n \to +\infty$ à la difficulté de mener un décodage efficace pour la plupart d'entre eux.

Focalisons-nous sur le contexte des *wet papers*, qui est le plus pertinent d'un point de vue applicatif. Aujourd'hui, les codes LGDM de [FF07, ZZW08, ZZW10, FF09b] sont ceux qui donnent la meilleure efficacité d'insertion. Par ailleurs, la construction ZZW et ses dérivées (présentées à la fin de cette introduction, page 22) conservent la quasi-optimalité en termes d'efficacité d'insertion. Les codes convolutifs ont permis d'obtenir des schémas également quasi-optimaux en termes d'efficacité d'insertion, mais capables de surcroît de s'adapter à n'importe quelle mesure de distorsion additive, y compris lorsqu'elle n'est pas identique sur l'ensemble du document de couverture [FJF10, FF10b, FJF11].

Formalisation des problèmes à résoudre et analyse (suite)

Reprenons notre analyse là où nous l'avions laissée. Minimiser la distorsion au maximum revient à chercher le vecteur $\mathbf{e}$ de plus petit poids satisfaisant les équations (1.6) et (1.7). Une stratégie naturelle est donc de s'appuyer sur un décodage unique (par syndrome ou plus élaboré selon les cas) d'un mot de syndrome $\mathbf{m} - \mathbf{x}\cdot\mathbf{H}^t$, décodage qui va justement nous retourner un vecteur $\mathbf{e}$ de plus petit poids satisfaisant l'équation (1.6). Pour prendre en compte l'équation (1.7), il suffit de considérer une fonction de décodage de rayon $T' = T$. C'est cette stratégie qui est adoptée en général, et qui est dévelopée dans les paragraphes suivants. Ses limites et quelques alternatives seront discutées plus loin.

Concentrons-nous sur les considérations purement topologiques de la couverture de l'union $\mathcal{U}_{\mathcal{C},T}$ des boules de rayon T centrées en les mots du code. Le système consitué des équations (1.6) et (1.7) admet une solution unique pour tout message $\mathbf{m}$ et tout vecteur de couverture $\mathbf{x}$ si et seulement si $T \leq \lfloor (d-1)/2 \rfloor$ (unicité) et $T = \varrho$ (pour que tous les syndromes $\mathbf{m} - \mathbf{x}\cdot\mathbf{H}^t$ puissent être atteints). Autrement

10. Ce qui est parfois judicieux, mais pas toujours, comme nous le discuterons page 21, ainsi qu'à la fin de la section 1.3

dit il admet une solution unique pour tout message $\mathbf{m}$ et tout vecteur de couverture $\mathbf{x}$ si et seulement si $\varrho = \lfloor \frac{d-1}{2} \rfloor$, c'est-à-dire si et seulement si le code $\mathcal{C}$ est parfait et que l'on autorise $T = \varrho = \lfloor \frac{d-1}{2} \rfloor$ modifications. Dans tous les autres cas, il y aura des messages qui ne pourront être insérés. Les codes parfaits sont donc très intéressants pour la stéganographie, d'autant que l'on sait très bien les décoder. Ainsi, c'est un code parfait, le code de Hamming, qui est utilisé dans `F5` [Wes01]. Malheureusement, ils donnent lieu à une efficacité d'insertion très en deçà de la borne (1.8) [BF08]. C'est pourquoi on utilise plutôt des codes non parfaits, même si ces derniers ne peuvent résoudre le problème 1.1 que pour des valeurs de $T < \varrho$. Pour ces codes non parfaits, il existe donc des messages qui ne pourront pas être insérés.

Formalisons maintenant ce qui se passe dans le cas d'un canal de sélection non partagé (*wet papers*), qui permet d'assurer une meilleure indétectabilité *via* le verrouillage de certaines composantes de $\mathbf{x}$. Le problème 1.2 prend ainsi en compte les contraintes (1.1), (1.2) et (1.3).

Problème 1.2 (codage par syndrome dans le cas des *wet papers*) *Soient $\mathcal{C}$ un code linéaire de paramètres $[n, k]$, $\mathbf{H}$ une matrice de parité de $\mathcal{C}$, $\mathbf{x} \in \mathbb{F}_q^n$ le vecteur de couverture, $\mathbf{m} \in \mathbb{F}_q^{n-k}$ le message à insérer dans $\mathbf{x}$, $T \in \mathbb{N}^*$ le nombre maximal de modifications autorisées dans $\mathbf{x}$, et $\mathcal{W} \subset \{1, \ldots, n\}$ l'ensemble des composantes de $\mathbf{x}$ considérées comme verrouillées, $\ell_{\mathcal{W}} = |\mathcal{W}|$. Le* problème du codage par syndrome dans le cas des *wet papers consiste à trouver $\mathbf{y} \in \mathbb{F}_q^n$ tel que $\mathbf{y} \cdot \mathbf{H}^t = \mathbf{m}$, $d_H(\mathbf{x}, \mathbf{y}) \leq T$, et $\mathbf{y}_i = \mathbf{x}_i \ \forall i \in \mathcal{W}$.*

Ceci revient à réécrire le système constitué des équations (1.6) et (1.7) en rajoutant la contrainte $\mathbf{e}_i = 0 \ \forall i \in \mathcal{W}$:

$$\begin{aligned} \mathbf{e} \cdot \mathbf{H}^t &= \mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t \\ \mathbf{e}_i &= 0 \ \forall i \in \mathcal{W} \\ w_H(\mathbf{e}) &\leq T \end{aligned} \tag{1.9}$$

On force donc, dans l'équation (1.4), le vecteur $\mathbf{e}$ renvoyé par la fonction D à être nul sur toutes les composantes spécifiées par $\mathcal{W}$. L'équation (1.4) est alors remplacée par l'équation (1.10) :

$$\begin{aligned} \mathbf{y} = \mathrm{Emb}(\mathbf{x}, \mathbf{m}) &= \mathbf{x} + \mathrm{D}^*_{\mathcal{W}}(\mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t) \\ \mathrm{Ext}(\mathbf{y}) &= \mathbf{y} \cdot \mathbf{H}^t \end{aligned} \tag{1.10}$$

La fonction $D^*_{\mathcal{W}}()$ se doit de retourner un vecteur $\mathbf{e}$ de poids suffisamment petit dont les composantes données par l'ensemble $\mathcal{W}$ sont nulles. Du point de vue de la théorie des codes, le calcul de $D^*_{\mathcal{W}}()$ s'apparente au décodage du code *raccourci* $\mathcal{C}^*_{\mathcal{W}}$ obtenu à partir de $\mathcal{C}$ en supprimant les composantes données par l'ensemble $\mathcal{W}$ ($\mathcal{C}^*_{\mathcal{W}}$ est donc de longueur $n - \ell_{\mathcal{W}}$), et en ne conservant que les mots de $\mathcal{C}$ dont la restriction à $\mathcal{W}$ était nulle.

On peut également adopter une autre vision et réécrire la contrainte $\mathbf{y} \cdot \mathbf{H}^t = \mathbf{m}$ comme :

$$\begin{aligned} \mathbf{y} \cdot \mathbf{H}^t &= \mathbf{m} \\ \mathbf{y}_{|\mathcal{D}} \cdot \mathbf{H}_{|\mathcal{D}}{}^t + \mathbf{y}_{|\mathcal{W}} \cdot \mathbf{H}_{|\mathcal{W}}{}^t &= \mathbf{m} \\ \mathbf{y}_{|\mathcal{D}} \cdot \mathbf{H}_{|\mathcal{D}}{}^t &= \mathbf{m} - \mathbf{y}_{|\mathcal{W}} \cdot \mathbf{H}_{|\mathcal{W}}{}^t \\ \mathbf{y}_{|\mathcal{D}} \cdot \mathbf{H}_{|\mathcal{D}}{}^t &= \mathbf{m} - \mathbf{x}_{\mathcal{W}} \cdot \mathbf{H}_{|\mathcal{W}}{}^t \end{aligned}$$

avec $\mathbf{y}_{|\mathcal{W}}$ (respectivement $\mathbf{x}_{\mathcal{W}}$) le vecteur de longueur $\ell_{\mathcal{W}}$ obtenu par restriction de $\mathbf{y}$ (respectivement $\mathbf{x}$) aux composantes spécifiées par $\mathcal{W}$; $\mathbf{H}_{|\mathcal{W}}$ la matrice de taille $(n-k) \times \ell_{\mathcal{W}}$ obtenue par restriction de $\mathbf{H}$ aux colonnes spécifiées par $\mathcal{W}$; $\mathbf{y}_{|\mathcal{D}}$ (vecteur de longueur $n - \ell_{\mathcal{W}}$) et $\mathbf{H}_{|\mathcal{D}}$ (matrice [11] de taille $(n-k) \times (n - \ell_{\mathcal{W}})$) sont définis de manière similaire, en notant $\mathcal{D} = \{1, \ldots, n\} \setminus \mathcal{W}$ l'ensemble complémentaire de $\mathcal{W}$, autrement dit l'ensemble des composantes modifiables (*dry components*).

Ce système admet une solution si et seulement si la matrice $\mathbf{H}_{|\mathcal{D}}$ est inversible. Or, si on considère un code de matrice de parité $\mathbf{H}$, on sait que dire que toute sous-matrice de taille $(n-k) \times (n-k)$ de $\mathbf{H}$ est inversible est équivalent à dire que le code est MDS [HP03, Cor 1.4.14]. Ainsi, les codes MDS permettent d'assurer l'insertion lorsque l'on verrouille au plus k composantes. Malheureusement, les codes MDS binaires n'existent que pour des longueurs 2 ou 3. Si l'on souhaite utiliser un code MDS plus long, on est obligé de passer en q-aire. C'est ce qui nous a poussés à étudier avec Fabien Galand le cas des codes de Reed-Solomon, qui sont MDS.

La résolution du problème 1.2 est donc bien plus difficile que celle du problème 1.1, et même les codes parfaits peuvent échouer [FGS06]. Ainsi, si on se restreint à une approche classique par décodage unique, on sait que la probabilité d'échec de l'insertion ne sera jamais nulle.

De plus, la probabilité que le message soit inséré avec succès décroît exponentiellement quand sa longueur augmente. Supposons que compte tenu de sa longueur, le message doive être découpé, disons, en L blocs, chaque bloc étant inséré indépendamment dans des documents de couverture différents. Si on note P la probabilité que l'insertion d'un bloc de message réussisse, alors la probabilité que le message total réussisse à être inséré est P^L.

En dehors du cas particulier des codes MDS, qui sont les seuls à assurer l'insertion, on ne peut s'appuyer sur une structure particulière de la matrice $\mathbf{H}$ pour estimer la résolvabilité du système. On peut donc tout aussi bien considérer que $\mathbf{H}$ est aléatoire, ce qui permet par ailleurs de maximiser asymptotiquement l'efficacité d'insertion moyenne [GK03, FGS06]. Dans ce cas, on peut utiliser pour estimer P la borne donnée dans [BGL01] : $\mathbf{H}_{|\mathcal{D}}$ étant considérée comme une matrice aléatoire de $n-k$ lignes et $n - \ell_{\mathcal{W}}$ colonnes, avec $n - \ell_{\mathcal{W}} \geq n - k$, on a alors

$$P \geq \begin{cases} 0.288, & \text{si } n - \ell_{\mathcal{W}} = n - k \text{ et } q = 2, \\ 1 - \frac{1}{q^{k - \ell_{\mathcal{W}}}(q-1)}, & \text{sinon.} \end{cases}$$

11. $\mathbf{H}_{|\mathcal{D}}$ est précisément la matrice de parité du code raccourci $\mathcal{C}^*_{\mathcal{W}}$

Dans la grande majorité des cas, l'insertion a donc une probabilité non nulle (et non négligeable si le message est long) d'échouer. Que peut alors faire l'émetteur s'il souhaite à tout prix envoyer son message ? S'il le peut, il choisira probablement de l'insérer dans d'autres documents de couverture. Et s'il n'en a pas suffisamment à sa disposition ? On trouve alors dans la littérature deux propositions : soit verrouiller moins de composantes (et dégrader légèrement la furtivité) [FJF10], soit découper le message en blocs plus petits pour faciliter leur insertion (quitte à perdre en efficacité d'insertion). Mais dans ces deux cas, même s'il augmente ses chances de voir l'insertion réussir, il n'en a aucune garantie.

On peut remarquer que ce risque d'échec est renforcé par le fait que la stratégie de décodage envisagée par défaut est un décodage unique. On peut alors être tenté d'adopter une autre approche. On peut par exemple autoriser la fonction $D()$ (ou $D^*_{\mathcal{W}}()$ dans le cas d'un canal de sélection non partagé) à retourner non pas un unique vecteur $\mathbf{e}$, mais un ensemble de vecteurs parmi lesquels on pourra choisir celui qui donne lieu à la plus faible distortion. Ceci permet de prendre en compte un deuxième modèle de distorsion, et de ne pas se limiter au cas de la minimisation du nombre de modifications comme critère final de choix. Une troisième approche est de résoudre le système autrement que par décodage, pour obtenir si possible toutes les solutions satisfaisant l'ensemble des contraintes, et choisir ensuite celle qui sera la plus adaptée en termes de furtivité. Si on arrive ainsi à résoudre le système de manière effective (en décodant jusqu'à un rayon $T = \varrho$, ou en résolvant directement le système), alors on est même certains de pouvoir insérer le message, même si le code n'est pas parfait. C'est par exemple ce qui se passe dans la section 1.2 avec les codes de Reed-Solomon, où l'on propose de résoudre le système grâce à une interpolation de Lagrange.

Une dernière solution pour assurer l'insertion est de modifier le schéma classique de codage par syndrome. C'est ce que nous avons proposé avec Daniel Augot et Morgan Barbier, comme discuté dans la section 1.3.

Quel code choisir ?

Comme nous l'avons vu, choisir un code approprié est délicat. Car il faut que le code nous donne un moyen efficace de calculer $\mathbf{y}$, que ce soit par des techniques de décodage ou par une résolution directe du système. Si on considère le cas du décodage, il faut non seulement que le code ait des paramètres connus et appropriés, pour assurer l'existance de solutions sur le plan topolagique, mais aussi que l'on dispose d'un algorithme de décodage efficace. On se heure alors à des problèmes difficiles célèbres : détermination du rayon de recouvrement, décodage, etc.

Plusieurs familles de codes ont été étudiées. Il a tout d'abord été démontré que si les codes aléatoires ont un comportement asymptotique intéressant [Bar98], leur utilisation se heurte à des problèmes difficiles, car le décodage par syndrome et le calcul du rayon de recouvrement de ces codes ont été prouvés respectivement NP-complet et Π_2-complet [Var97, McL84]. De plus, on ne connaît pas d'algorithme de décodage efficace pour ces codes en général, ni même pour des sous-familles. Leur utilisation dans des systèmes de stéganographie effectifs n'est donc pas envisageable.

L'attention s'est donc naturellement portée vers des codes plus structurés, et surtout pour lesquels nous disposons d'algorithmes de décodage efficaces : codes de Hamming [Cra98,FGS05a,ZW06,FL07], codes de Golay [ZW06,Mie06,FL07], codes du Simplex [FS06], codes BCH [SW06,SW07,ZSK09,SKZ09,OMS10], codes de Reed-Solomon [FG07b, FG09], codes produits parfaits [RPR09, RR10], codes à matrice génératrice creuse (LGDM) [FF07,ZZW08,ZZW10,FF09b], codes convolutifs [FJF10, FF10b,FJF11].

Focalisons-nous sur le contexte des *wet papers*, qui est le plus pertinent d'un point de vue applicatif. Les codes LGDM de [FF07,ZZW08,ZZW10,FF09b] sont ceux qui donnent la meilleure efficacité d'insertion, mais avec une probabilité d'échec à l'insertion non négligeable. Les codes convolutifs ont permis d'obtenir des schémas également quasi-optimaux en termes d'efficacité d'insertion, mais capables de surcroît de s'adapter à n'importe quelle mesure de distorsion additive, y compris lorsqu'elle n'est pas identique sur l'ensemble du document de couverture [FJF10,FF10b,FJF11]. Les codes de Reed-Solomon sont les seuls à offrir une assurance du succès de l'insertion [FG07b,FG09].

Quel alphabet utiliser : binaire, ternaire, q-aire ?

La plupart des techniques d'insertion s'appuient sur des vecteurs binaires (suites de bits de poids faibles des pixels ou de coefficients DCT par exemple), même s'ils sont parfois mis en relation avec des vecteurs sur d'autres alphabets à un moment ou un autre de l'algorithme. La possibilité d'utiliser une information ternaire peut s'appliquer sur la plupart des insertions binaires, et permet alors d'augmenter l'efficacité d'insertion. Ainsi, au lieu de pouvoir par exemple augmenter ou conserver la valeur d'un pixel ou d'un coefficient de transformée (*e.g.* DCT, Fourier, ondelettes), on peut choisir de le diminuer. On parle alors souvent de stéganographie « ± 1 ». Dans ce contexte où les données de base sont binaires, il a été montré que lorsque l'on s'attache à minimiser le nombre de modifications la meilleure efficacité d'insertion est justement obtenue avec de tels codes ternaires [FLS06]. Willems *et al.* ont été les premiers à étudier ce cas de figure et poser les limites de ce qu'elle permet, notamment en terme d'efficacité d'insertion [WvD05]. Ils ont par ailleurs proposé l'utilisation des codes de Hamming et de Golay ternaires pour réaliser un codage par syndrome ternaire. Deux autres travaux ont généralisé indépendamment cette étude [ZW06,Mie06,FL07]. Zhang *et al.* ont par ailleurs proposé dans [ZZW07] une construction dite en double couche permettant de construire à partir de n'importe quel stégo-code binaire un schéma d'insertion de stéganographie « ± 1 ». Ils ont également prouvé que si le stégo-code binaire a une efficacité d'insertion presque optimale, alors le schéma ternaire ainsi construit a lui aussi une efficacité d'insertion presque optimale. Les constructions ZZW et ses variantes, présentées dans la section suivante, suivent le principe développé dans [ZZW07], et conservent cette propriété de préservation d'une efficacité d'insertion quasi-optimale, dans les cas binaires comme ternaires.

Mais ceci ne veut pas dire que le cas q-aire ne présente pas d'intérêt lorsque $q > 3$, car pourquoi faire l'hypothèse que l'on part de données binaires ? C'est effectivement

le cas pour la plupart des schémas, mais on peut très bien envisager des fonctions d'extraction de données offrant nativement des données de couverture q-aires. Dans ce cas, on n'a plus à relier la modification des bits et celle des symboles, puisque l'on raisonne directement au niveaux du symbole lui-même, et les résultats d'optimisation mentionnés dans le paragraphe précédent ainsi que dans l'équation (1.8) ne sont plus valables. Un exemple de fonction d'extraction permettant de constituer le vecteur de couverture de cette manière est donnée dans [FL07]. Cette piste reste aujourd'hui peu envisagée, alors qu'elle ouvre des possibilités nouvelles. Comme nous le verrons dans la section 1.2, les codes de Reed-Solomon pourraient très bien être utilisés dans ce contexte, d'autant qu'ils présentent d'excellentes propriétés.

L'insertion à deux niveaux de Zhang, Zhang et Wang (ZZW)

En 2008, Zhang *et al.* ont proposé une construction permettant, à nombre de modifications moyen R_a fixé, d'utiliser un code binaire $\mathcal{C}_1$ prévu pour insérer m bits de message dans des données de longueur n pour construire une famille de codes permettant d'insérer en moyenne $m + bR_a$ bits de message dans des données de longueur $n2^b$ avec le même nombre R_a de modifications [ZZW08].

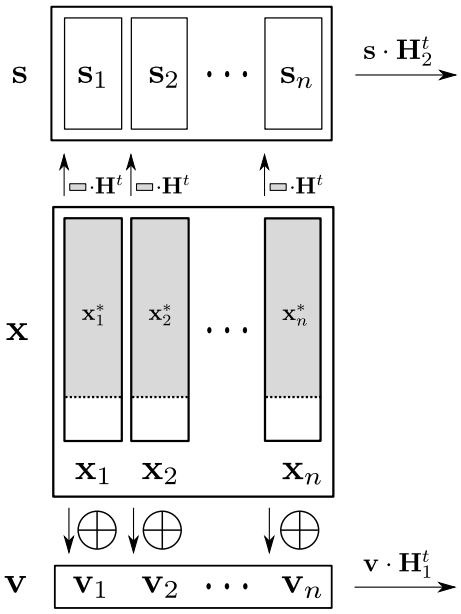


FIGURE 1.3: Insertion à deux niveaux de Zhang, Zhang et Wang (ZZW)

L'idée générale de leur construction, illustrée par la figure 1.3, est de découper le vecteur des données de couverture $\mathbf{x}$, de longueur $n2^b$, en n vecteurs $\mathbf{x}_i$ de longueur 2^b, $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, et d'associer à chacun de ces blocs un bit résultant de leur XOR : $\mathbf{v}_i = \oplus_{j=1}^{2^b} \mathbf{x}_i[j]$. Ils obtiennent ainsi un vecteur binaire artificiel $\mathbf{v}$ de longueur n, grâce auquel ils procèdent à une insertion par syndrome de m bits de message grâce au code $\mathcal{C}_1$ de matrice de parité $\mathbf{H}_1$. Le syndrome final $\mathbf{v} \cdot \mathbf{H}_1^t$ sera précisément égal à ce message de m bits. Ce premier niveau d'insertion détermine quelles sont les

R_a composantes de $\mathbf{v}$ qui doivent être modifiées pour porter les m bits de message. Soit $\mathcal{S}$ l'ensemble des indices de ces composantes : $\mathcal{S} = \{i, \mathbf{v}_i \text{ doit être modifié}\}$. Leur objectif est ensuite de modifier une composante de chaque $\mathbf{x}_i, i \in \mathcal{S}$ pour effectuer cette insertion. Mais ce choix n'est pas opéré au hasard, et va permettre de cacher de nouveaux bits de message. Voici comment. Considérons la matrice de parité $\mathbf{H}$ d'un code de Hamming de longueur $2^b - 1$, et calculons les syndromes associés à tous les $\mathbf{x}_i$: $\mathbf{s}_i = \mathbf{x}_i^* \cdot \mathbf{H}^t$, $i = 1, \ldots, n$, $\mathbf{x}_i^*$ désignant le vecteur constitué des $2^b - 1$ premières composantes de $\mathbf{x}_i$. Ce vecteur $\mathbf{s} = (\mathbf{s}_1, \ldots, \mathbf{s}_n)$ de longueur nb sert alors de support pour une deuxième insertion. A l'aide d'un deuxième code $\mathcal{C}_2$ de matrice de parité $\mathbf{H}_2$ on modifie $\mathbf{s}$ pour y insérer de nouveaux bits de message : le syndrome final $\mathbf{s} \cdot \mathbf{H}_2^t$ sera précisément égal à ce nouveau message de nb bits. Cette deuxième insertion est contrainte par le fait que les portions $\mathbf{s}_i$ de $\mathbf{s}$ correspondant aux $i \notin \mathcal{S}$ ne doivent pas être modifiées car les $\mathbf{x}_i$ doivent rester inchangés. En revanche, les autres portions $\mathbf{s}_i$ de $\mathbf{s}$ correspondant aux $i \in \mathcal{S}$ peuvent être modifiées. Pour gérer cette contrainte, les auteurs suggèrent d'utiliser un code de type *wet paper* $\mathcal{C}_2$, en considérant toutes les composantes des portions $\mathbf{s}_i$, $i \notin \mathcal{S}$, comme verrouillées. Par le résultat sur la capacité rappelé dans le paragraphe de présentation de l'approche des *wet papers*, comme le vecteur $\mathbf{s}$ contient bR_a composantes modifiables, cette insertion permet de transmettre en moyenne bR_a bits de message supplémentaires. On peut par exemple utiliser pour $\mathcal{C}_2$ un code LT [FGS05a] [12]. On peut remarquer que le nombre de modifications que le code $\mathcal{C}_2$ introduit dans le vecteur $\mathbf{s}$ n'a en fait pas d'importance car le code de Hamming va permettre de répercuter toutes les modifications en changeant seulement 1 bit sur chacun des $\mathbf{x}_i^*$, $i \in \mathcal{S}$, concernés. Notons $\mathbf{s}^{\text{new}}$ le vecteur de syndromes obtenu après insertion, et regardons maintenant précisément comment répercuter ces modifications sur $\mathbf{x}$. Rappelons que pour tous les $i \notin \mathcal{S}$, $\mathbf{s}_i^{\text{new}} = \mathbf{s}_i$ car ces composantes ont été verrouillées. En revanche les autres peuvent avoir été modifiées par l'insertion du message, et il reste à modifier les $\mathbf{x}_i$, $i \in \mathcal{S}$, en conséquence. Pour chacun de ces $\mathbf{x}_i$, $i \in \mathcal{S}$: soit le syndrome $\mathbf{x}_i^* \cdot \mathbf{H}^t$ a déjà la bonne valeur $\mathbf{s}_i^{\text{new}}$, et on modifie alors le bit $\mathbf{x}_i[2^b]$ qui n'intervenait pas dans $\mathbf{x}_i^*$; soit le syndrome $\mathbf{x}_i^* \cdot \mathbf{H}^t$ n'a pas la bonne valeur $\mathbf{s}_i^{\text{new}}$, et on modifie alors 1 bit de $\mathbf{x}_i^*$ pour que le syndrome prenne la bonne valeur. Dans tous les cas, on aura changé 1 bit de chaque $\mathbf{x}_i$, $i \in \mathcal{S}$. Au total, $\mathcal{S}$ étant de cardinal R_a, on aura donc changé R_a bits de $\mathbf{x}$.

Ainsi, en partant d'un schéma (code $\mathcal{C}_1$) permettant d'insérer des messages de longueur m dans un support de longueur n en effectuant R_a modifications, on peut construire un schéma permettant d'insérer des messages de longueur $m + bR_a$ dans un support de longueur $n2^b$ en effectuant R_a modifications. Cette technique permet donc, à nombre de modifications fixé, d'améliorer l'efficacité d'insertion en insérant plus de bits de message grâce à cette deuxième couche d'insertion. Bien sûr, cette amélioration nécessite que l'on puisse se permettre d'utiliser des vecteurs de données de couverture suffisamment longs. Plusieurs codes ont été testés pour le premier

12. On notera que la probabilité d'échec de l'insertion d'un tel code diminue avec la longueur du vecteur de données support, et que l'on a donc intérêt à concaténer un maximum de syndromes avant d'effectuer l'insertion à ce deuxième niveau, par exemple en traitant plusieurs blocs d'une image d'un seul coup plutôt que successivement.

niveau d'insertion, d'abord des codes structurés comme les codes de Hamming, de Golay, ou les BCH, puis avec des codes à matrice génératrice creuse. Zhang *et al.* ont montré que les codes LGDM proposés par Fridrich *et al.* dans [FF07] sont ceux qui permettent aujourd'hui la meilleure efficacité d'insertion, frôlant la borne donnée plus haut. Ces résultats expérimentaux de [ZZW08] sont conformes à ceux exposés et discutés dans [FF07] lorsque ces codes étaient utilisés seuls sur un schéma classique en une couche. Mais ici, associés à la deuxième couche, ils permettent d'aller encore plus loin. De plus, alors qu'en couche simple l'efficacité d'insertion s'éloignait de la borne lorsque α était petit, elle reste ici dans ce schéma à double couche proche de la borne, même lorsque le α général est petit (par construction, le « α » du schéma global est plus petit que le « α » du schéma utilisé au premier niveau).

Les auteurs ont également abordé le cas de la stéganographie « ± 1 », qui utilise des codes ternaires, et pour lequel il suffit de modifier très légèrement la construction proposée dans le cas binaire. Les résultats obtenus dans le cas ternaire sont tout-à-fait similaires à ceux du cas binaire : l'efficacité d'insertion s'avère expérimentalement là aussi proche de la borne théorique. Cette stratégie d'insertion présente donc un très fort intérêt en pratique car elle est presque optimale en terme d'efficacité d'insertion. Fridrich a apporté dans [Fri09a] une preuve théorique de ces observations, explicitant l'écart entre l'efficacité d'insertion et la borne théorique. Bien sûr, cette efficacité s'obtient au détriment de la rapidité d'insertion, puisque l'on doit appliquer plusieurs codages par syndrome.

La construction ZZW originale permet de construire de grandes familles de codes appropriés pour la stéganographie, mais ne couvre qu'un ensemble assez clairsemé de longueurs de messages relatives α. Zhang *et al.* ont proposé une généralisation de cette construction permettant de couvrir un ensemble plus dense de longueurs de message relatives [ZW09].

Fridrich *et al.* et Zhang *et al.* ont ensuite proposé en parallèle deux généralisations équivalentes de la construction ZZW permettant d'insérer les bits de message supplémentaires dans des blocs de longueur N autre que 2^b, et permettant également de gérer des composantes verrouillées. Ces deux généralisations, qui ne sont pas détaillées ici, sont connues sous les noms respectifs de *Wet ZZW* [FF09b] pour la première, et *Paper Folding* [ZZ09] puis *N-page* [ZLWY10] pour la seconde, qui s'est déroulée en deux étapes. Les expérimentations ont montré que l'efficacité d'insertion de ces solutions est excellente, mais qu'elle décroit lorsque le nombre de composantes verrouillées augmente. Une preuve de l'efficacité d'insertion de la construction 2^b-page (qui correspond au cas du *Paper Folding*), ainsi qu'une estimation de l'influence du nombre de composantes verrouillées sur l'efficacité d'insertion sont données dans [ZLWY10]. Ainsi, pour cette construction l'écart estimé entre l'efficacité d'insertion et la borne varierait entre 0.05 pour 10% de composantes verrouillées et 0.32 pour 99% de composantes verrouillées, ces résultats asymptotiques étant à comparer avec des valeurs absolues pour l'efficacité d'insertion comprises entre 6 et 8 pour α variant de 1/6 à 1/20.

Ces constructions sont très intéressantes du point de vue de l'efficacité d'insertion. Elles le sont évidemment d'autant plus que l'on dispose de longs ou de nombreux vecteurs de données $\mathbf{x}$. Par ailleurs, la *square root law* de Ker nous impose, pour

conserver un niveau de sécurité similaire lorsque l'on fait croître n, de faire décroître le paramètre α plus vite que $1/\sqrt{n}$ [FKF09]. Et ces constructions permettent de construire des schémas à efficacité d'insertion quasi-optimale pour ce type de longueur relative de message α.

On peut remarquer que ce type de construction n'a été envisagé ici que sous l'angle binaire ou ternaire. Là encore, le cas q-aire reste non étudié. Pourtant, comme nous le verrons dans la section suivante, les codes de Reed-Solomon (qui sont q-aires) offrent toutes les propriétés requises pour être alors utilisés dans ce type de construction, y compris dans le cas des *wet papers*.

1.2 Une meilleure gestion des composantes verrouillées grâce aux codes de Reed-Solomon : séjour post-doctoral de F. Galand (2006-2007)

Avec Fabien Galand, nous nous sommes particulièrement intéressés à l'article de Schönfeld *et al.* qui traite de l'utilisation des codes BCH binaires dans un schéma de codage par syndrome [SW06]. Les auteurs y ont montré que si l'on ne souhaite pas verrouiller de composantes, les codes BCH offrent une bonne solution, mais soulignent néanmoins la difficulté à déterminer quel BCH est le plus approprié, car nous ne connaissons pas de bon algorithme de décodage complet par syndrome pour ces codes. Considérant ensuite la gestion de composantes verrouillées, ils montrent que les codes BCH de dimension k ne permettent pas de gérer plus de $k/2$ composantes verrouillées sans échec, et ne sont donc pas optimaux.

Nous avons étudié le cas des codes de Reed-Solomon et montré que, bien que de la même famille que les BCH, ils permettent néanmoins une gestion optimale de ces composantes verrouillées puisqu'ils sont capables d'en gérer jusqu'à k sans échouer. Ils offrent par ailleurs de nombreux avantages :

- leur distance minimale les rend idéaux (ils sont MDS),
- leur rayon de recouvrement est connu et sa valeur est intéressante pour une application telle que la stéganographie,
- le décodage complet est peu coûteux, et on dispose par ailleurs de la possibilité d'utiliser un décodage en liste, ce qui nous donne plusieurs candidats, parmi lesquels on peut alors choisir le plus approprié (celui qui donnera la plus petite distortion).

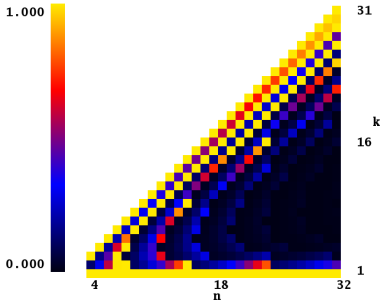
Nous avons proposé une solution algorithmique pour l'insertion d'un bloc de message à l'aide des codes de Reed-Solomon. Cette solution comporte une première proposition reposant sur une interpolation de Lagrange, qui permet une utilisation efficace de ces codes pour un faible coût algorithmique. Elle est optimale au regard de la gestion des positions verrouillées, puisqu'elle permet d'en gérer jusqu'à k. Elle offre par ailleurs l'avantage de ne jamais échouer, même si ces codes ne sont pas des codes parfaits, car on s'autorise ici des solutions de poids non minimal, tant que ce dernier reste inférieur ou égal à T. Nous avons également proposé une deuxième version reposant sur l'algorithme de décodage en liste de Guruswami-Sudan. Cette deuxième stratégie est plus lente et plus incertaine, mais plus intéressante en termes

de nombre de modifications, et donc d'efficacité d'insertion. Pour certaines valeurs des paramètres (longueur n, dimension k), on peut insérer le même bloc de message en introduisant jusqu'à 50% de modifications en moins si on utilise cette stratégie plutôt qu'une simple interpolation, tout en préservant la gestion des composantes verrouillées. Nous avons donc finalement proposé une solution hybride, qui s'appuie sur le décodage en liste quand il est possible, et se rabat sur l'interpolation sinon, offrant ainsi le meilleur compromis entre : une gestion sans échec d'au plus k composantes verrouillées, et une efficacité d'insertion et une distorsion aussi faibles que possible. Nous avons par ailleurs donné une estimation de la probabilité que le décodage en liste aboutisse, en fonction des paramètres n et k. La figure 1.4 illustre, pour $q = 64$ (les codes de Reed-Solomon sont des codes q-aires), les paramètres pour lesquels le décodage en liste est intéressant, ainsi que le gain en termes de modifications par rapport à une simple interpolation (le cas $q = 32$ est donné ici pour montrer plus en détail ce qui se passe sur les bords du triangle). Ces figures sont représentatives d'autres valeurs de q.

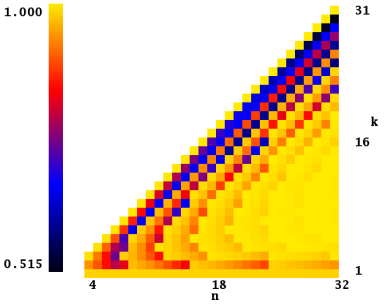
Ces résultats ont été présentés lors de la conférence internationale *International Workshop on Information Hiding, IH'07* [FG07b], ainsi que dans un article de revue publié dans le numéro spécial *Secure Steganography in Multimedia Content* du journal international *EURASIP Journal on Information Security* [FG09] [13]. Ces travaux ont été valorisés dans le cadre du réseau d'excellence européen ECRYPT.

Nous avons ainsi validé la pertinence de ces codes dans la conception de systèmes stéganographiques. Néanmoins, deux remarques sont à souligner quant à la possible utilisation de ces résultats dans des schémas opérationnels. La première est liée à la taille de l'alphabet, qui ne peut être binaire ; ce n'est pas vraiment un problème, même si les systèmes opérationnels s'appuient traditionnellement sur des vecteurs binaires, puisqu'il existe des techniques d'extraction pour obtenir $\mathbf{y}$ à partir du document qui donnent des symboles non binaires. La deuxième remarque concerne l'implémentation des stratégies d'insertion. Car s'il est aisé d'implémenter la stratégie reposant sur l'implémentation de Lagrange, aucune implémentation optimisée de l'algorithme de Guruswami-Sudan n'est actuellement disponible ; il faudrait donc en réaliser une, ce qui n'est pas chose aisée.

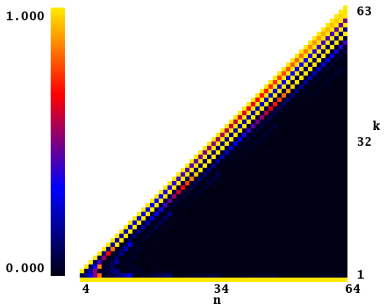
13. Article joint en annexe D.



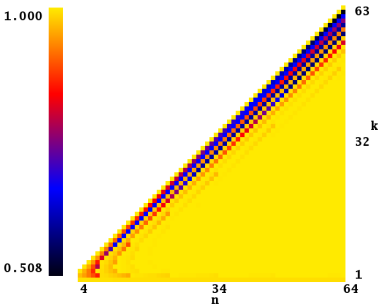
(a) Probabilité de succès du décodage en liste pour une entrée aléatoire, pour $q = 32$



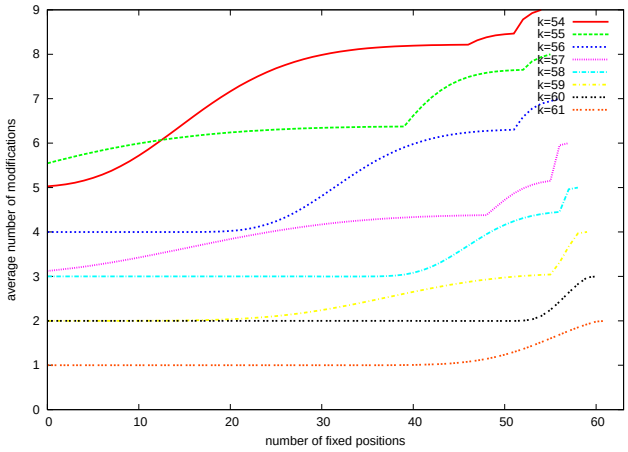
(b) Nombre moyen de modifications, pour $q = 32$



(c) Probabilité de succès du décodage en liste pour une entrée aléatoire, pour $q = 64$



(d) Nombre moyen de modifications, pour $q = 64$



(e) Nombre moyen de modifications en fonction du nombre de composantes verrouillées avec l'approche hybride, pour $q = 64$. On ne trace que les cas où le gain entre décodage en liste et interpolation de Lagrange (correspondant aux valeurs à droite en bout de courbes) est significatif. Pour chacune de ces courbes, une interpolation de Lagrange seule donnerait une valeur constante égale à la valeur la plus grande observée.

FIGURE 1.4: Quelques repères concernant l'efficacité du décodage en liste par rapport à l'interpolation de Lagrange.

1.3 Assurer l'insertion, coûte que coûte : thèse de M. Barbier (2008-2011)

J'ai poursuivi mes recherches sur les liens entre codes correcteurs et stéganographie en collaboration avec Morgan Barbier et Daniel Augot. Partis du constat que l'approche classique du codage par syndrome connaît une probabilité d'échec (la plupart du temps non nulle) qui augmente exponentiellement avec la longueur du message à cacher, *i.e.* avec le nombre de blocs à cacher [FGS05a], nous avons cherché comment garantir son succès.

On peut pour cela soit utiliser un codage par syndrome classique avec un code parfait et sans verrouiller de composantes, soit conserver la structure du codage par syndrome et s'affranchir de l'utilisation d'algorithmes de décodage unique si cela est possible, comme dans la section précédente, soit repenser la structure du schéma. C'est cette approche que nous avons suivie, proposant une nouvelle construction qui garantit l'insertion quelque soit le code utilisé. Notre idée a été de proposer une variante de l'approche classique, dans laquelle le bloc de message à cacher $\mathbf{m}$ ne représente qu'une partie du syndrome, l'autre partie $\mathbf{R}$ étant laissée libre, comme dans le schéma de signature proposé en 2001 par Courtois *et al.* [CFS01] :

$$\begin{aligned} \mathrm{Emb}(\mathbf{x}, \mathbf{m}) &= \mathbf{x} + \mathrm{D}^*_{\mathcal{W}}((\mathbf{m}||\mathbf{R}) - \mathbf{x} \cdot \mathbf{H}^t) \\ \mathrm{Ext}(\mathbf{y}) &= \mathbf{y} \cdot \mathbf{H}^t \end{aligned}$$

A code fixé, la taille du syndrome est la même dans cette nouvelle approche que dans le schéma classique, soit $n-k$; la longueur du bloc de message que nous pouvons cacher est donc plus petite qu'avant, et est donnée par $n-k-r$, où r est la longueur de $\mathbf{R}$. C'est le prix à payer en termes d'efficacité d'insertion et de longueur relative de message pour être certain de pouvoir cacher le bloc de message.

La perte relative en termes d'efficacité d'insertion par rapport à l'approche classique est donnée par

$$\frac{e - e'}{e} = \frac{r}{n-k},$$

avec e l'efficacité d'insertion du schéma classique, et e' l'efficacité d'insertion de notre nouveau schéma. Pour estimer cette perte, il faut donc être capable d'estimer pour quelles valeurs de r notre schéma admet toujours une solution. Nous nous sommes tout d'abord focalisés sur le cas des codes parfaits linéaires : codes de Hamming et de Golay, exhibant des conditions suffisantes sur r pour garantir l'insertion.

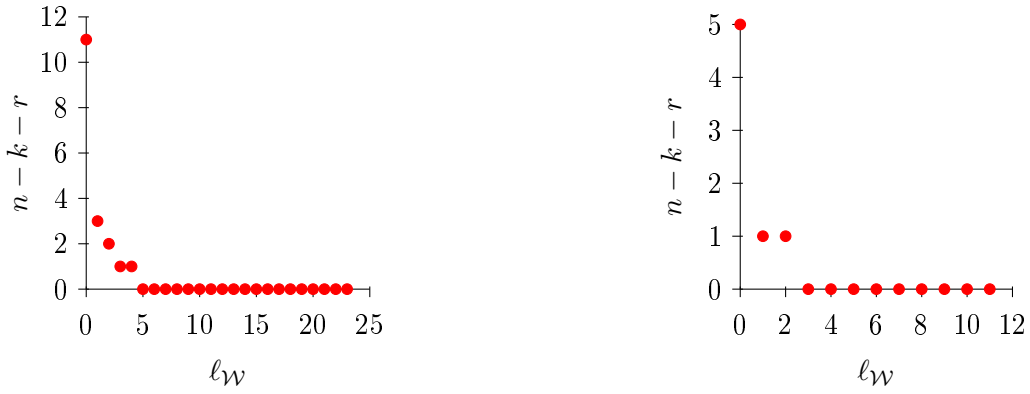
Nous avons ainsi montré que pour le code de Golay binaire de paramètres $[23, 12, 7]$ le schéma admet toujours une solution si

$$r \geq \log_2 \left(1 + \frac{796}{3}\ell_{\mathcal{W}} - \frac{23}{2}\ell_{\mathcal{W}}^2 + \frac{1}{6}\ell_{\mathcal{W}}^3\right).$$

Pour le code de Golay ternaire de paramètres $[11, 6, 5]$, le schéma admet toujours une solution si

$$r \geq \log_3 \left(1 + 44\ell_{\mathcal{W}} - 2\ell_{\mathcal{W}}^2\right).$$

La figure 1.5 montre que pour ces deux codes, le nombre $n-k-r$ de symboles utilisables pour l'insertion décroît très vite lorsque le nombre de composantes verrouillées $\ell_{\mathcal{W}}$ augmente.



(a) Code de Golay binaire

(b) Code de Golay ternaire

FIGURE 1.5: Évolution du nombre $n-k-r$ de symboles de message que l'on peut insérer, en fonction du nombre de composantes verrouillées $\ell_{\mathcal{W}}$.

Dans le cas des codes de Hamming q-aires, de paramètres $[(q^p-1)/(q-1), n-p, 3]$, nous avons montré que l'on peut insérer au moins un symbole de message tant que $\ell_{\mathcal{W}} \leq n/q$, et que l'insertion est garantie si

$$r \geq \lceil \log_q((q-1)\ell_{\mathcal{W}} + 1) \rceil.$$

Asymptotiquement, la perte relative en efficacité d'insertion est alors de

$$\frac{\lceil \log_q((q-1)\ell_{\mathcal{W}} + 1) \rceil}{p}.$$

Cette perte est tout-à-fait acceptable, compte tenu des contraintes que nous avons imposées au système.

Ce schéma est donc intéressant. Mais pour qu'il puisse être utilisé, il faut que le destinataire du stégo-document sache quelle valeur de r a été utilisée à l'insertion (r est calculé dynamiquement lors de l'insertion pour minimiser la perte en efficacité d'insertion). Nous avons donc proposé d'encapsuler ce schéma dans un schéma plus complexe, reposant sur la construction dite ZZW présentée page 22, qui permet de transmettre la valeur de r utilisée. Cette étape, assez technique, n'est pas présentée ici.

Ce travail a été présenté en janvier 2011 lors des journées nationales *Codes et Stéganographie*. Il sera publié en décembre 2011 dans les actes de la conférence internationale *IMA Conference on Cryptography and Coding 2011* [ABF11] [14].

14. Article joint en annexe D.

Comme nous l'avons vu, les codes de Hamming offrent une assez bonne solution pour ce schéma. Néanmoins, il serait intéressant de voir si d'autres codes donnent lieu à une perte en efficacité d'insertion acceptable, et de voir comment affiner au mieux l'estimation des valeurs de r qui garantissent le succès de l'insertion.

1.4 Conclusion et perspectives

Les deux contributions présentées dans ce chapitre s'attachent à montrer comment l'on peut assurer l'insertion du message dans un document de couverture donné. Elles proposent deux pistes bien distinctes, offrant toutes deux de nombreuses perspectives.

Comme nous l'avons souligné dans l'introduction de ce chapitre, ainsi que dans la section 1.2, l'utilisation de codes q-aires dans des schémas stéganographiques n'a pour l'instant pas retenu l'attention de la communauté. Pourtant, les codes de Reed-Solomon possèdent toutes les qualités requises pour motiver l'élaboration de tels schémas. C'est donc une perspective naturelle que de tenter l'expérience, en s'appuyant par exemple sur l'extraction proposée dans [FL07] pour constituer le vecteur de couverture. Une deuxième perspective concerne l'extension de la constrution ZZW au cas q-aire.

Dans la section 1.3, nous avons introduit une nouvelle variante du codage par syndrome, qui garantit le succès de l'insertion. Cette étude, récente et prometteuse, doit être poussée plus avant afin d'en explorer tout le potentiel.

Chapitre 2

Protéger le droit d'auteur grâce à un tatouage robuste et sûr

> Attacks always get better ;
> they never get worse.
>
> Bruce Schneier

Le tatouage relève, comme la stéganographie, de la dissimulation d'information. Il en est cependant assez éloigné : d'abord parce que l'exigence de furtivité est moindre, et que l'on se contente d'une marque imperceptible pour l'utilisateur, qui ne pertubera pas l'utilisation normale du document ; ensuite parce qu'il demande en contrepartie une bonne maîtrise de la robustesse de la marque insérée [1]. Le tatouage sert en général à insérer des informations de protection, visant par exemple à :

- identifier l'ayant droit du document,
- ou identifier son utilisateur légitime,
- ou détecter d'éventuelles modifications de ce document.

Dans les deux premiers cas, on utilisera un tatouage *robuste*, avec une marque la plus indélébile possible, car sans la présence de la marque on ne sera plus en mesure d'identifier l'ayant droit ou l'utilisateur du document. Dans le dernier cas, en revanche, on s'orientera vers un tatouage *fragile* ou *semi-fragile*, dont la marque disparaîtra dès qu'une modification non autorisée aura été appliquée au document ; c'est alors l'absence de la marque qui révélera la falsification.

Mes travaux s'étant concentrés sur le cas du tatouage robuste, seul ce type de tatouage est considéré ici.

1. On ne s'intéresse pas ici au cas de l'enrichissement de contenu, pour lequel la robustesse n'a pas vraiment d'importance. Ce dernier relève de la dissimulation d'information, mais ne relève pas du tatouage tel que nous l'avons défini.

2.1 Introduction au tatouage robuste

Apposer sa marque sur un objet physique pour en attester sa propriété, ou son authenticité (marque de fabrique par exemple) n'a rien de nouveau. La plupart de ces marques sont visibles, ce qui permet à tout le monde de les voir aisément, et au propriétaire ou fabriquant de se faire connaître. Néanmoins, si on souhaite les supprimer, on sait tout de suite sur quelles zone de l'objet concentrer ses efforts. Ceci facilite également la contrefaçon, puisque l'on sait quelle marque copier. On peut bien sûr recourir à des procéder de marquage difficiles à reproduire, comme pour les billets de banque.

Le tatouage de documents numériques s'est développé au début des années 1990 pour la protection du droit d'auteur des documents multimédia. Les mécanismes plus anciens de DRM (Digital Rights Managements), déployés par les professionnels de la distribution de contenus, offrent en effet de nombreux inconvénients. Ils reposent en notamment sur le chiffrement des données durant la transaction, le déchiffrement n'étant possible que lorsque l'utilisateur s'est acquitté des droits d'usage du document. Une licence d'utilisation sert à l'arbitrage : elle spécifie les droits d'usage et délivre la clé de déchiffrement associée. Mais pour qu'un tel système fonctionne correctement, il faut l'intégrer dans une architecture spécifique, suffisamment robuste aux attaques. Cette architecture est extrêmement contraignante, et est souvent propriétaire, empêchant toute interopérabilité. De plus, une fois le document déchiffré, il n'est plus du tout protégé et peut être facilement redistribué de manière illégale par l'utilisateur.

C'est pour contrer ce manque de protection en bout de chaîne que le tatouage a été élaboré. Son rôle initial a été d'offrir — grâce à l'insertion d'une marque dans le document, qui fait corps avec lui et n'en modifie pas l'aspect ou la signification — une protection du contenu une fois que celui-ci est déchiffré. Il offre par ailleurs l'avantage de garantir une bonne interopérabilité, car le document tatoué conserve le format du document original et peut donc être « joué » par les mêmes appareils ou logiciels que celui-ci.

Son rôle est donc aujourd'hui de compléter les outils cryptographiques, qui visent à empêcher techniquement la fraude mais ne le peuvent que dans certaines limites, en offrant un moyen de détecter la fraude *a posteriori*, une fois qu'elle pu avoir lieu. Malheureusement, sa détection est toujours sujette à erreurs, même si elles sont faibles, et il n'est aujourd'hui pas juridiquement reconnu. Il peut cependant constituer un élément dans un dossier judiciaire, et l'accumulation de tels éléments peut être déterminante lors d'une instruction. Et surtout, son utilisation peut dissuader certains utilisateurs malhonnêtes, de peur d'être finalement attrapés.

Le concept de tatouage peut s'appliquer à tout type de données : textes, images fixes, films, fichiers audio, objets 3D, plans et cartes, bandes dessinées, programmes informatiques, bases de données, etc. Néanmoins, le cas le plus largement étudié est celui des images fixes, avec un nombre de publication colossal, puis viennent les cas des films, fichiers audio et objets 3D, les autres n'ayant donné lieu qu'à de très rares publications.

2.1.1 Principes généraux et algorithmiques

Capacité, imperceptibilité, et robustesse

Depuis la « naissance » du tatouage il y a une vingtaine d'années, la qualité des schémas de tatouage est évaluée en prenant en compte, outre la rapidité de traitement, trois critères essentiels à sa mission. Deux d'entre eux, la *capacité* et la *robustesse*, sont liées à la présence implicite d'un canal de communication.

Capacité. La *capacité* d'un schéma de tatouage mesure la quantité d'information que ce dernier est capable de transmettre. Les exigences varient beaucoup d'une application à l'autre : une capacité de 1 bit peut s'avérer suffisante pour de la protection de copie (droit ou non de copier), tandis qu'on demande une capacité d'au moins 64 bits dans le cadre de la protection du droit d'auteur, et jusqu'à plusieurs centaines de bits dans le cadre de la personnalisation de copie.

Imperceptibilité. L'*imperceptibilité* mesure la discrétion de l'insertion. L'exigence est que l'insertion ne doit pas perturber l'utilisation du document, et doit subjectivement être transparente pour l'utilisateur. Pour chaque type de document il convient de définir une mesure adéquate pour estimer la distorsion induite par l'insertion (distorsion visuelle pour les images ou les films, auditive pour les fichiers audio, distorsion comportementale pour les programmes informatiques, distorsion dans les réponses aux requêtes pour les bases de données). Intéressons-nous plus précisément au cas des documents multimédia. Pour estimer le plus fidèlement possible la distorsion induite par l'insertion, on peut s'appuyer sur des modèles psychovisuels ou acoustiques. Mais cela s'avère gourmand en calcul. Une alternative souvent employée est de s'appuyer sur des mesures moins fidèles mais aussi moins coûteuses, comme le traditionnel *PSNR (Peak Signal to Noise Ratio)*, ou l'indice *MSSIM (Mean Structural SIMilarity)*, plus récent [WBSS04].

Robustesse. L'objectif d'un attaquant est avant tout de rendre la marque indétectable, tout en conservant un document de qualité suffisante[2]. La *robustesse* mesure l'aptitude de la marque à persister lorsque le document tatoué subit des transformations génériques. Ces transformations peuvent relever d'intentions malhonnêtes (attaques) ou non (simples manipulations), et être de natures très diverses : filtrage, ajout de bruit, quantification, compression, translation spatiale ou temporelle, suppression de petits morceaux du document, découpage et recomposition du document (mosaïque), insertion d'une nouvelle marque, re-numérisation du document tatoué analogique (scan, photo ou enregistrement au magnétophone ou caméscope selon le type de document), etc.

2. D'autres objectifs ont été abordés dans la littérature, comme le fait d'estimer une marque pour la copier dans un autre document (*copy attack*), ou encore le fait de faire croire qu'un document est tatoué alors qu'il ne l'est pas (*ambiguity attack*). Ces derniers sont marginaux, et ne sont pas abordés ici.

Ces trois critères sont malheureusement en partie incompatibles, et le concepteur d'un schéma de tatouage doit avant tout déterminer en fonction du contexte applicatif quelles sont ses priorités.

Principe de Kerckhoffs et clés

Le tatouage, comme la stéganographie, a hérité de la cryptographie quelques principes de sécurité. Le premier est le *principe de Kerckhoffs*, énoncé par Kerckhoffs en 1883 [Ker83a, Ker83b], qui stipule que la sécurité d'un système ne doit pas reposer sur la tenue secrète de la technique utilisée (algorithme). L'histoire nous a en effet montré qu'il est fort probable que les techniques propriétaires — autrement dit conservées secrètes — se trouvent dévoilées un jour, et que le non respect de ce principe peut mener à la catastrophe. Les procédés d'insertion et d'extraction du message sont donc considérés comme publics. Ils sont en revanche paramétrés par des clés, qui vont notamment déterminer quels éléments ou coefficients du document vont être modifiés pour porter le message, et dans quel ordre ils vont être traités. La plupart des schémas de tatouage sont symétriques (à clé secrète, et utilisent la « même » clé pour insérer et extraire le message. Seules de très rares publications ont étudié la possibilité de construire des schémas asymétriques (à clé publique) [HG97, SD99, VSTS99, FD00a, FMD00, EG00, FD00b, ST01, SHHD01, FVD01, PR01, HJJ02, FD03, CWH04, THC05].

Étant donné que l'attaquant sait que le document contient une marque, et que son insertion a été paramétrée par une clé, le rôle de cette clé est essentiel. Et la difficulté pour l'attaquant à l'estimer l'est tout autant, comme nous le verrons aux paragraphes 2.1.2 et 2.2. Car connaître la clé, ou même seulement 40% de la clé, permet d'effacer la marque, et de pouvoir usurper l'identité d'un ayant droit !

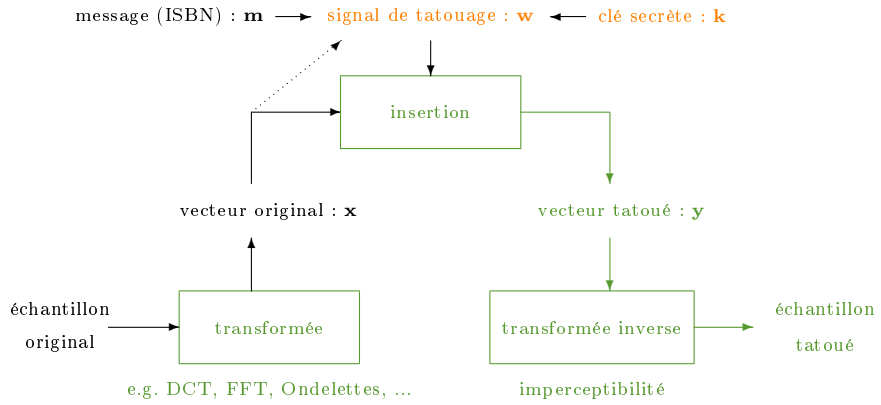
Modèles, principes algorithmiques et notations

Tout schéma de tatouage s'appuie sur une phase d'*insertion* et une phase de *détection* ou *extraction de message*, comme illustré globalement par la figure 1, et plus précisément par la figure 2.1.

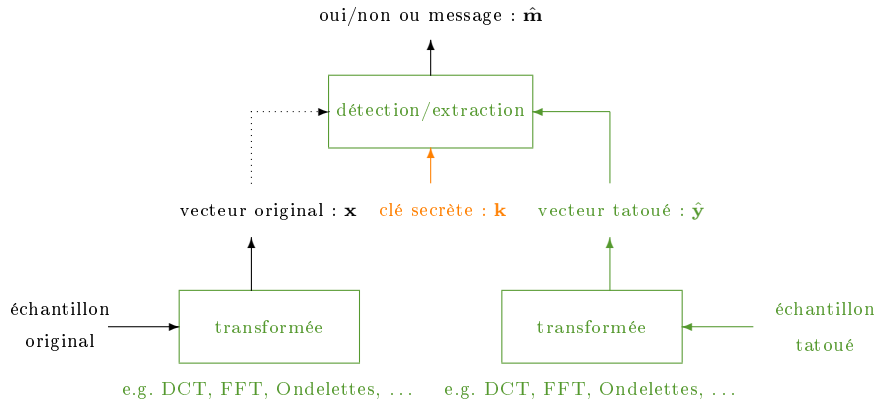
Détection *vs.* extraction. On parle de *détection* lorsque l'on souhaite seulement décider si le document contient ou non une marque correspondant à la clé testée. La technique de tatouage est alors dite *zéro-bit*. La détection s'appuie sur des tests d'hypothèses, et sa fiabilité est estimée à l'aide de probabilités : de détection correcte, de fausse alarme et de non détection.

On parle d'*extraction* lorsque l'on cherche à extraire du document un message à proprement parler. Sa fiabilité est estimée à l'aide du taux d'erreur binaire (Bit Error Rate), qui donne une mesure du nombre de symboles de message erronés à l'extraction.

Bien entendu, cette différence de capacité a un impact fort sur la robustesse, un schéma zéro-bit étant potentiellement plus robuste qu'un schéma de forte capacité. Des codes correcteurs d'erreur peuvent être utilisés pour l'insertion et l'extraction



(a) Insertion



(b) Détection/extraction

FIGURE 2.1: Ces deux schémas modélisent les opérations d'insertion et de détection/extraction. Ils posent par ailleurs les notations utilisées dans ce manuscrit. Les lignes en pointillés montrent les entrées optionnelles. Pour l'insertion, la ligne pointillée correspond à l'utilisation d'une information adjacente. Pour la détection/extraction, elle correspond à la notion de détection/extraction aveugle ou informée. En retrouve en vert les informations connues de l'attaquant, et en orange les informations tenues secrètes. Les données qui peuvent être connues de l'attaquant ou non selon les contextes sont laissées en noir.

de messages. Ils peuvent également servir à améliorer la robustesse du tatouage en introduisant une redondance moins coûteuse qu'une simple répétition.

La détection comme l'extraction s'appuient sur des mesures de biais statistiques (corrélations) liés à l'utilisation de la bonne clé : si la clé utilisée est la bonne, alors un fort biais statistique est mesuré ; si la clé utilisée n'est pas la bonne, alors on ne mesure pas de biais statistique particulier. L'extraction de symboles de message vient en sus.

La détection/extraction est dite *aveugle* lorsqu'elle opère sans la connaissance du

En faisant abstraction du domaine dans lequel l'insertion s'effectue, le document tatoué est obtenu de la manière suivante :

$$\begin{array}{lllll} 1. & \mathbf{y}_i & \leftarrow & \mathbf{x}_i & \forall\ 1 \leq i \leq n \\ 2. & \mathbf{y}_{k_t} & \leftarrow & \mathbf{m}_t & \forall\ 1 \leq t \leq N \end{array}$$

La clé est un vecteur de N entiers. Elle indique quelles positions du document doivent être modifiées pour être remplacées par les symboles du message, lui aussi de longueur N. Ainsi, le t-ième symbole du message est inséré en lieu et place de la $\mathbf{k}_t$-ième composante du document.

Exemple : $n = 8, N = 4$, alphabet binaire :

$\mathbf{m} = (1101)$ $\quad$ $\mathbf{k} = [2, 8, 5, 3]$

$\mathbf{x} = (01001011)$ $\quad$ $\mathbf{y} = (0\mathbf{1}10\mathbf{0}01\mathbf{1})$

FIGURE 2.2: Principe d'une insertion substitutive.

document original, et *informée* si elle s'appuie sur la connaissance de ce document. Bien entendu, la connaissance du document original la rend plus fiable. Dans certains contextes, comme le traçage de documents diffusés dans le cadre d'un schéma de VOD (vidéo à la demande), le distributeur connaît les documents originaux, et peut les utiliser lors de la détection/extraction. Dans d'autres scénarios, cela n'est pas possible.

Insertion substitutive *vs.* additive. Algorithmiquement parlant, on distingue deux types d'insertion. Les méthodes *substitutives* remplacent *in situ* certaines composantes (*e.g.* valeurs de luminance, coefficients DCT, de Fourier ou ondelettes pour des images) du document original $\mathbf{x}$ par d'autres valeurs, suivant le message $\mathbf{m}$ que l'on souhaite cacher et la clé $\mathbf{k}$. Parmi les méthodes de tatouage d'image substitutives, la plus célèbre est celle de Koch et Zhao [ZK95,KZ95], qui a ensuite été adaptée à la vidéo. Cette méthode a pour principe de considérer certains blocs de pixels de l'image, déterminés par la clé utilisée ; chacun de ces blocs va porter un bit de message : pour un bloc donné, le bit correspondant du message est inséré en forçant une relation d'ordre entre deux coefficients DCT. La figure 2.2 donne une formalisation de ce type d'insertion, qui fait abstraction du domaine dans lequel elle s'effectue.

La plupart des méthodes de tatouage sont *additives* : à l'aide de la clé $\mathbf{k}$ et du message $\mathbf{m}$, elles élaborent une marque $\mathbf{w} = \mathrm{Gen}(\mathbf{m}, \mathbf{k})$ pour finalement l'ajouter au document original $\mathbf{x}$, produisant ainsi le document tatoué $\mathbf{y} = \mathbf{x} + \mathbf{w}$. La génération de la marque $\mathbf{w}$ fait par ailleurs intervenir un coefficient pondérateur ou *masque* qui permet de modérer l'insertion en pour garantir l'imperceptibilité. Les approches les plus populaires pour la conception de tels schémas s'appuient sur des techniques de modulation avec étalement de spectre *e.g.* [CKLS97, DDGM97, OP98, DDVM98, PL03, MF04] ou quantification avec dictionnaires (Quantized Index Modulation), *e.g.* [CW99a, MCB00, CW01, EBTG03, MDC04]. Une comparaison de l'efficacité de ces deux approches a été publiée en 2005 [PFPG05]. La figure 2.3 propose à titre

À titre d'exemple, et en faisant abstraction du domaine dans lequel l'insertion s'effectue, la marque $\mathbf{w}$ générée dans le cas d'un schéma de base s'appuyant sur une modulation avec étalement de spectre pour insérer un message binaire $\mathbf{m}$ est de la forme :

$$\mathbf{w} = \frac{\gamma}{\sqrt{N}} \sum_{t=1}^{N} (-1)^{\mathbf{m}_t} \mathbf{u}_t$$

avec γ un coefficient pondérateur, et les vecteurs $\mathbf{u}_t \in \{-1, +1\}^n$ les N porteuses secrètes (autrement dit la clé secrète du schéma ; ces porteuses sont de norme égale à 1 et deux-à-deux orthogonales). On a ainsi :

$$\mathbf{y} = \mathbf{x} + \frac{\gamma}{\sqrt{N}} \sum_{t=1}^{N} (-1)^{\mathbf{m}_t} \mathbf{u}_t$$

FIGURE 2.3: Principe d'une insertion additive avec étalement de spectre.

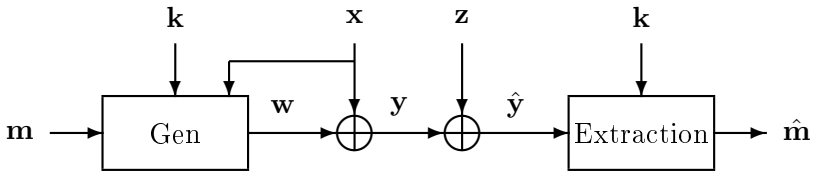


FIGURE 2.4: Le tatouage perçu comme une transmission avec information adjacente à l'émission. Deux bruits interviennent successivement pour perturber le signal transmis (le message) : tout d'abord le document original $\mathbf{x}$ (qui va forcer à modérer l'insertion par souci d'imperceptibilité), puis le bruit du canal de transmission $\mathbf{z}$ correspondant aux transformations et attaques que le document tatoué va subir.

d'exemple une formalisation d'une insertion de base par étalement de spectre. Les techniques d'insertion additives ont tiré parti dès le début des années 2000 des travaux menés dans un autre cadre par Costa [Cos83], et qui permettent d'exploiter au mieux la connaissance que l'on a du document original au moment de la génération de la marque que l'on va y ajouter. Comme remarqué en 1999 [CW99b, CMM99], et comme l'illustre la figure 2.4, le tatouage peut être modélisé comme une transmission avec *information adjacente*, telle que définie par Shannon en 1958 [Sha58] : le document original apparaît alors comme un bruit qui perturbe la marque, porteuse du message à transmettre, mais bruit que l'émetteur connaît. Costa a montré que si ce bruit, connu de l'émetteur mais pas du récepteur, est i.i.d. gaussien et que le bruit de canal est un bruit blanc gaussien, alors la capacité du canal global est la même que si ce bruit n'existait pas. Il a par ailleurs donné une construction effective qui se transpose très bien en tatouage et permet de construire la marque $\mathbf{w} = \mathrm{Gen}(\mathbf{m}, \mathbf{k}, \mathbf{x})$ de manière (sub)optimale en termes de capacité et robustesse.

Quelle que soit la technique d'insertion choisie, elle est ensuite complétée par

des procédés visant à en améliorer la robustesse, l'imperceptibilité ou la capacité si nécessaire, selon le contexte applicatif : codes correcteurs, tatouage dans un espace transformé invariant par certaines manipulations du document, insertion de motifs de resynchronisation, exploitation de points d'intérêt, etc.

Une vision géométrique. L'insertion peut être représentée et interprétée géométriquement. Prenons comme espace ambiant l'ensemble de tous les documents possibles, *e.g.* toutes les images fixes, muni d'une distance relative à la distorsion appropriée (visuelle pour des images, auditives pour des signaux audio, etc). L'insertion revient à projeter le document dans un espace secret, déterminé par la clé d'insertion. Pour une clé donnée, l'espace secret associé correspond à l'ensemble des documents considérés comme tatoués, le reste de l'espace correspondant aux documents considérés comme non tatoués. Selon le type d'insertion et le type de corrélation utilisée à la détection/extraction, l'espace secret peut s'avérer être un hyperplan, un hypercône, etc. Tant que le document tatoué restera dans cet espace, il sera considéré comme tatoué. S'il en sort, pour une raison ou une autre, il sera considéré comme non tatoué. Pour être la plus robuste possible, il faut donc que l'insertion l'entraîne le plus loin possible du « bord » de cet espace. Mais elle doit également ne pas l'entraîner trop loin du document original, pour ne pas introduire une trop forte distorsion.

Études théoriques et optimisation. Le tatouage relève majoritairement d'une approche et d'un savoir-faire empirique. Certains travaux ont néanmoins visé à en dresser les limites, et à étudier par exemple la capacité maximale des schémas, à formaliser des critères de *sécurité*, etc. Ces travaux s'appuient sur les outils issus de la théorie de l'information, et suivent les jalons posés par les études antérieurs liées aux problèmes de transmission en général.

Mais modéliser le canal de transmission s'avère très difficile, chaque manipulation du document tatoué ayant des conséquences très différentes. Quelques travaux ont néanmoins tenté de modéliser le canal [Mit99, MO03, SBM04, YJXPLML08] pour en dériver des bornes sur la capacité et la robustesse maximales des schémas [ZZ04, YJXPLML05, AKS06, YJXPLML08]. La modélisation du tatouage comme un jeu entre l'émetteur qui insère la marque et l'attaquant qui tente de la retirer a également permis de mieux comprendre le problème, et de guider certains choix de paramètres [MI03, PL03].

Comme mentionné plus haut, les travaux de Costa [Cos83] ont eux aussi cherché à optimiser la capacité, sous des hypothèses différentes. Cette étude, théorique mais aussi constructive, a eu un impact en tatouage très important.

Les travaux traitant de *sécurité* sont décrits dans le paragraphe suivant.

2.1.2 Sécurité

Aux trois critères historiques que sont la capacité, l'imperceptibilité ou la robustesse s'en ajoute depuis une dizaine d'années un quatrième, la *sécurité*. Car robustesse et sécurité sont bien distinctes, comme nous le discuterons ici et dans la section 2.2. Avant tout, il convient de bien garder à l'esprit que l'objectif de l'attaquant est de rendre la marque indétectable (voire de la modifier à son gré), tout en conservant un document de qualité suffisante. Pour ce faire, il peut agir à l'aveuglette, en appliquant au document tatoué un ensemble de transformations génériques : compressions, filtrages, ajout de bruit, etc, comme évoqué plus haut dans le paragraphe dédié à la robustesse. Contrer ce type d'attaque relève de la robustesse. Mais il peut aussi agir plus méthodiquement, tirant partie de toutes les informations qu'il peut collecter : ensemble de documents tatoués avec la même clé, connaissance de la technique d'insertion, etc. Son objectif peut alors aller jusqu'à estimer la marque, l'espace de tatouage, ou même la clé qui a servi lors de l'insertion. Plus son estimation sera précise, plus son attaque sera efficace. Contrer ce type d'attaque relève de la *sécurité*.

Cette distinction entre robustesse et sécurité est apparue peu à peu dans la littérature [Kal01,FR02,BBF03]. Elle semble aujourd'hui évidente, et on peut se demander pourquoi elle n'a pas été envisagée plus tôt. Tout simplement parce qu'il ne servait à rien de se poser la question tant que les trois critères de base n'étaient pas suffisamment bien gérés, en particulier la robustesse. Mais aujourd'hui, les techniques d'insertion offrent une robustesse tout-à-fait satisfaisante, et il convient d'étudier ce que l'attaquant peut effectuer comme attaque pour malgré tout retirer la marque.

L'attaquant peut par exemple, s'il a la possibilité de soumettre des documents à un détecteur/extracteur utilisable en boîte noire, effectuer des attaques par oracle appelées *sensitivity attacks*, *e.g.* [CL97, KLvD98, LvD98, CPFPG05, VJ05, Wes06, CPG07b, ECM07b, ECM07a, ECM09b, ECM09a, Wes09]. Leur principe est, si l'on adopte la vision géométrique, de trouver le chemin le plus court pour sortir le document tatoué de l'espace secret et obtenir ainsi un document considéré comme non tatoué, mais qui présente le moins de distorsion possible par rapport au document tatoué donné en entrée.

L'attaquant peut aussi, puisqu'il connaît la technique d'insertion, l'analyser pour y chercher des failles. Une série d'articles a ainsi tiré parti, pour plusieurs techniques de tatouage très spécifiques, de la connaissance et de l'analyse de l'algorithme d'insertion pour élaborer une attaque dédiée, qui permet de retirer la marque à l'aide d'une unique image tatouée, sans estimer ni la marque, ni l'espace de tatouage, ni la clé (ce qui n'était manifestement pas nécessaire, les techniques souffrant de faiblesses structurelles.) [DM02, DMM05, DM04, DM06, DKM05, DMZ06]. Le même genre de faiblesse a permis dans [DZ05] de retrouver la clé de tatouage à partir d'une seule image tatouée, et dans [DZM05] de retrouver une partie de l'espace secret d'insertion, suffisante pour retirer la marque.

Il existe cependant de nombreux algorithmes d'insertion qui ne présentent pas *a priori* de failles majeures. Mais ils ne sont pas forcément pour autant à l'abri d'attaques dédiées. En effet, l'attaquant peut alors, s'il dispose d'un ensemble de documents tatoués tous avec la même clé, essayer d'estimer l'espace secret d'insertion,

la marque, ou même la clé. Cette opération sera facilitée s'il connaît l'algorithme d'insertion. Cette approche a été formalisée dans les travaux que j'ai menés avec Teddy Furon et François Cayre de 2002 à 2004 [CFF05b, CFF05c, CFF05d, CFF05a, CFF05e, CFF07], et qui sont résumés dans le paragraphe 2.2. Nous avons mené une étude théorique de la sécurité de l'ensemble des schémas de tatouage substitutifs, et de l'ensemble des schémas additifs reposant sur de l'étalement de spectre. Nous avons pour cela suivi la méthodologie introduite par Shannon en 1949 pour l'étude de la sécurité des schémas de chiffrement symétriques [Sha49] : l'attaquant observe tout d'abord un ensemble de documents tatoués avec la même clé ; de ces observations, il tire des informations sur la clé en question. Si aucune information sur la clé ne fuit, alors la sécurité est parfaite. Si au contraire on constate une fuite d'information, alors il convient d'estimer à partir de combien de documents observés cette fuite est dangereuse. Ces outils nous ont permis d'étudier la sécurité des schémas de tatouage substitutifs, mais nous ont parus insuffisants pour étudier la sécurité des schémas reposant sur de l'étalement de spectre. Pour ces derniers, nous avons utilisé des outils issus de la théorie de Fisher, qui permettent d'estimer le paramètre d'un système (ici la clé). Cette étude théorique a été complétée par une étude opérationnelle des techniques algorithmiques permettant concrètement de retrouver tout ou partie de la clé pour les cas où une fuite d'information théorique avait été constatée. Cette étude théorique et pratique a été menée dans trois contextes d'attaques, que nous avons calqués sur les contextes d'attaque usuels en cryptographie : *WOA (Watermark Only Attack)* lorsque l'attaquant n'a accès qu'à des documents tatoués, *KMA (Known Message Attack)* lorsque l'attaquant a en sus accès aux messages insérés, et *KOA (Known Original Attack)* lorsque l'attaquant a accès en sus des documents tatoués aux documents originaux correspondants. Une étude algorithmique indépendante a aussi abordé la question de l'intérêt pour l'attaquant d'observer une série de documents tatoués tous avec la même clé, dans le cas de la vidéo [DD04a, DD04b], cherchant à estimer la marque, mais pas à retrouver la clé d'insertion. Notre méthodologie a été reprise et étendue dans une série d'études qui ont suivi, et qui ont étudié à leur tour d'autres techniques d'insertion [CPG05, BH05, PFCPG05, PFPGFC06, PFCTPPG06, PFPG07, BD07, PFPG08, CB08, PFPG09, BW09, GFB10, BCG11, BG10, FGB11, BCG12].

Parmi les travaux les plus récents, P. Bas et F. Cayre ont développé dans [CB08] une analyse et une formalisation de la sécurité unifiée pour l'ensemble des techniques de dissimulation d'information, couvrant ainsi tous les scénarios de stéganographie (prenant en compte le modèle de gardien passif mais aussi celui du gardien actif), tatouage, et par extension *fingerprinting*. Cette analyse propose quatre niveaux de sécurité : du plus fort au plus faible la *stego-security* (impossibilité de distinguer les documents porteurs d'information de ceux qui n'en contiennent pas), la *subspace-security* (impossibilité de distinguer l'espace secret d'insertion, toutes les clés donnant lieu à une distribution de données correspondant aux observations de l'attaquant), la *key-security* (certaines clés d'insertion correspondent à la distribution des données observées par l'attaquant, et peut donc être privilégiées dans l'attaque, tandis que d'autres clés correspondent à d'autres distributions), et l'*insecurity* (une seule clé correspond aux données observées ; elle peut donc être isolée).

Algorithmiquement parlant, les attaques qui tirent partie de multiples observations pour estimer la clé ou l'espage secret d'insertion s'appuient par exemple sur : une analyse en composantes principales (PCA) ou analyse en composantes indépendantes (ICA) pour retrouver une base de l'espace secret [CFF05e, DD04b, BW09], des techniques de classification (*clustering*) [DD04b, BW09], des techniques d'estimation à maximum de vraissemblance (MLE) [CFF05e], ou encore des techniques de maximisation d'espérance (EM) [DD04b].

Pour stimuler la recherche dans le domaine, deux concours internationaux ont été organisés. Le premier, BOWS-1 (*Break Our Watermarking System*), s'est déroulé en 2005-2006 [PB07] [3]. Le deuxième, BOWS-2, s'est déroulé en 2007-2008 [4]. Constitué de trois épisodes dédiés successivement à des attaques de type robustesse, à des attaques de type *sensitivity attack*, puis à des attaques exploitant la connaissance de l'algorithme et l'observation d'un grand nombre de documents tatoués, BOWS-2 a mis à l'épreuve une technique de tatouage zéro-bit très robuste utilisant un détecteur aléatoire, et conçue spécialement pour le concours, dénommée `Broken Arrows` [FB08]. Ces deux concours ont effectivement stimulé la communauté, et donné lieu à la publication de nouvelles attaques, *e.g.* [Wes06, CPG07b, BW09, Wes09]. Ils ont également illustré à quel point l'accès à un oracle de détection/extraction, la connaissance de la technique d'insertion, et l'accès à un grand nombre de documents tatoués tous avec la même clé peuvent aider l'attaquant.

On voit donc aujourd'hui comment aborder l'étude *a posteriori* de la sécurité des schémas. Mais un point reste globalement ouvert : comment garantir dès la conception qu'un schéma atteindra tel ou tel niveau de sécurité ? En ce qui concerne les *sensitivity attacks*, une solution proposée pour contrer ce type d'attaque est d'utiliser un détecteur aléatoire (*randomized detector*) approprié [LvD98, CPFPG05, VJ05, ECM09b], mais le choix du détecteur s'avère délicat [ECM09b, Wes09]. Pour ce qui est des niveaux de sécurité mentionnés plus haut (*key-security*, *subspace-security*, etc), seuls quelques rares articles (comparativement au grand nombre d'articles publiés sur la dissimulation d'information) se sont penchés sur la question [BC06a, BC06b, CB08, MCB07, MBCM09a, MBCPG08, MBCM09b, XFF10b, GFB10, BG10, BCG11, FGB11, BCG12], et il est apparu qu'apporter de la sécurité nécessite souvent de baisser le niveau de robustesse, sans que l'on comprenne bien pourquoi et surtout dans quelle mesure. Le paragraphe 2.4 présente les travaux que j'ai menés avec Teddy Furon et Fuchun Xie sur la sécurisation de `Broken Arrows` et les questions que cette sécurisation a soulevées. Citons enfin pour terminer [MBCM10b, MBCM10a], qui se sont penchés sur les contraintes de robustesse et de sécurité imposées par l'utilisation du tatouage comme support pour un code anti-collusion comme celui de Tardos (voir chapitre suivant pour la définition de ces codes).

Liens entre tatouage et cryptologie

Cryptologie et tatouage sont complémentaires par nature, et s'enrichissent mutuellement. La cryptographie offre toute une panoplie de protocoles qui permettent

3. `http://lci.det.unifi.it/BOWS/`
4. `http://bows2.ec-lille.fr/`

d'utiliser les schémas de tatouage dans des architectures réelles, du simple protocole d'échange de clés aux protocoles dits *Buyer-Seller* ou *asymétriques*, qui permettent de s'assurer qu'aucune des parties ne triche, voire qui assurent l'anonymat des utilisateurs. Ces derniers ont principalement été développés dans le contexte du *fingerprinting*, et sont présentés dans le paragraphe 3.4 du chapitre suivant.

Mais au-delà de cette simple complémentarité, où chaque technique se cotoie, les liens entre tatouage et cryptologie sont plus profonds : utilisation de clés, objectif d'authentification (du propriétaire ou de l'utilisateur) pour le tatouage robuste, ou parfois d'intégrité pour le tatouage (semi-)fragile. Ces similitudes, ainsi que les apports mutuels de ces deux domaines ont par exemple été discutés dans [FR02]. Ainsi, on a vu des algorithmes de tatouage utiliser des techniques d'authentification *zero-knowledge* issues de la cryptographie, *e.g.* [Cra99,PL05,YL05,ARS06,TPPG06], et des méthodologies d'analyse se transposer parfaitement au tatouage, comme par exemple dans [CFF05e], dont les résultats ont été mentionnés plus haut et sont présentés ci-dessous à la section 2.2. Mais il faut prendre garde à ne pas se laisser tenter par toutes les analogies trop rapidement, car elles sont parfois trompeuses. Cox *et al.* ont par exemple mentionné dans [CDF06] quelques pièges à éviter. Ils soulignent par exemple le fait qu'en tatouage robuste avec étalement de spectre, utiliser une clé secrète binaire de n bits ne signifie en aucun cas qu'il y a 2^n clés possibles, car cette clé est constituée de porteuses 2-à-2 orthogonales, ce qui réduit l'espace des clés à $2^{n-0.5\log_2 n}$. Par ailleurs, l'étude que nous avons mené avec T. Furon et F. Cayre et qui est résumée ci-après, a montré qu'avec une estimation de seulement 40% de la clé on peut retirer la marque très proprement, sans dégrader le document. Ainsi, le défi que doit relever l'attaquant face à une clé secrète de n bits n'a ici rien à voir avec celui qu'il devrait relever face à une clé secrète de n bits d'un algorithme de chiffrement (avec un espace de clés de 2^n et la nécessité de retrouver toute la clé pour pouvoir parvenir à ses fins). P. Bas, T. Furon et F. Cayre poursuivent actuellement dans cette voix, cherchant des estimations plus précises de la taille de l'espace des clés en tatouage, mettant en exergue l'équivalence de certains ensembles de clés, équivalences que l'attaquant peut utiliser pour accéder au canal de transmission. Ces travaux devraient être publiés en 2012.

2.2 Cryptanalyse des schémas de tatouage (2002-2005)

Les premiers travaux que j'avais menés en 2001 avec Frédéric Raynal sur les liens entre tatouage et cryptographie [FR02] m'avaient fortement poussée à étudier la possibilité de mener des attaques structurelles sur les algorithmes de tatouage. Je parle ici d'attaques visant non pas à lessiver le filigrane (comme le ferait une compression, une conversion numérique-analogique-numérique par exemple), mais à retrouver la clé secrète utilisée lors de son insertion. En effet, si l'attaquant réussit à estimer cette clé, alors il sera en mesure d'effacer proprement le filigrane (sans dégrader le document, et sans laisser de trace de son action), ou de modifier le message qu'il contient (accès au canal caché). Si cette approche avait été évoquée dans [FR02,Kal01,BBF03], personne n'avait à l'époque encore abordé techniquement

cette question. Teddy Furon et François Cayre étant eux aussi intéressés par cette approche, il nous a semblé opportun de conjuguer nos efforts, et surtout de profiter de nos profils scientifiques complémentaires pour aborder cette question. Notre ambition était d'étudier la cryptanalyse des schémas les plus populaires de tatouage d'images fixes : schémas substitutifs classiques (de la famille de [ZK95, KZ95]), et schémas reposant sur l'étalement de spectre.

Tout d'abord, conformément au pricipe de Kerckhoffs, nous avons supposé que l'attaquant sait quel système de tatouage a été utilisé. Nous avons ensuite suivi la même démarche que C. Shannon lorsqu'il avait modélisé la sécurité des systèmes de chiffrement dans [Sha49] : nous nous sommes placés dans le cas où l'attaquant observe N_o documents qui ont tous été tatoués avec la même clé, mais qui peuvent contenir des messages différents ; nous avons alors étudié d'une part la quantité d'information sur la clé qui pouvait potentiellement fuire de ces observations (au sens de la théorie de l'information), et d'autre part quels moyens concrets pouvaient permettre au pirate de récupérer cette information et de reconstituer la clé. Conformément aux études cryptanalytiques, nous avons considéré différents contextes : 1) WOA l'attaquant n'a accès qu'aux documents tatoués, 2) KOA il a accès aux documents tatoués et aux documents originaux correspondants, et 3) KMA il a accès aux document tatoués et aux messages qui y sont cachés. Il nous a fallu une bonne année pour résoudre la question théorique ; les outils proposés par Shannon, comme l'*entropie* ou l'*equivocation*, se sont avérés adaptés à l'étude des schémas substitutifs mais pas à celui des schémas reposant sur l'étalement de spectre, pour lesquels nous avons utilisé les outils proposés par Fisher, comme la *matrice d'information de Fisher* et le *théorème the Cramer-Roa*, qui permettent d'estimer le paramètre d'un système (ici la clé), et fournissent une borne sur la fiabilité de cette estimation.

Nous avons alors pu démontrer que seul le schéma substitutif présente une *couverture parfaite* (*i.e.* ne laissait filtrer aucune information sur la clé, par analogie avec la notion de *perfect secrecy* de Shannon), et seulement dans le cas WOA. Dans tous les autres cas, il y a une fuite théorique d'information, que l'attaquant peut essayer d'exploiter. Voici le nombre d'observations à partir duquel l'attaquant a, en théorie, suffisamment d'information pour reconstituer la clé. Pour le schéma substitutif, dans le cas KOA il est de $\log_2 N$ avec N la taille de la clé et du message, et dans le cas KMA de $\log_2 n$ avec n la longueur du vecteur extrait. La figure 2.5 illustre ces résultats. Pour les schémas par étalement de spectre, dans les cas WOA et KMA il suffit que l'attaquant puisse réaliser $\mathcal{O}(N\sigma_{\mathbf{x}}^2/\gamma^2)$ observations, avec N le nombre de porteuses, $\sigma_{\mathbf{x}}^2$ la variance du signal hôte et γ la force d'insertion. Dans le cas KOA $\mathcal{O}(N)$ observations suffisent. Nous avons également démontré que pour les deux types d'algorithmes la clé ne peut être complètement retrouvée que lorsque des messages sont connus (KMA). Dans les autres cas (WOA et KOA pour les techniques additives par étalement de spectre, et KOA pour les techniques substitutives), on peut connaître toutes ses composantes, mais à l'ordre près (et au signe près pour les techniques additives utilisant de l'étalement de spectre, à l'ordre près pour les techniques substitutives). L'attaquant peut alors sans difficulté altérer un message, mais pas forcément inscrire celui de son choix.

Enfin, nous avons élaboré des algorithmes permettant à l'attaquant de mener effi-

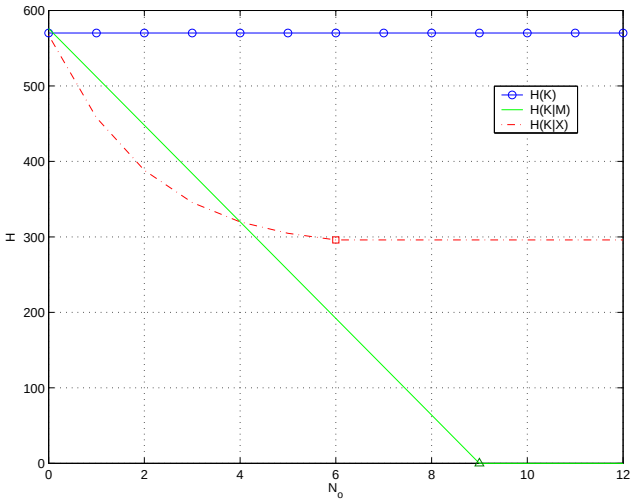


FIGURE 2.5: Évolution de l'*equivocation*, qui mesure l'incertitude sur la clé, en fonction du nombre d'observations, pour des schémas de tatouages substitutifs. La courbe $H(K)$ correspond au cas WOA, la courbe $H(K|M)$ au cas KMA, et la courbe $H(K|X)$ au cas KOA.

cacement la cryptanalyse. Dans le cas des schémas substitutifs, les algorithmes étaient très naturels, procédant par comparaison composante-à-composante des messages ou des documents originaux connus pour en déduire où la clé a agi. Le cas des schémas par étalement de spectre est nettement plus complexe. Le contexte le plus simple est KMA, pour lequel une estimation à maximum de vraisemblance (MLE) fonctionne très bien. Pour les deux autres contextes, KOA et WOA, nous nous sommes appuyés sur les techniques de séparation de sources, sans bruit pour KOA, avec bruit pour WOA. Vient ensuite le cas KOA, utilisant l'analyse en composantes principales (PCA) et l'analyse en composantes indépendantes (ICA). Le cas KOA est relativement aisé à gérer avec les techniques connues d'ICA. Le cas WOA est en revanche beaucoup plus délicat. Deux astuces nous ont néanmoins permis d'obtenir une attaque opérationnelle : couper les vecteurs en sous-vecteurs plus courts plus faciles à manipuler par ICA, et utiliser itérativement une estimation à maximum de vraisemblance (MLE) reposant sur des estimations succesives de porteuses et de messages. Nous avons mené des expérimentations sur des signaux synthétiques et sur de vraies images, avec différentes techniques d'insertion par étalement de spectre, dont des techniques avec information adjacente. Nos résultats, non spécifiques à l'image, ont démontré l'efficacité de cette approche, qui a depuis été suivie dans la communauté. Ils sont illustrés par les figures 2.6 et 2.7. Ils ont également montré à quel point il peut s'avérer dangereux d'utiliser trop souvent la même clé. Il faut donc en changer relativement souvent, même si cela induit un coût de gestion supplémentaire.

La fin de ce travail a été soutenue par le réseau d'excellence européen ECRYPT, du 6ème PCRD. Ce travail a été présenté à la communauté internationale de trai-

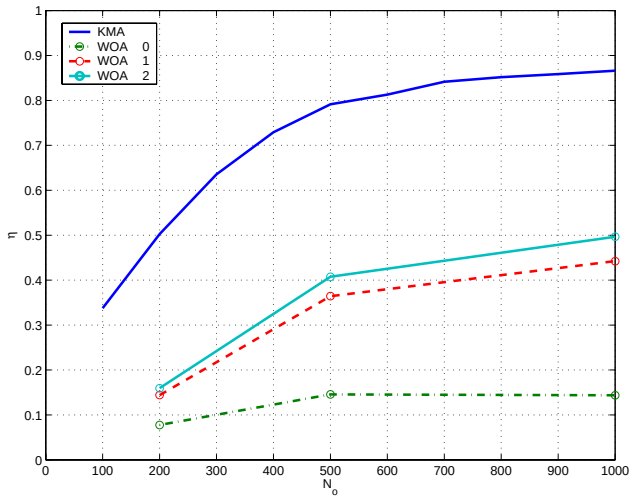


FIGURE 2.6: Efficacité de l'estimation des porteuses pour les techniques d'insertion par étalement de spectre : corrélation entre les porteuses estimées et les vraies, en faisant abstraction de l'incertitude sur l'ordre des porteuses (on les met dans le bon ordre pour calculer les corrélations). On remarque que dans le cas WOA il ne sert à rien de procéder à un grand nombre d'itérations pour le calcul des porteuses. Avec 3 itérations et 500 images, on estime correctement 40% des porteuses. La figure 2.7 montre à quel point cette estimation est suffisante pour retirer la marque tout en conservant une excellente qualité à l'image.

tement du signal lors des conférences *IWDW : International Workshop on Digital Watermarking* en 2004 [CFF05b] – où il a reçu le *Best Paper Award* – et *IS&T/SPIE International Symposium on Electronic Imaging'2005* [CFF05c, CFF05d] en janvier 2005 ; l'analyse théorique a été plus spécifiquement présentée à la communauté internationale de théorie de l'information lors du *IEEE-International Symposium on Information Theory'2005* [CFF05a]. L'ensemble de ce travail fait également l'objet d'un article dans la revue *IEEE Transactions on Signal Processing* pour un numéro spécial *Supplement on Secure Media* [CFF05e] [5]. Un chapitre de livre a également été publié [CFF07].

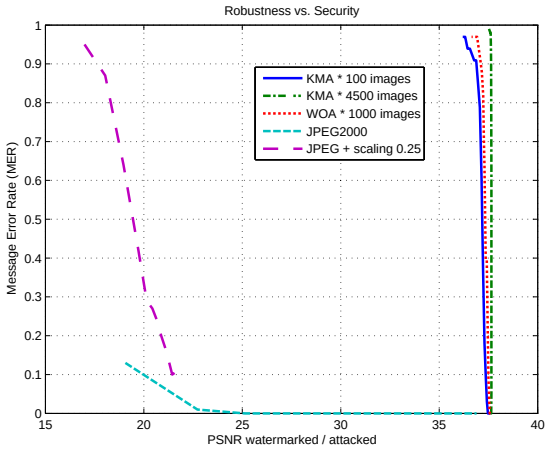
2.3 Utilisation de tatouage audio robuste pour sécuriser la diffusion de musique sur les téléphones mobiles 3G (2003-2006)

Dans le cadre du projet RNRT SDMO (Sécurisation de la Diffusion de Musique sur les mObiles), qui s'est déroulé de 2003 à 2006, nous avons étudié la possibilité de sécuriser la diffusion de morceaux de musique sur les téléphones mobiles 3G. J'ai

5. Article joint en annexe D.

(a) Meilleure qualité obtenue pour une attaque à l'aveuglette (robustesse) : 21.8 dB

(b) Meilleure qualité obtenue pour une attaque par estimation de la clé (sécurité) : 35.8 dB



(c) Qualité (PSNR) des images attaquées avec succès par des attaques à l'aveuglette, et par une estimation de la clé, pour différents contextes.

FIGURE 2.7: Expérimentations effectuées sur des images de taille 512×512, tatouées avec une technique de tatouage par étalement de spectre avec information adjacente, réputée pour sa robustesse [PL03]. Les paramètres sont $n = 258058$ et $N = 8$. Comparaison de l'efficacité de deux stratégies d'attaque : (a) l'attaquant procède traditionnellement, à l'aveuglette ; (b) il estime la clé par notre méthode, en observant environ 1000 images tatouées avec la même clé. L'image « Lena » tatouée présentait un PSNR de 38 dB. (c) Nous avons mené ces tests sur 50 images de taille 512×512, et le PSNR de l'image obtenue avec l'estimation de la clé est en moyenne de 15 dB supérieure à celle de l'image obtenue avec une attaque à l'aveuglette.

été responsable du sous-projet *sécurité*, dont le rôle était de définir l'architecture de sécurité, ainsi que de fournir une analyse complète de la sécurité finale du projet, tant au niveau cryptographique qu'en ce qui concerne le tatouage. J'ai également participé au sous-projet *architecture* qui a défini l'architecture globale du système, ainsi qu'au sous-projet *tatouage* qui a élaboré les algorithmes de tatouage adéquats.

Dans le sous-projet *sécurité*, nous avons défini l'ensemble de l'architecture et des protocoles permettant d'assurer la meilleure sécurité possible, compte tenu des contraintes de temps lors des échanges entre les entités (serveur, carte SIM dans le téléphone, ...), et des capacités de traitement de ces entités. Nous avons ainsi pu mettre en place une architecture originale, reposant sur l'utilisation de la carte SIM pour la gestion des licences, des échanges sécurisés (confidentialité et authentification assurées grace à des protocoles dédiés, tout en conservant une compatibilité avec les normes en cours d'élaboration par l'OMA – Open Mobile Alliance) entre la carte, le logiciel qui gèrera les morceaux et le matériel qui les jouera réellement ; nous avons ainsi décidé d'insérer deux types de tatouages : le premier permet de cacher un identifiant long, qui sera comparé dans le terminal (téléphone) à celui qui est extrait de la licence, assurant un lien non falsifiable entre le fichier audio lui-même (dans son format d'origine ou un autre, ayant éventuellement subi des tranformations) ; le deuxième ne permet qu'une détection binaire et permettra de détecter les fichiers frauduleux qui circulent (indifféremment au niveau des terminaux ou sur le réseau) ; cette détection est la garantie que le fichier a été originellement protégé par SDMO. En parallèle, j'ai travaillé dans le cadre du sous-projet *tatouage* aux réflexion menées sur la pertinence ou non d'utiliser des procédés stéganographiques dans SDMO, et enfin à l'analyse globale de la sécurité du système, prenant en compte à la fois les aspects de tatouage et cryptographiques. Notons que les travaux de cryptanalyse exposés à la section suivante peuvent s'adapter aux schémas de tatouage audio par étalement de spectre et sont donc alimenté notre analyse de la sécurité du système.

Ce projet a donné lieu à la réalisation d'un prototype et une démonstration en a été faite au Ministère de l'Industrie en mai 2006. Ces travaux ont été présentés lors de salons industriels, et de conférences académiques [ABTD+06, BTFF+06], et un brevet a été déposé sur le procédé liant le fichier audio à sa licence [GFF+06]. Un article a été publié en 2008 dans la revue *ISAST Transactions on Communication and Networking* [FDD+08].

2.4 Amélioration de la robustesse et de la sécurité de Broken Arrows : thèse de F. Xie (2007-2010)

Comme mentionné au paragraphe 2.1.2, différentes études ont été menées depuis l'introduction de la notion de sécurité pour estimer *a posteriori* la sécurité des techniques de tatouage, notamment encouragées par les deux concours internationaux BOWS-1 et BOWS-2 ; mais seuls de rares articles se sont penchés sur la question de la sécurité dès la conception du système. Ils ont néanmoins constaté qu'apporter de la sécurité nécessite souvent de baisser le niveau de robustesse, sans que l'on maîtrise encore dans quelle mesure ni pourquoi.

Avec Fuchun Xie et Teddy Furon, nous nous sommes focalisés sur la technique de tatouage zéro-bit robuste `Broken Arrows` [FB08], qui avait été développée pour être mise à l'épreuve lors du concours BOWS-2. Elle a donc été conçue avec un fort souci de robustesse. Le concours était organisé en plusieurs étapes, chacune permettant de mener un type d'attaque. Lors de la première phase du concours BOWS-2, dédiée aux attaques de type robustesse, A. Westfeld a mis à jour une faiblesse de `Broken Arrows` [Wes08]. Lors de la troisième phase de BOWS-2, dédiée aux attaques de type sécurité, les participants avaient une connaissance parfaite du système, et pouvaient observer beaucoup d'images tatouées avec la même clé secrète. A. Westfeld a alors conçu une nouvelle attaque, plus efficace mais qui s'appuie néanmoins sur la première. P. Bas, un des concepteurs de `Broken Arrows`, a de son côté réfléchi aux éventuelles failles de sécurité du système. Ils ont publié l'ensemble de leurs conclusions en 2009 [BW09].

Nous avons dans un premier temps cherché à renforcer `Broken Arrows` pour le rendre invulnérable à la première attaque de Westfeld. Dans un deuxième temps, nous avons cherché à améliorer sa sécurité pour contrer les attaques publiées par P. Bas et A. Westfeld dans [BW09].

Amélioration de la robustesse de `Broken Arrows`. L'attaque menée par A. Westfeld lors de la première phase du concours BOWS-2, dédiée aux attaques de type robustesse, peut être considérée comme un processus de débruitage. Elle repose sur l'estimation de l'amplitude de chaque coefficient de la transformée en ondelettes, estimation réalisée *via* une régression linéaire portant sur les coefficients situés dans son voisinage [Wes08].

Il est nécessaire pour présenter nos résultats de décrire certains aspects de `Broken Arrows`. Globalement, l'insertion se déroule dans le domaine des ondelettes. Elle s'appuie sur une modulation et une projection dans un plan dit MCB, paramétrées par la clé. La modulation implique N_v porteuses secrètes, tandis que le sous-espace secret final est constitué de N_c hypercônes. La détection s'appuie un détecteur aléatoire, pour limiter prévenir les attaques de type *sensitivity attacks*. Nous décrirons ici uniquement les points nécessaire à la compréhension de nos résultats, et renvoyons le lecteur désireux d'en savoir plus à l'article [FB08] et à l'implémentation téléchargeable sur le site du concours BOWS-2 [6].

L'insertion s'effectue dans le domaine des ondelettes. En se référant aux notations de l'introduction de ce chapitre, le signal tatoué $\mathbf{y}$ est obtenu par l'addition du signal orginal $\mathbf{x}$ comme avec un signal de tatouage $\mathbf{w}$: $\mathbf{y} = \mathbf{x} + \mathbf{w}$. Le signal de tatouage est de la forme $\mathbf{w} = \text{mask}\ \mathbf{s_w}$, avec $\mathbf{s_w}$ le signal généré dans le domaine des ondelettes en fonction de la clé secrète, et mask le masque perceptuel qui assure l'imperceptibilité. Dans le schéma original, on a $\text{mask}_{\texttt{BA}} = |\mathbf{x}|$, où $|\mathbf{x}|$ désigne la valeur absolue des coefficients d'ondelette du signal hôte $\mathbf{x}$, qui sont les coefficients choisis parmi toutes les sous-bandes, sauf la sous-bande de basse fréquence LL. Pour renforcer la robustesse de `Broken Arrows`, nous avons pris en compte la dépendance entre les coefficients voisins lors de l'insertion. Nous avons proposé de remplacer

6. `http://bows2.ec-lille.fr/`

chaque coefficient (sauf dans la sous-bande LL) par une moyenne de cinq coefficients : lui-même $\mathbf{x}[k,\ell]$ et ses quatre voisins $\mathbf{x}[k-1,\ell]$, $\mathbf{x}[k,\ell-1]$, $\mathbf{x}[k+1,\ell]$, et $\mathbf{x}[k,\ell+1]$, obtenant ainsi un nouveau masque :

$$\mathrm{mask}_{\texttt{AWC}}[k,\ell] = \frac{1}{5}\left|\sum_{s=k-1}^{k+1}\sum_{t=\ell-1}^{\ell+1}\mathbf{x}[s,t]\right|.$$

En collectant tous les $\mathrm{mask}_{\texttt{AWC}}[m,n]$ ensemble, nous obtenons le masque $\mathrm{mask}_{\texttt{AWC}}$ de l'insertion proportionnelle `BA-AWC`. Intuitivement, cette insertion renforce la dépendance entre les coefficients voisins du signal tatoué. Pour une position donnée, le masque a une valeur plus grande que l'amplitude d'au moins un des cinq coefficients considérés. Selon la valeur du signal tatoué, l'insertion pourrait par conséquent modifier le signe du coefficient d'ondelette. Ainsi, la présence de la marque est non seulement cachée dans les amplitudes des coefficients, mais également dans certains de leurs signes. Dans nos expériences, environ 6% des coefficients d'ondelette voient leur signe modifié par cette nouvelle insertion.

Afin d'évaluer notre proposition, nous l'avons tout d'abord confronté à la série de tests de robustesse utilisée lors de l'évaluation du schéma original [FB08]. Ces tests ont porté sur le même ensemble de 2000 images, et ont consisté en des attaques résultant de combinaisons de compressions JPEG ou JPEG 2000 aux facteurs de qualité variés, avec des changements d'échelle. La version `BA-AWC` est un peu moins robuste que le `Broken Arrows` original pour certaines attaques, mais de même niveau de robustesse pour d'autres. En tout les cas, la perte éventuelle en robustesse est tout-à-fait acceptable.

A. Westfeld a utilisé dans ses expériences un ensemble de 10000 images, incluant les 2000 images que nous avons utilisées dans toutes nos expériences [Wes08]. Néanmoins, par souci d'homogénéité avec les tests de robustesse, nous avons conservé cet ensemble de 2000 images pour les tests liés à l'attaque de Westfeld. Cette différence donne lieu à de légères variations entre ses résultats et les notres quant aux performances de son attaque sur `Broken Arrows`, mais sans conséquence sur les conclusions de ce travail. Le PSNR des images attaquées s'étend de 19.9 à 46.2 dB (il varie de 19.7 à 45.0 dB dans [Wes08]). La figure 2.8 montre la diminution du pourcentage des images attaquées avec succès quand le PSNR augmente. Pour l'insertion originale, l'attaque de Westfeld est vraiment dévastatrice, avec un succès de 100% pour les images lorsque le PSNR est inférieur à 30 dB, et même si son efficacité diminue lorsque le PSNR s'augmente, elle reste très efficace pour 40% des images quand le PSNR est autour de 35 dB. Notre variante `BA-AWC` résiste par contre très bien à cette attaque, puisque le pourcentage des images attaquées avec succès est proche de 0 pour n'importe quel PSNR.

Ainsi, `BA-AWC` remplit parfaitement son rôle : offrir, comme `Broken Arrows`, une très bonne robustesse générale, tout en résistant efficacement à l'attaque de Westfeld.

Ces travaux ont été présentés à la conférence internationale *IS&T/SPIE International Symposium on Electronic Imaging'2009* [CXFF09] ainsi qu'au GRETSI 2009 [XFF09]. Ils ont été valorisés dans le cadre des projets ANR-RIAM ESTIVALE et MEDIEVALS.

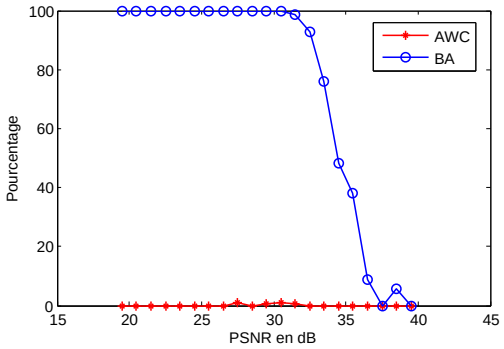


FIGURE 2.8: Résistance à la première attaque de Westfeld [Wes08] : comparaison du `Broken Arrows` original (`BA`) et denotre variante `BA-AWC`. On trace le pourcentage d'images attaquées avec succès en fonction du PSNR des images attaquées.

Amélioration de la sécurité de `Broken Arrows`. Lors de la phase de BOWS-2 dédiée aux attaques de type sécurité, les participants avaient une connaissance parfaite du système, et pouvaient observer beaucoup d'images tatouées avec la même clé secrète. A. Westfeld a alors conçu une nouvelle attaque, qui s'appuie sur la première mais est plus efficace. P. Bas, un des concepteurs de `Broken Arrows`, a de son côté réfléchi aux éventuelles failles de sécurité du système. Ils ont publié l'ensemble de leurs conclusions dans [BW09]. Avec Fuchun Xie et Teddy Furon, nous avons proposé des améliorations de `Broken Arrows` pour contrer ces attaques, et avons poussé la réflexion plus avant vers la théorisation des critères de sécurité.

Dans son attaque, A. Westfeld procède par regroupement (clustering), et s'appuie sur son attaque par débruitage. Comme notre variante `BA-AWC` neutralise cette attaque par débruitage, elle contre aussi *a priori* l'attaque par regroupement. Nous avons néanmoins tenu à évaluer expérimentalement les performances de `BA-AWC` face à l'attaque par regroupement, sur le même lot d'images. Nous avons mesuré la précision du regroupmenet grâce à l'information mutuelle ajustée. Pour la technique d'insertion originale, l'information mutuelle ajustée est de 0.7 ; cela signifie que les groupes estimés sont très similaires aux vrais groupes, et donc l'attaque de regroupement réussit à estimer les cônes secrets (un document est considéré comme tatoué tant qu'il se situe dans un de ces cônes) avec une bonne précision. Toutefois, ce classifieur ne fonctionne pas avec `BA-AWC`, car l'information mutuelle ajustée est alors inférieure à 0.05. La technique d'insertion proportionnelle `BA-AWC` est donc effectivement une solution efficace pour bloquer l'attaque de regroupement de Westfeld.

L'attaque d'estimation de sous-espace de Bas ne s'appuie pas, elle, sur l'attaque de débruitage de Westfeld. Elle s'appuie en revanche sur une estimation de sous-espace par des techniques de type PCA, *via* l'implémentation OPAST. Elle exploite le fait que l'insertion modifie la distribution des puissances du signal dans le sous-espace secret, ce qui donne une piste pour évaluer ce sous-espace. Nous avons proposé un outil de mesure de l'efficacité de l'attaque au travers de la distance chordale car-

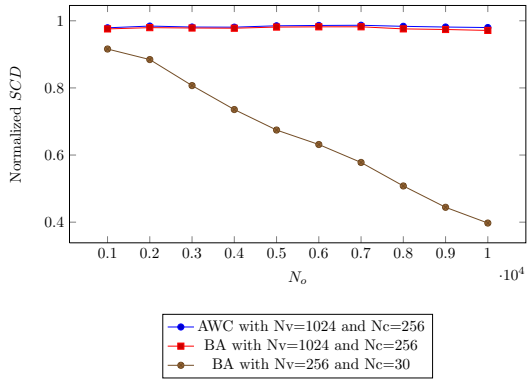


FIGURE 2.9: Distances chordales normalisées SCD_{norm} comparées pour le `Broken Arrows` original (`BA`) avec différentes valeurs des paramètres N_v et N_c, et notre variante `BA-AWC+`. Rappelons que les paramètres originaux de `Broken Arrows` sont $N_v = 256$ et $N_c = 30$.

rée normalisée SCD_{norm} appliquée au sous-espace secret et au sous-espace estimé. Cette distance vaut 0 lorsque le sous-espace estimé est égal au sous-espace secret, et 1 lorsque les sous-espaces sont orthogonaux et que, par conséquent, l'attaque a échoué. Une fois cet outil en place, nous avons proposé un nouveau réglage des paramètres du système pour réduire cette différence, prenant $N_v = 1024$ et $N_c = 256$ alors que dans sa version originale, `Broken Arrows` utilisait les paramètres $N_v = 256$ et $N_c = 30$. Nos résultats expérimentaux, présentés dans la figure 2.9, ont montré que, pour la technique d'insertion originale, SCD_{norm} est décroissante avec le nombre d'observations, et que l'estimation converge très rapidement. Cela confirme les résultats de P. Bas [BW09]. Toutefois, lorsque l'on considère les paramètres $N_v = 1024$ et $N_c = 256$ au lieu de $N_v = 256$ et $N_c = 30$, SCD_{norm} diminue très lentement : même après 3.10^4 observations, SCD_{norm} est toujours très proche de 1 (supérieure à 0.9), ce qui montre que l'attaque ne peut plus estimer les sous-espaces efficacement. C'est vrai lorsque l'on applique ce changement de paramètres à `Broken Arrows`, mais aussi lorsqu'on l'applique à `BA-AWC`. Dans la suite, on appellera `BA-AWC+` cette amélioration de `BA-AWC`. En appliquant à nouveau le même benchmark que dans l'article original de `Broken Arrows` pour examiner l'incidence sur la robustesse causée par les nouvelles améliorations proposées, on constate que la robustesse est légèrement diminuée, mais pas dramatiquement. Ainsi, notre contre-mesure donne une grande amélioration des niveaux de sécurité en sacrifiant seulement peu de robustesse. On sacrifie en revanche un peu plus sur le temps d'insertion qui, avec ces nouveaux paramètres, est multiplié par 4.

Ces résultats étaient d'ores et déjà satistfaisants. Néanmoins, nous avons voulu aller plus loin afin d'anticiper de futures attaques potentielles. Il est par exemple possible que le pirate dispose un jour d'implémentations de PCA plus puissantes que OPAST, et qu'il puisse collecter plus d'images tatouées pour rechercher le sous-espace secret. Nous avons donc choisi de renforcer encore l'insertion pour contrer des

attaques basées sur des statistiques de second ordre, en restituant une distribution de puissance parfaitement uniforme. Ces améliorations ne sont pas détaillées ici. Elles comprennent : 1) l'introduction d'un critère de sécurité, 2) un processus d'insertion mis en œuvre comme une maximisation de la robustesse, sous les contraintes de la perception et de la sécurité, et 3) une détection de tatouage reposant sur un critère de décision *a contrario*. Nous sommes arrivés à une formalisation de la sécurité de `Broken Arrows` au regard de ce type d'attaque, et en avons ainsi proposé une nouvelle version, que j'appelerai ici `BA-AWC++`, qui offre une meilleure sécurité que les précédentes. Malheureusement, nous avons constaté qu'en augmentant ainsi encore le niveau de sécurité on perd beaucoup en robustesse, et que la probabilité de fausse alarme augmente fortement. Il faut donc là encore trouver un compromis acceptable, en fonction du contexte applicatif. On peut par exemple remarquer que dans certains contextes, comme le traçage de documents multimédia abordé au chapitre suivant, le gain global est néanmoins très intéressant. Tout d'abord, dans ce contexte précis la probabilité de fausse alarme n'a pas d'importance car tous les documents diffusés sont tatoués. Par ailleurs, notre test de décision *a contrario* peut alors fournir une probabilité de la présence d'un symbole donné dans le tatouage, donnant lieu à non plus à une information dure (présence ou absence), mais à une information souple, plus riche, et qui peut donner lieu à une amélioration du processus d'accusation du code traçant.

Ces travaux ont été présentés lors des conférences internationales *IS&T/SPIE International Symposium on Electronic Imaging'2010* [XFF10a] et *ACM Multimedia & Security, MM&SEC'10* [XFF10b]. Ils ont été valorisés dans le cadre du projet ANR-RIAM MEDIEVALS. Ils mériteraient de faire l'objet d'une publication dans une revue.

2.5 Conclusion et perspectives

Le tatouage offre de nombreux avantages fonctionnels, qui en font un allié précieux des techniques classiques de protection comme la cryptographie. Développées depuis une vingtaine d'années, les techniques de tatouage ont aujourd'hui atteint une bonne maturité, et concilient capacité et imperceptibilité avec une très bonne robustesse. Néanmoins, l'étude de la sécurité des schémas de tatouage que nous avons menée (section 2.2) a montré la nécessité de bien distinguer robustesse et sécurité, car même très robuste, un schéma de tatouage peut s'avérer peu sûr si les clés ne sont pas changées suffisamment souvent. Par ailleurs, comme l'ont illustré d'une part notre travail sur l'amélioration de la robustesse et de la sécurité de la technique de tatouage `Broken Arrows` (section 2.4), et d'autre part les travaux menés par Bas *et al.* et Guyeux *et al.*, de nombreux points restent à éclaircir concernant les compromis à réaliser entre ces deux notions (sans oublier bien sûr les traditionnelles imperceptibilité et capacité). Et au-delà de l'analyse *a posteriori*, la question de la prise en compte de la notion de sécurité dès la conception d'un système, que ce soit de manière empirique ou formelle, reste en grande partie ouverte, même si quelques travaux ont commencé à creuser cette question. Il reste donc, au-delà des améliorations des

critères classiques, à prendre en compte la sécurité de manière plus systématique qu'elle ne l'est aujourd'hui, pour améliorer encore les techniques de tatouage.

Pour ma part, les points que je souhaite développer dans l'avenir portent plus spécifiquement sur l'articulation de ces techniques de tatouage avec des codes anti-collusion, afin de permettre l'identification d'utilisateurs qui distribuent de manière illicite des documents en leur possession (cf. chapitre suivant). Deux pistes retiennent mon attention dans ce contexte : l'articulation des deux techniques à proprement parler, en prenant en considération des préoccupations de sécurité, et leur intégration dans des protocoles de distribution.

Chapitre 3

Identifier la provenance de fraudes grâce à la personnalisation de copies (*active fingerprinting*)

> Des traces de lion, toutes fraîches ; on voit que le fauve s'est couché là ; ces demi-cercles ont été tracés par sa queue.
>
> André Gide

Le mot *fingerprinting* peut avoir plusieurs sens, y compris quand on traite de sécurité des contenus. Je m'intéresse pour ma part au *fingerprinting* dont l'objet est de personnaliser chaque document délivré en y insérant un identifiant pour chaque transaction, afin de pouvoir tracer les utilisateurs indélicats qui le mettrait à disposition de personnes non autorisées. Des scénarios typiques sont par exemple la vente de vidéos à la demande, la distribution de films aux membres du jury des oscars, la sous-traitance lors de la production de films, la diffusion de données sur les réseaux sociaux, etc. L'insertion de l'identifiant s'effectue par tatouage, et doit être robuste aux manipulations classiques comme la compression, la renumérisation après passage en analogique, etc. Mais on doit ici en sus faire face à une nouvelle catégorie d'attaques, les attaques par *collusion*. En effet, si on utilise une technique d'insertion assez robuste, il est relativement facile d'identifier une copie provenant directement d'un unique utilisateur. Une stratégie d'attaque consiste donc pour un ensemble d'utilisateurs, qui possèdent chacun une copie d'un même document, à former une coalition et à mélanger leurs documents pour forger un document dont l'identifiant ne correspondra a aucun des leurs, afin de brouiller les pistes. C'est ce qu'illustre la figure 3.1. On cherche donc à élaborer des identifiants tels que si les documents qui les contiennent se retrouvent mélangés pour produire un nouveau document, l'identifiant extrait de ce dernier permette quand même de remonter à au moins un des identifiants d'origine, et donc à un des membres de la coalition.

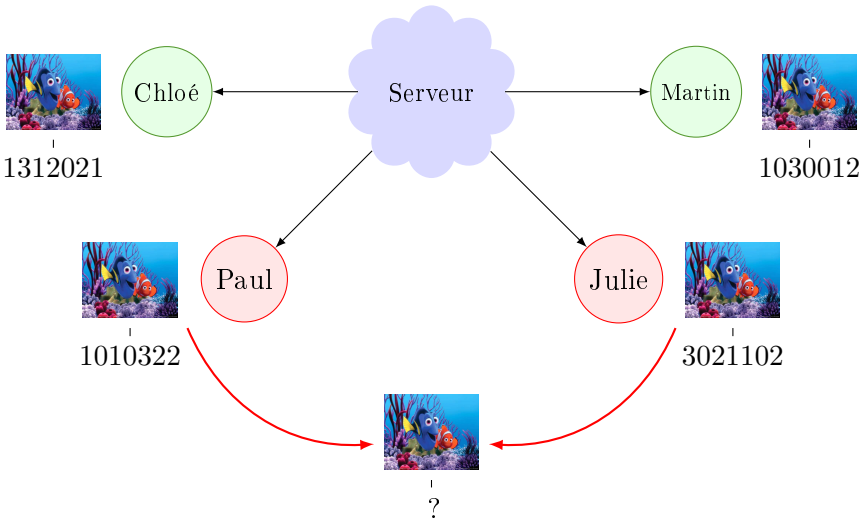


FIGURE 3.1: Principe d'une attaque par *collusion*.

3.1 Introduction à la personnalisation de copies

3.1.1 Historique

Personnaliser des copies lors de leur diffusion pour pouvoir ensuite identifier une redistribution illicite n'est pas nouveau. John Neper a ainsi protégé au 17ème siècle ses tables de logarithmes en en modifiant quelques décimales insignifiantes par-ci par-là selon les copies qu'il en distribuait. Plus récemment, suite à la publications de documents confidentiels dans la presse, Margareth Thatcher a personnalisé les documents qu'elle distribuait à ses ministres. Plus récemment encore, suite à la redistribution illicite de films confiés au jury des oscars, ou à des entreprises de sous-traitance en production, les producteurs de cinéma personnalisent aujourd'hui les versions numériques des films qu'il leur confient. Mais ces premières protections n'envisageaient pas de coalition.

Les premiers travaux formalisant la personnalisation de documents numériques en vue de l'identification d'utilisateurs indélicats ont été publiés par Wagner en 1983 [Wag83] et Blakley *et al.* en 1985 [BMP86] (donc avant l'avénement des techniques de tatouage). Depuis, de très nombreux travaux ont été publiés. La plupart d'entre eux considèrent que la protection s'appuie sur deux couches distinctes : la génération d'un identifiant à l'aide d'un *code anti-collusion*, puis son insertion par une technique de tatouage appropriée.

Plusieurs modèles d'attaques ont été proposés dans la littérature pour la conception des tels codes anti-collusion. Le plus répandu est celui introduit par Boneh et Shaw en 1995 [BS95, BS98]. Un modèle alternatif, très peu étudié, a ensuite été introduit en 2005 par Somekh-Baruch et Merhav [SBM05]. Ces deux modèles sont présentés et discutés dans les paragraphes suivants. Les codes anti-collusions associés à ces modèles ont des structures combinatoires particulières, et leur étude a en

grande partie été menée par des chercheurs en théorie des codes (les codes de Tardos, par leur nature probabiliste, font exception à la règle), qui se sont focalisés sur les propriétés de traçage sans se préoccuper de l'insertion en elle-même.

Enfin, quelques schémas complets ont été proposés, prenant en compte l'insertion des identifiants, ce qui donne lieu à une vision différente des attaques qui peuvent être menées. Ils sont présentés à la fin de cet historique. Ils ont montré comment les deux couches peuvent tirer parti l'une de l'autre, mais aussi quelles contraintes elles génèrent l'une pour l'autre.

Dans les pas de Boneh et Shaw. La plupart des travaux menés sur les codes anti-collusion s'appuient sur le modèle d'attaque proposé par Boneh et Shaw en 1995 [BS95,BS98] : les attaquants comparent leurs copies et forgent une copie pirate en reprenant (a) telles qu'elles les composantes qui sont identiques pour toutes les copies considérées, et (b) en prenant pour les autres des éléments dans chacune des copies. Quelques variantes du modèle ont été considérées, prenant en compte la possibilité pour les attaquants de remplacer (b) par : (b′) n'importe quel symbole pour l'identifiant pour les composantes non toutes identiques, autorisant ainsi les erreurs, ou même des effacements [GP00] [1]. On retrouve dans toutes ces études une hypothèse commune, connue sous le nom de *marking assumption*, qui formalise le point (a) en stipulant que si les membres de la coalition ont tous le même symbole d'identifiant à une certaine position, alors la copie qu'ils vont forger ensemble aura elle aussi ce même symbole à cette position [2] est parfois utilisé dans un sens différent, stipulant que les membres de la coalition ne peuvent produire pour une composante donnée un symbole qu'ils n'ont pas, regroupant ainsi les points (a) et (b). Nous nous en tiendrons ici à la définition qui concerne seulement le point (a), moins restrictive pour ce qui est des attaques, et qui est donc la plus réaliste. C'est la plus utilisée aujourd'hui. Ces modèles sont rappelés dans la figure 3.2, car nous nous y référerons souvent.

Notons que ce modèle peut à première vue sembler peu réaliste, car il ne s'appuie que sur la couche « message », en faisant abstraction du support lui-même, et qu'il limite les actions des pirates (en imposant par exemple la *marking assumption* (a)). Néanmoins, lors de la mise en place d'un système réel de protection, on peut s'arranger pour rendre le modèle étendu (a)+(b′) effectif. Classiquement, on va découper le document à personnaliser en blocs, et cacher un symbole de l'identifiant par bloc. Ainsi, le mélange opéré par les pirates aura des répercussions bloc par bloc, et on pourra raisonner de manière indifférenciée sur les blocs de document ou sur les symboles associés, la préservation d'un bloc donnant lieu directement à la préservation du symbole correspondant. Ceci permet par ailleurs de préparer des versions

1. On pourrait encore raffiner ce modèle étendu pour différencier les cas où l'on autorise par exemple les effacements mais pas d'autres symboles de l'alphabet, comme Boneh et Shaw l'avaient proposé pour certains de leurs résultats dès le départ. Certains articles se placent ainsi dans des sous-modèles de notre modèle (a)+(b′). Néanmoins, le modèle (a)+(b′) tel que nous l'avons défini est le plus réaliste, et puisqu'il existe aujourd'hui des codes qui s'en accomodent très bien, par souci de concision il ne m'a pas semblé utile ici d'en proposer des raffinements.

2. Le terme *marking assumption*

Les membres de la coalition élaborent une copie pirate à partir des différentes copies légitimes du document qu'ils ont en leur possession. Chacune de ces copies contient un identifiant différent. L'identifiant de la copie pirate est considéré comme satisfaisant les propriétés (a)+(b) pour le modèle initial, ou (a)+(b′) pour le modèle étendu.

(a) *Marking assumpion* : pour les composantes où les identifiants des copies légitimes sont identiques, l'identifiant de la copie pirate leur sera également identique.

(b) Pour les composantes où les identifiants des copies légitimes ne sont pas tous identiques, le symbole présent dans la copie pirate est égal au symbole provenant d'une des copies d'un membre de la coalition.

(b′) Pour les composantes où les identifiants des copies légitimes ne sont pas tous identiques, le symbole présent dans la copie pirate peut être égal à n'importe quel symbole de l'alphabet (ce qui peut s'apparenter à une erreur), ou un effacement.

Le modèle (a)+(b) correspond par exemple à la situation suivante :

Paul	1	0	1	0	3	2	2
Julie	3	0	2	1	1	0	2
Martin	1	0	3	0	0	1	2
copie pirate	1	0	3	1	0	2	2

Tandis que le modèle (a)+(b′) peut donner lieu à :

Paul	1	0	1	0	3	2	2
Julie	3	0	2	1	1	0	2
Martin	1	0	3	0	0	1	2
copie pirate	0	0	?	0	2	?	2

En jaune : la *marking assumption* (a).

FIGURE 3.2: Les modèles issus des travaux de Boneh et Shaw [BS95, BS98, GP00].

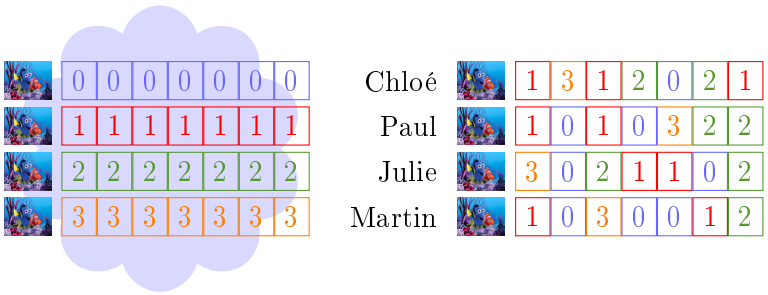


FIGURE 3.3: Principe du découpage en blocs et aiguillage. Chaque document est découpé en une série de blocs. Dans chacun de ces blocs est inséré un symbole de l'identifiant. On pré-tatoue autant de copies maîtres qu'il y a de symboles dans l'alphabet, chaque copie maître contenant le même symbole dans tous ses blocs. Les copies personnalisées sont ensuite composées à la volée lors de la transaction (vidéo à la demande), ou du rendu utilisateur (lecture d'un disque blu-ray, ou décodeur TV par exemple).

maîtres tatouées hors-ligne avant toute transaction, chacune contenant toujours le même symbole de l'alphabet ; ces versions maîtres serviront ensuite à composer les documents personnalisés à la volée lors des transactions : il suffira d'intercaler les bons blocs des versions maîtres pour composer le document personnalisé, en fonction de l'identifiant associé.

Les études menées dans ce modèle ((a)+(b) pour sa version initiale, (a)+(b′) dans sa version étendue) reposent pour la plupart sur les structures combinatoires des codes correcteurs d'erreur, les identifiants étant précisément des mots de code. Ils ont abouti à la définition de niveaux de sécurité quant au traçage, niveaux regroupés dans deux grandes catégories : la *traçabilité forte* (qui comprend par exemple les classes de codes *c-IPP* et *c-traceable*), et la *traçabilité faible* (qui comprend par exemple les classes de codes *c-secure* et *c-frameproof secure*). Dans les deux cas, on essaie d'éviter l'incrimination d'innocents, tout en s'assurant d'attraper un des coupables.

En traçabilité forte, on ne tolère aucun faux pas lors de l'accusation. Elle n'est malheureusement accessible, dès que $n \geq 3$ et $c \geq 2$, que dans le modèle (a)+(b), et jamais dans le modèle étendu (a)+(b′) [BS98]. Par ailleurs, il a été prouvé que pour satisfaire aux exigences de la traçabilité forte, les codes anti-collusion doivent avoir une très grande longueur, et être définis sur de gros alphabets [HvLLT98, BCE+01, SSW01], ce qui rend l'implémentation et le décodage très lourds [HW05], et même impossibles en pratique à l'aide des algorithmes de tatouage actuels. Certaines constructions de tels codes ont néanmoins été étudiées, comme par exemple [HvLLT98, SW98, CFNP00, BCE+01, SSW01, SNW02, SSW03, FS04b, FS04c, BK04, FS04a, SFC05].

En traçabilité faible, on s'autorise quelques erreurs de jugement maîtrisées (borne sur la probabilité d'accuser à tort), et on peut alors utiliser des codes beaucoup plus faciles à manipuler. Ce niveau de traçabilité est accessible dans les deux modèles, (a)+(b) et (a)+(b′). On dispose de candidats binaires déterministes issus

de la théorie des codes, parfois mâtinés de permutations aléatoires, comme par exemple [BS98, GP00, CFNP00, BBK03, Sch04, ZFZ05] (et beaucoup d'autres qui ne sont pas cités ici). Tardos a de son côté proposé des codes radicalement différents, puisque probabilistes, sur lesquels j'ai travaillé ces dernières années. En 2003, Peikert *et al.* [PSS03] et Tardos [Tar03, Tar08] ont publié indépendamment la meilleure borne inférieure connue pour la longueur m d'un code à traçabilité faible admettant une probabilité ε que globalement on accuse un utilisateur à tort, en gérant n identifiants/utilisateurs et une taille de coalition d'au plus c : m doit alors être au moins en $O(c^2 \ln(n/\varepsilon))$. Tardos, statisticien jusqu'alors inconnu de la communauté du fingerprinting, a alors surpris la communauté en publiant un code traçant probabiliste pour le cas binaire, qui est à ce jour le seul à atteindre cette borne [3]. Ce code est asymptotiquement optimal en longueur, très efficace en termes de traçage, et de surcroît très facile à mettre en œuvre. Et même si dans certains contextes, il est possible de construire des codes encore plus courts, comme le montrent par exemple [KH07, KHN+08] pour des coalitions de 3 utilisateurs, on peut dire sans exagérer que sa publication a donné lieu à une véritable révolution.

Le modèle de Somekh-Baruch et Merhav. Un deuxième modèle a été proposé en 2005 par Somekh-Baruch et Merhav [SBM05], qui offre un champ d'actions beaucoup plus large aux pirates. Dans ce modèle, on suppose que les copies sont toutes à distance (de Hamming) bornée de l'original, et que la copie pirate forgée par une coalition est à distance bornée d'une des copies légitimes appartenant à la coalition. Les auteurs ont étudié dans leur article la proposition de solutions de traçage dans ce modèle ; ils ont montré que sous des hypothèses de régularité et de composition constante le nombre d'utilisateurs qu'un schéma de traçage peut asymptotiquement gérer tend vers zéro ; ils considèrent que les hypothèses de régularité et composition constante n'interviennent pas de manière significative dans la preuve, et sont donc très pessimistes sur la possibilité de proposer des solutions concrètes dans ce cadre. Néanmoins, Galand a montré depuis [Gal05] que si on considérait des codes ne vérifiant pas ces hypothèses alors on pouvait construire des schémas de traçage réalistes et tout-à-fait efficaces algorithmiquement parlant ; ils nécessitent cependant que la distorsion maximale entre les copies légitimes et la copie pirate soit très faible, ce qui est assez restrictif et pas toujours réaliste. Ce modèle n'a pas donné lieu à d'autres publications depuis.

Lier le code anti-collusion au signal. Mais pour concevoir un schéma opérationnel, on ne peux considérer indépendamment les deux couches, car elles vont fortement interagir. Il faut donc choisir ou concevoir le code anti-collusion en même temps que l'on choisit ou conçoit l'insertion des identifiants.

La prise en compte de la nature du signal et d'une technique de tatouage implique de prendre en compte des manipulations du signal lui-même. Les attaques ne sont donc plus formalisées au niveau des identifiants, mais des composantes du signal.

3. Les codes proposés par Boneh et Shaw dans [BS95, BS98] avaient une longueur en $\mathcal{O}(n^3 \ln(n/\varepsilon))$ pour le premier, et $\mathcal{O}(c^4 \ln^2(1/\varepsilon))$ avec $\varepsilon < 1/n$ pour le second.

Néanmoins, l'hypothèse de la *marking assumption* (a) est considérée ici encore comme incontournable, et les attaques vont être classées en fonction de leur impact au niveau des identifiants, selon qu'elles donnent lieu aux situations (a)+(b) ou (a)+(b′) du modèle de Boneh et Shaw. Ainsi, le choix au hasard d'un bloc d'un des utilisateurs, pris tel quel, correspond par exemple au modèle (a)+(b), tandis que les attaques non-linéaires opérées sur les blocs des copies, comme par exemple le moyennage de ces blocs pour obtenir le bloc pirate, correspondent au modèle étendu (a)+(b′), leur impact sur les identifiants étant beaucoup plus difficile à prévoir et à exploiter. Une typologie plus précise de ces attaques est donnée plus loin.

Quelques travaux ont été menés dans ce sens, principalement dans la communauté de traitement du signal/images. Certains ont tenté d'obtenir les propriétés de traçage directement par la nature des signaux cachés, en utilisant par exemple des signaux orthogonaux comme identifiants, et en les insérant par des modulations avec étalement de spectre [WWZ+03, WWZ+05]. Malheureusement, cette approche donne lieu à des calculs trop coûteux dès que le nombre d'utilisateurs n et la taille de la coalition c augmente, et à une probabilité de fausse alarme élevée. Une variante récursive a été proposée dans [TWWL03], qui permet de mieux faire face à un grand nombre d'utilisateurs. Néanmoins, elle souffre comme la version originale d'une trop forte complexité pour faire face à de grandes coalitions, et sa probabilité de fausse alarme est difficile à évaluer, et donc à maîtriser.

Une extension de ces travaux a conduit à l'utilisation de signaux orthogonaux comme porteuses, modulées par des mots de codes anti-collusion *frameproof* [TWWL03]. Il s'agit alors de procéder précisément comme dans un schéma de tatouage habituel, avec comme message un mot de code anti-collusion. Trappe *et al.* ont proposé dans cet article de modéliser l'impact de la coalition sur les mots de code comme un ET logique, opérant composante par composante. Ils proposent également de nouveaux codes anti-collusion, reposant sur les *designs*, dont ils estiment les performances en s'appuyant sur des attaques par moyennage et ajout de bruit.

He *et al.* se sont inspirés de ces travaux en les combinant avec des codes anti-collusion de Safavi-Naini *et al.* [SNW02], pour en réduire la complexité de traçage [HW06, HW07]. Elles proposent également un comparatif, reproduit ici dans la figure 3.4, qui permet de cerner rapidement les ordres de grandeur des schémas que nous venons de mentionner, en termes de taille de coalition, de longueur d'identifiants (et donc de durée de vidéo nécessaire à l'insertion), et de complexité de traçage, pour un contexte applicatif réaliste de service de vidéo à la demande avec un très grand nombre d'utilisateurs. Zhao *et al.* [ZWWL03, ZWWL05] ayant montré que lorsque les identifiants suivent une distribution gaussienne, la plupart des attaques non-linéaires s'apparentent à un moyennage avec du bruit, He *et al.* se sont appuyés sur cette assertion pour estimer la robustesse de leur schéma. Néanmoins, Schaathun a proposé dans [Sch08a] une nouvelle attaque non-linéaire (MMX) mettant à mal ce système. Il a montré que la nature du bruit (gaussien ou uniforme) ajouté à la moyenne des blocs dans la modélisation de Zhao *et al.* a de fortes implications sur le comportement de l'attaque, et que si globalement cette assertion peut être considérée comme vraie il faut néanmoins rester prudent. Il a poursuivi l'étude par une réflexion sur l'équivalence (ou la non-équivalence) entre les différentes attaques

Schémas	c	durée de vidéo nécessaire	complexité de traçage
code binaire de [BS98, Yac01]	10	22 heures (10^7 bits)	$\mathcal{O}(N_{ech}+n)$
codes q-aires de [SNW02, FS04c]	< 10	quelques minutes	$\mathcal{O}(\sqrt[k]{n}(N_{ech}+n))$
sigaux orthogonaux [WWZ+05]	~ 100	quelques minutes	$\mathcal{O}(N_{ech}n)$
Trappe *et al.* [TWWL03]	~ 100	quelques minutes	$\mathcal{O}(\sqrt{N_{ech}}N_{ech}+n)$
He *et al.* [HW06]	~ 100	quelques minutes	$\mathcal{O}(\sqrt[k]{n}(N_{ech}+n))$

FIGURE 3.4: Comparaison de quelques schémas, proposée dans [HW07] : n est le nombre d'utilisateurs, ici pris égal à 10^7, N_{ech} le nombre d'échantillons de signal hôte, la probabilité d'accuser à tort de 10^{-3}.

non-linéaires [Sch08b].

En parallèle, Schaathun a essayé de tirer parti du modèle de Boneh et Shaw, qui permet de s'assurer au moins que dans le cas d'attaques de type (a)+(b) tout se passe bien. Il a bien entendu également considéré des attaques de type (a)+(b′). Il a ainsi étudié comment la construction de codes anti-collusion de Boneh-Shaw avec décodage souple peut être conciliée avec une technique d'insertion, dans le modèle réaliste (a)+(b′) [Sch08b]. Dans sa conclusion, il suggère de mener des études similaires avec d'autres codes anti-collusion issus du modèle de Boneh et Shaw, pour lesquels on dispose d'arguments théoriques quand à leur comportement et à leurs performances, comme les codes de Tardos.

C'est précisément ce que nous avions étudié en parallèle, avec Teddy Furon et Fuchun Xie. Nous avons montré dans [XFF08] que les codes de Tardos se combinent très à une technique d'insertion zéro-bit, comme `Broken Arrows`. Cette technique est en effet tellement robuste qu'elle permet de détecter plusieurs symboles dans le cas d'attaques de type fusion, facilitant ainsi le traçage. Nous avons adapté les codes de Tardos à cette détection multiple, pour en tirer le meilleur parti. Ces résultats sont présentés dans le paragraphe 3.2.

Plus récemment, Mathon *et al.* ont étudié quelles contraintes l'utilisation des codes de Tardos impliquent pour la technique de tatouage utilisée pour l'insertion des identifiants [MBCM10b, MBCM10a].

On dispose donc aujourd'hui de techniques de traçage efficaces, que l'on arrive à combiner avec des techniques d'insertion présentant toutes les qualités requises.

3.1.2 Quelques considérations supplémentaires.

Avant de présenter plus en détail les codes de Tardos, il convient de discuter d'un certain nombre de questions.

Fingerprinting* vs. *Traitor Tracing Les deux termes sont parfois utilisés indifféremment par la communauté du *fingerprinting*. Néanmoins, les cryptographes s'intéressent depuis plus longtemps au *traçage de traitres*, et préfèrent distinguer les deux. Le scenario du traçage de traitres est en effet différent de celui considéré ici dans le sens où un document est « broadcasté »à l'ensemble des utilisateurs, chiffré de telle sorte que seuls les utilisateurs qui ont acquis légalement une clé de déchiffrement

puissent le déchiffrer. Ce sont ces clés personnelles qui sont mélangées par des coalitions de pirates pour forger de fausses clés qui déchiffrent correctement mais sont plus difficiles à tracer. Cette différence de scenario donne lieu à des contraintes techniques légèrement différentes. Néanmoins, certains travaux sur le *fingerprinting* se sont inspirés des codes développés dans le contexte du traçage de traitres [BS95,BS98], et le traçage de traitres a su lui aussi tirer parti des avancées en fingerprinting multimédia, *e.g.* [BHP08].

Découpage en blocs et aiguillage. Comme discuté précédemment dans le paragraphe dédié au modèle de Boneh et Shaw, le principe du découpage en blocs et aiguillage est illustré par la figure 3.3. Chaque document est découpé en une série de blocs. Appliquant le principe de Kerckhoffs, on suppose que ce découpage est public, donc connu de tous. Dans chacun de ces blocs est inséré un symbole de l'identifiant. On pré-tatoue autant de copies maîtres qu'il y a de symboles dans l'alphabet, chaque copie maître contenant le même symbole dans tous ses blocs. Les copies personnalisées sont ensuite composées à la volée lors de la transaction (vidéo à la demande), ou du rendu utilisateur (lecture d'un disque blu-ray, ou décodeur TV par exemple).

Cette technique offre de nombreux avantages. Tout d'abord elle permet de faire coïncider les symboles de l'identifiant avec la suite des blocs tatoués, si bien que les actions des utilisateurs, qui vont s'effectuer au niveau de ces blocs, se répercutent directement au niveau des symboles. On peut ainsi raisonner en termes d'attaques sur les symboles directement, comme dans le modèle proposé par Boneh et Shaw, le rendant ainsi totalement réaliste et effectif. Par là même, ce procédé pousse la *marking assumption* (a) à être vérifiée. Ainsi, le bon fonctionnement de tous les codes anti-collusions conçus pour ce modèle, comme les codes de Tardos par exemple, est assuré. Le deuxième avantage de cette technique est qu'elle permet d'effectuer l'insertion par tatouage, tâche la plus coûteuse, hors-ligne.

Quelques hypothèses et notations. On note n le nombre d'utilisateurs du système, et c la taille maximum pour une coalition d'utilisateurs malhonnêtes. On désigne par $\mathbf{X}_j$, $1 \leq j \leq n$, les identifiants des n utilisateurs. Ces identifiants $\mathbf{X}_j = (X_{j1}, \ldots, X_{jm})$ sont des mots q-aires de longueur m, mots de code du code anti-collusion choisi. On note $\mathbf{Y} = (Y_1, \ldots, Y_m)$ le vecteur extrait de la copie pirate forgée par la coalition.

On suppose que lors de la création de la copie pirate, les membres de la coalition opèrent bloc par bloc, chaque bloc étant traité indépendamment des autres (canal sans mémoire).

Quelle taille de coalition doit-on envisager de contrer ? On ne dispose pas à l'heure actuelle d'exemples d'actions de coalitions d'utilisateurs faisant jurisprudence. Est-ce que des coalitions effectives peuvent raisonnablement compter 5, 10, 20 membres ? Probablement. Est-ce qu'elles peuvent en contenir beaucoup plus ? Cela dépendra grandement des applications. Intéressons-nous tout d'abord à la redistribution illicite de films téléchargés sur des plate-formes de VOD. Avec le développement

des réseaux sociaux on peut envisager que de grands groupes se créent facilement, sans lien préalable entre les membres. Si on fait abstraction des réseaux sociaux, de grands groupes peuvent également se créer au sein par exemple de campus universitaires ou d'écoles. Mais pour qu'une telle coalition puisse agir, encore faut-il que tous les membres soient en possession d'un exemplaire du film. Ce facteur peut être limitant dans le cas où la coalition regroupe des étudiants d'un même campus. Il faut par ailleurs que chacun ait confiance en les autres membres. Car si chaque membre de la coalition met à disposition de l'ensemble de la coalition son exemplaire du film, il court le risque que cet exemplaire se retrouve distribué tel quel, le mettant directement en cause. Dans le cas où la coalition s'est créée *via* un réseau social, c'est peut-être ce facteur de confiance qui risque de limiter la taille des coalitions. Néanmoins, il suffit de procéder à la création de la copie pirate sans obliger les membres de la coalition à dévoiler entièrement leurs copies légitimes, pour coutourner cette restriction. C'est par exemple ce que propose [BHP09]. Si l'on s'intéresse à d'autres contextes, comme la distribution de films aux membres du jury des Oscars, la sous-traitance en post-production, ou bien encore à la diffusion des données mises sur les réseaux sociaux, la taille des coalitions est beaucoup plus réduite.

Comme nous l'avons vu dans l'historique et dans la figure 3.4, les schémas qui permettent de contrer les plus grandes coalitions souffrent aujourd'hui d'une trop forte complexité lors du traçage. Le meilleur compromis actuel semble être celui des codes de Tardos, qui combinent les avantages d'être courts, simples à comprendre et implémenter, et faciles à tracer. Ils permettent aisément, pour des longueurs réalistes, de tracer des coalitions d'une vingtaine d'utilisateurs, voire plus.

Quelles attaques peut mener une coalition ? Les attaques peuvent être classées de différentes manières. On distinguera ici principalement trois classes d'attaques : les attaques qui tirent partie de la coalition et conservent les blocs tels quels, les attaques qui tirent partie de la coalition et fusionnent les blocs, et enfin les attaques qui visent la robustesse ou la sécurité de la technique de tatouage pour effacer l'identifiant (compression, filtrage, cryptanalyse, etc), et qui peuvent être menées individuellement.

La première catégorie d'attaques pourrait s'intituler « échange de blocs », ou « copier-coller » : les membres de la coalition choisissent comme i-ème bloc de la copie pirate qu'ils forgent l'un des i-ème blocs de leurs copies légitimes, tel quel, comme illustré par la figure 3.5. On peut prédire qu'à l'extraction, les composantes du vecteur $\mathbf{Y}$ qui sera extrait de la copie pirate seront précisément égales à des composantes d'identifiants de membres de la coalition, comme dans le modèle de base (a)+(b) de Boneh et Shaw. Plusieurs stratégies sont possibles, comme par exemple de prendre l'un des blocs au hasard suivant une distribution uniforme, prendre le bloc le plus fréquent, ou le moins fréquent, etc. La *marking assumption* (a) est alors respectée. En effet, si les blocs sont tous identiques, alors le bloc forgé leur sera identique, ce qui se traduit au niveau des identifiants par : si $X_{ji} = a$ pour tout $j \in \mathcal{C}$, alors $Y_i = a$.

La deuxième catégorie d'attaques pourrait s'intituler « fusion de blocs » : les membres de la coalition répartissent leur contribution à la création de la copie pirate, comme illustré par la figure 3.6. Plusieurs stratégies sont possibles. Si on prend

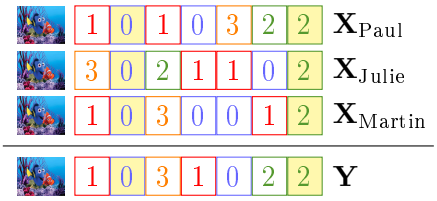


FIGURE 3.5: Attaques de type « échange de blocs », « copier-coller-fg. On peut prédire que les symboles extraits de la copie pirate sont issus directement des identifiants des membres de la coalition. En jaune : la *marking assumption* (a) est respectée. On est dans le modèle (a)+(b).

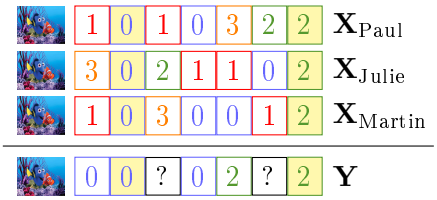


FIGURE 3.6: Attaques de type « fusion de blocs ». On ne peut pas prédire quels symboles seront extraits de la copie pirate, cela dépendra de la technique de tatouage utilisée. En jaune : la *marking assumption* (a) est cependant toujours respectée. On est dans le modèle (a)+(b′), qui autorise des erreurs et des effacements.

comme exemple des images et qu'on les considère comme des tableaux de pixels (ou encore de coefficients DCT, Fourier ou ondelettes), on peut créer le bloc de la copie pirate en prenant pour chacun de ses pixels (ou coefficients DCT, Fourier ou ondelettes) la moyenne — ou le min, ou le max, ou la valeur médiane, etc — des pixels des copies légitimes des membres de la coalition. On peut même envisager que le bloc de la copie pirate est constitué de morceaux issus des blocs des copies des membres de la coalition. Comme évoqué page 61, ces attaques non-linéaires ont été proposées et étudiées notamment dans [ZWWL03,ZWWL05,Sch08a], qui ont cherché à voir si on peut les modéliser dans leur ensemble. Avec de telles attaques, on ne peut prédire quels symboles seront extraits : ce seront tantôt des symboles correspondant aux copies des membres de la coalition, tantôt des symboles différents, tantôt des effacements. Cela dépendra beaucoup de la technique de tatouage employée. On se retrouve alors dans le modèle étendu de Boneh et Shaw proposé par Guth *et al.*, soit (a)+(b′), autorisant tous les symboles de l'alphabet (et donc des erreurs), ainsi que des effacements. Mais là encore la *marking assumption* (a) est respectée. En effet, si les blocs sont tous identiques, alors le bloc forgé leur sera identique, ce qui se traduit au niveau des identifiants par : si $X_{ji} = a$ pour tout $j \in \mathcal{C}$, alors $Y_i = a$.

La troisième catégorie d'attaques, illustrée par la figure 3.7, n'est pas spécifique à la constitution d'une coalition. Elle s'apparente à un « lessivage » de la marque à l'aide de techniques compression avec perte, de débruitage, etc, comme on en

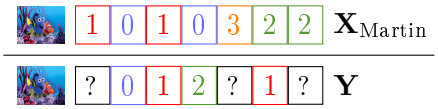


FIGURE 3.7: Attaques de type « lessivage ». On ne peut pas prédire quels symboles seront extraits de la copie pirate, cela dépendra de la technique de tatouage utilisée. On est donc dans le modèle (b′).

considère habituellement pour ébranler la robustesse des techniques de tatouage. On ne peut pas prédire quels symboles seront extraits de la copie pirate, cela dépendra de la technique de tatouage utilisée.

Quelles contraintes pour la technique de tatouage ? Il est clair que dans un contexte d'identification de copies l'identifiant doit être inséré à l'aide d'une technique de tatouage robuste. C'est cette forte robustesse qui sera garante de l'extraction d'un identifiant exploitable pour remonter la piste des membres de la coalition. Cette question est abordée de manière heuristique dans toutes les publications de schémas complets, tels ceux mentionnés page 60 [WWZ+03, WWZ+05, TWWL03, ZWWL03, ZWWL05, HW06, HW07, Sch08a, Sch08b, XFF08]. La question n'a pour l'instant été abordée formellement que dans [MBCM10b, MBCM10a], qui étudient les contraintes imposées sur la technique d'insertion par l'utilisation des codes de Tardos, en termes de robustesse mais aussi de sécurité.

La capacité d'insertion de la technique de tatouage peut, par contre, ne pas être déterminante, pour peu que l'on s'en accomode. L'utilisation de la technique zéro-bit `Broken Arrows` mentionnée à la section 3.2 en est un exemple.

On notera également qu'il n'est pas nécessaire ici que la détection/extraction soit aveugle, la présence du document original lors de la phase d'extraction de l'identifiant étant tout-à-fait envisageable, puisqu'elle sera probablement effectuée par son propriétaire ou un tiers de confiance, et qu'il n'y aura de plus en général pas d'ambiguïté sur l'identification de l'original. Si on cherche à remonter la piste de la distribution illégale d'un film au box office, ce film est identifié sans contest, et son propriétaire probablement enclain à mettre l'original à la disposition de l'entité de confiance qui procèdera à l'extraction.

Les vitesses de tatouage et d'extraction ne sont pas non plus essentielles, l'insertion étant réalisées hors-ligne, et la détection étant opérée de manière ponctuelle.

Comment intégrer ces techniques dans un protocole complet de distribution de documents ? Dans une utilisation naïve d'un système de fingerprinting, le fournisseur de contenus se charge de générer lui-même les identifiants des utilisateurs au fur et à mesure que ces derniers effectuent des requêtes pour obtenir des documents. Mais dans ce cas, puisque le fournisseur connaît les identifiants de tous les utilisateurs, il peut décider de distribuer lui-même la copie d'un de ces utilisateurs de manière illégale, pour ensuite accuser ce dernier. Il peut également, comme

nous l'avons montré dans [CFFC11] pour les codes de Tardos, s'arranger pour biaiser l'accusation afin d'accuser un utilisateur de son choix. Par ailleurs, même si le fournisseur de contenu reste honnête, son honnêteté n'est pas garantie. Ainsi, tout utilisateur accusé peut se prétendre victime d'un fournisseur de contenu malhonnête.

Pour empêcher ceci, on intègre les primitives du système dans des protocoles de fingerprinting qualifiés d'*asymétriques*[4]. Dans ces protocoles, les utilisateurs participent à la génération de leurs identifiants, parfois même les génèrent intégralement eux-même, et le fournisseur ne connaît aucun identifiant dans son intégralité. Le fournisseur délivre alors par un moyen adéquat les morceaux de documents tatoués avec les bons symboles, sans pour autant connaître leur valeur exacte ! Ces protocoles asymétriques ont été introduits en 1996 par Pfitzmann *et al.* [PS96]. Les publications se sont alors rapidement succédées [PW97b, PW97a, Bie97], pour ensuite se focaliser sur un cas particulier des protocoles asymétriques, les protocoles asymétriques *anonymes*, introduits dans [PW97a], qui assurent de surcroît l'anonymat de l'utilisateur pour le fournisseur de contenus lors de la transaction, *e.g.* [DF99, Cam00, BPW07, HYF08, RDB+10]. Pour ce faire, ces protocoles s'appuient sur un tiers de confiance, qui lui connaît les liens entre les identités véritables des utilisateurs et leurs versions anonymisées. Une étude plus récente a également introduit la notion de protection de la vie privée dans les protocoles de fingerprinting asymétriques anonymes, pour non seulement masquer l'identité de l'utilisateur au moment de la transaction, mais également masquer la référence des objets qu'il a acquis auprès du fournisseur [AGC10, Abd11].

Ces protocoles asymétriques, élaborés par des cryptographes, ont longtemps fait abstraction des techniques d'insertion, parfois même des codes anti-collusions eux-mêmes, ces derniers étant considérés comme des boîtes noires qui remplissent correctement leur fonction. Ceci s'explique par le fait que lorsque ces protocoles ont été introduits, les techniques d'insertion et les codes anti-collusion n'offraient pas suffisamment de maturité pour qu'on les utilise en pratique. Aujourd'hui, leur maturité est suffisante, et on peut donc maintenant envisager la conception de tels protocoles avec une vision plus concrète et un soucis d'application réel.

D'autres publications ont poursuivi en parallèle le même objectif de prévenir tout comportement malicieux du fournisseur de contenus, en proposant des protocoles dits *Buyer-Seller*, *e.g.* [MW01,LYTC04,DBPP09,Kur10]. Ils se sont principalement focalisés sur les moyens d'insertion permettant de s'assurer que le fournisseur de contenu n'a jamais accès au contenu personnalisé délivré à l'utilisateur. Ces travaux sont donc finalement complémentaires des précédents, et la maturité des techniques développées dans chacun de ces courants permet aujourd'hui de les réunir pour proposer des protocoles complets, comme nous l'avons fait avec Ana Charpentier, Teddy Furon et Ingemar Cox en proposant un protocole asymétrique entièrement spécifié utilisant les codes de Tardos [CFFC11]. Ces travaux sont présentés au paragraphe 3.4.

Si on se limite à la notion d'asymétrie sans se préoccuper de l'anonymat, la conception de tels protocoles se heurte à trois étapes délicates.

4. Attention, cette terminologie n'a rien à voir avec la notion de schéma asymétrique en cryptographie, stéganographie ou tatouage.

1. L'utilisateur génère son identifiant, avec l'aide du fournisseur. Il y a alors deux points délicats : l'utilisateur doit générer son identifiant suivant les recommandations du fournisseur, et parfois même certains paramètres secrets connus seulement du fournisseur, sans tricher ; le fournisseur doit ensuite pouvoir s'assurer que l'utilisateur a suivi les recommandations, sans pour autant connaître l'identifiant dans son intégralité.
2. Le fournisseur doit fournir à l'utilisateur le contenu demandé, tatoué avec son identifiant. Le point délicat est que l'utilisateur ne doit pas rentrer en possession du contenu non tatoué, et le fournisseur ne connaît pas l'identifiant en entier.
3. Lorsque le fournisseur suspecte une fraude, il procède au traçage pour retrouver au moins un des coupables. Le point délicat est ici de s'assurer que le fournisseur ne triche pas pour biaiser le processus et accuser un innocent.

Pour résoudre ces problèmes délicats, les protocoles de fingerprinting asymétriques s'appuient sur des primitives cryptographiques, comme par exemple des protocoles de *transfert inconscient* (*oblivious transfer*), des *protocoles à divulgation nulle de connaissance* (*zero-knowledge protocols*), des techniques de PIR (*private information retrieval*), des techniques de *mise en gage* (*commitment*), des schémas de *signature*, ou encore des systèmes de *chiffrement homomorphiques*. Des techniques de *tatouage homomorphiques*, s'appuyant sur des systèmes de chiffrement homomorphiques, sont aussi très utiles. Elles permettent au fournisseur d'insérer un identifiant chiffré (dont il ne connaît pas le clair) dans un contenu chiffré ; il peut alors envoyer le document chiffré tatoué à l'utilisateur, qui le déchiffre pour obtenir le document en clair, tatoué avec son identifiant (en clair également) [DBPP09,Kur10].

La sécurité des protocoles ainsi élaborés est évaluée soit de manière heuristique par une énumération des attaques envisagées et une liste d'arguments montrant comment ces attaques sont contrées, soit formellement grâce à des preuves de sécurité menées dans des modèles de sécurité. Cette approche formelle s'appuie sur des modèles élaborés en cryptographie pour les preuves de protocoles, notamment en *multi-party computation*. Ces modèles spécifient la capacité de calcul des parties, ainsi que les comportements autorisés (*semi-honnête* : les entités suivent le protocole sans tricher, mais conservent des enregistrements des échanges réalisés ; *malicieux* : les entités peuvent ne pas suivre le protocole). Pour chaque modèle, on peut s'appuyer sur une vision idéale, plus simple à manipuler dans les preuves, ou une version plus réaliste, plus difficile à manipuler, mais plus proche de ce qui peut se passer dans la réalité.

Fingerprinting* et *Group Testing La problématique du *group testing* est d'identifier dans une large population quelles personnes sont, par exemple, infectées par un virus, mais en effectuant peu de tests. On effectue par exemple des prélèvements sanguins pour toute la population, prélèvements que l'on mélange ensuite par groupes, pour n'effectuer qu'un test par groupe.

Les liens entre la problématique du *fingerprinting* et celle du *group testing* ont été étudiés pour la première fois dans [SvTW00]. Kitagawa *et al.* ont ensuite montré comment des techniques de *group testing* permettent de construire des codes anti-collusion extrêmement courts lorsque la coalition ne comprend pas plus de 3

utilisateurs [KH07, KHN+08]. Plus récemment, Meerwald et Furon ont comparé les performances, dans un contexte de *group testing*, des solutions provenant des deux domaines. Ils ont ainsi montré que les codes de Tardos sont au moins aussi performants que les techniques utilisées habituellement en *group testing* [MF11b].

3.1.3 Codes de Tardos

L'idée de G. Tardos [Tar03, Tar08] a été d'utiliser comme mots de code des vecteurs binaires tirés aléatoirement suivant une certaine loi, mi-dense mi-creuse, générée aléatoirement et tenue secrète par le concepteur du système. Lorsqu'un vecteur pirate est récupéré, on peut calculer pour chaque utilisateur référencé un score qui tend à l'accuser ou au contraire à l'innocenter. Les paramètres du système garantissent que la fiabilité de la procédure d'accusation est indépendante de la stratégie avec laquelle la copie pirate a été constituée (tant que la *marking assumption* (a) est respectée). Le calcul de ce score a été amélioré et généralisé au cas d'alphabets non binaires par Škorić *et al.* en 2008 [ŠKC08]. C'est cette version qui sert de base aux nombreux travaux qui ont été menés depuis sur ces codes dits *de Tardos*. Notons que le terme de code peut donner lieu à confusion, car ce code n'est pas un code correcteur, et ne dispose pas d'une procédure d'encodage.

Mes travaux ayant porté sur les versions binaires et q-aires, elles sont toutes deux brièvement présentées ici, telles que présentées dans [ŠKC08]. Soient ε_1, ε et ε_2 les probabilités d'erreur maximum autorisées, désignant respectivement ε_1 la probabilité qu'un utilisateur innocent donné soit accusé, ε la probabilité que globalement on accuse un innocent, et ε_2 la probabilité de ne trouver aucun coupable. Les autres paramètres fixés par le concepteur du systèmes sont le nombre d'utilisateurs n et la taille maximum c pour une coalition. Ces paramètres déterminent directement la longueur m du code ainsi que le seuil Z qui permettra de déterminer la culpabilité ou l'innocence des utilisateurs en fonction de leur score. Rappelons que la borne inférieure sur la longueur m donnée par Peikert *et al.* [PSS03] et Tardos [Tar03] est en $O(c^2 \ln(n/\varepsilon))$. Comme $\varepsilon = 1 - (1 - \varepsilon_1)^{n-c}$, et qu'en pratique $\varepsilon_1 << 1$ et $c << n$, on peut considérer que $\varepsilon \simeq n\varepsilon_1$ et la borne inférieure sur la longueur est alors en $O(c^2 \ln(1/\varepsilon_1))$. Certains articles utilisent la formule faisant intervenir ε_1, d'autres celle qui s'appuie sur ε. Les codes de Tardos ont une longueur de la forme $m = Ac^2\lceil \ln(1/\varepsilon_1)\rceil$ et un seuil de la forme $Z = Bc\lceil \ln(1/\varepsilon_1)\rceil$, avec A et B des nombres réels. A titre d'exemple, Škorić *et al.* ont proposé dans [ŠVCT08] des valeurs de $A = 2\pi^2$ et $B = 2\pi$ qui garantissent dans le cas binaire que l'accusation de tous les utilisateurs dont le score sera supérieur à Z n'entraînera pas un taux d'erreur supérieur aux ε_1 et ε_2 spécifiés, sous les seules conditions que les probabilités d'erreur $\varepsilon_1 << \varepsilon_2$ indépendantes, et que la *marking assumption* (a) est respectée. On peut encore réduire la longueur (en adaptant le seuil) sous certaines conditions, notamment dans le cas q-aire, comme cela est mentionné dans le paragraphe « Dans les pas de Tardos ».

Les codes de Tardos, dont le cas binaire est illustré par la figure 3.8, fonctionnent alors ainsi. Les n identifiants distribués vont former une matrice $n \times m$ $\mathbf{X}$. L'utilisateur j, ou la transaction j selon les scénarios, sera identifié(e) par le mot

$\mathbf{X}_j = (X_{j1}, X_{j2}, \ldots, X_{jm})$, qui sera inséré dans le document correspondant. Avant de générer cette matrice, un vecteur $\mathbf{p}$ est tiré au hasard une fois pour toute, qui décrit la loi que suivra chacune des m composantes des vecteurs $\mathbf{X}_j$:

`[binaire]`	`[q-aire]`
$\mathbf{p} = (p_1, \ldots, p_m)$ chaque $p_i \in [0,1]$ est tiré indépendamment suivant la fonction de densité $f(p) = \frac{1}{\pi\sqrt{p(1-p)}}$	$\mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_m)$ chaque $\mathbf{p}_i = (p_i^0, \ldots, p_i^{q-1})$ est tiré indépendamment suivant un cas particulier d'une distribution de Dirichlet, de paramètre de forme κ

Chaque élément de la matrice $\mathbf{X}$ est ensuite tiré indépendamment :

`[binaire]`	`[q-aire]`
$\mathbb{P}(X_{ji} = 1) = p_i, \quad \mathbb{P}(X_{ji} = 0) = 1 - p_i$	$\mathbb{P}(X_{ji} = a) = p_i^a$

Chacun des mots X_j est caché dans la copie délivrée à l'utilisateur qui lui est associé. Lors de la phase d'accusation, on extrait la séquence $\mathbf{Y}$ de la copie pirate. Afin de savoir si l'utilisateur (ou la transaction) j est impliqué(e) dans la production de la copie pirate, on calcule un score d'accusation (ici donné dans sa version générique ; la version binaire, beaucoup plus simple, est donnée dans la figure 3.8)

$$S_j = \sum_{i=1}^{m} g(Y_i, X_{ji}, \mathbf{p}_i)$$

en prenant $g(Y_i, X_{ji}, \mathbf{p}_i) = \delta_{Y_i = X_{ji}} g_1(p_i^{Y_i}) + (1 - \delta_{Y_i \neq X_{ji}}) g_0(p_i^{Y_i})$
avec les fonctions d'accusation $g_1(p) = \sqrt{(1-p)/p}$ et $g_0(p) = -\sqrt{p/(1-p)}$,
et en notant abusivement $p^1 = p$ et $p^0 = 1 - p$ dans le cas binaire.

Si ce score est supérieur au seuil Z, alors on considère l'utilisateur j comme coupable. Ces fonctions d'accusation permettent de considérer, composante après composante, qu'une similitude (resp. une différence) entre le mot extrait de la copie pirate et celui d'un utilisateur donné tend à accuser (resp. innocenter) ce dernier, mais cette information est pondérée par la rareté d'avoir cette valeur précise pour cette composante. C'est là que la distribution $\mathbf{p}$ joue son rôle. Les fonctions d'accusation ont été définies de manière à ce que la distribution des scores ne dépende pas de la stratégie de la coalition dans la construction de la copie pirate, tant que la *marking assumption* (a) est respectée. Les scores des innocents sont décorrélés et centrés en 0. Les scores des coupables sont, eux, centrés en une valeur nettement plus élevée, $2\pi m/c$.

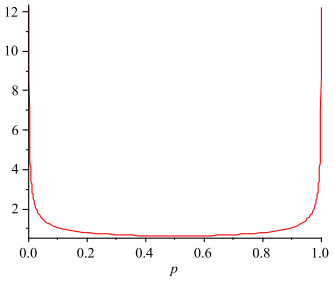
Pourquoi ces choix ? Les performances des codes de Tardos sont liées à l'allure des distributions des scores des coupables et des innocents. Le choix de chaque paramètre du code (longueur, seuil, fonction de densité, fonctions d'accusation, etc) a un impact sur ces distributions, et on peut chercher pour optimiser le code à minimiser sa longueur, mais aussi à augmenter la moyenne du score des coupables, ou encore à séparer encore mieux les distributions des coupables et des innocents. Il est important de noter que Škorić *et al.* ont montré dans [ŠKC08] que les scores des innocents (respectivement des coupables) suivent une distribution gaussienne dès que c est suffisamment grand. Dans la pratique, on le constate en effet, y compris pour

Paramètres : $c << n$, $\varepsilon_1 << \varepsilon_2$
Longueur et seuil : *e.g.* $m = 2\pi^2 c^2 \lceil \ln(1/\varepsilon_1) \rceil$, $Z = 2\pi c \lceil \ln(1/\varepsilon_1) \rceil$ [ŠVCT08]

/* Génération des identifiants : */
m probabilités secrètes p_i sont tirées indépendamment suivant la fonction de densité $f(p) = \frac{1}{\pi\sqrt{p(1-p)}}$, représentée ci-dessous à droite.

	p_1	p_2	p_3	$\dots$	p_m
$\mathbf{X}_1$	1	0	1	$\dots$	0
$\mathbf{X}_2$	0	1	0	$\dots$	1
$\mathbf{X}_3$	1	1	0	$\dots$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathbf{X}_n$	0	0	0	$\dots$	1

avec $\mathbb{P}(X_{ij} = 1) = p_i \ \forall i \ \forall j$



/* Accusation : */
pour chaque utilisateur j suspecté on calcul un score S_j et on le compare au seuil Z : $S_j = \Sigma_{i=1}^{m} g(Y_i, X_{ji}, p_i) >^{?} Z$

$\mathbf{Y}$	1	1	0	$\dots$	1
$\mathbf{X}_j$	1	0	1	$\dots$	0

$S_j = g(1,1,p_1) + g(1,0,p_2) + g(0,1,p_3) + \ldots + g(1,0,p_m)$

$$g(1,1,p) = g(0,0,1-p) = +\sqrt{(1-p)/p}$$
$$g(1,0,p) = g(0,1,1-p) = -\sqrt{p/(1-p)}$$

Si on trouve dans l'identifiant extrait de la copie pirate le même symbole que dans l'identifiant de l'utilisation suspecté, alors on tend à l'accuser en faisant augmenter son score : de peu si ce symbole est fréquent à cette position (l'information est peu pertinente), beaucoup s'il est rare (l'information est beaucoup plus pertinente). De même, si le symbole extrait de la copie pirate est différent de celui de d'indentifiant de l'utilisateur, alors on tend à l'innocenter en faisant diminuer son score : de peu ou de beaucoup, selon la fréquence de ce dernier dans les identifiants des utilisateurs en général.

FIGURE 3.8: Résumé du fonctionnement des codes de Tardos binaires tels que présentés dans [ŠKC08].

des valeurs de c relativement petites, comme l'illustre la figure 3.9. Ceci nous permet de considérer dans tous les raisonnements ces scores comme suivant des distributions gaussiennes.

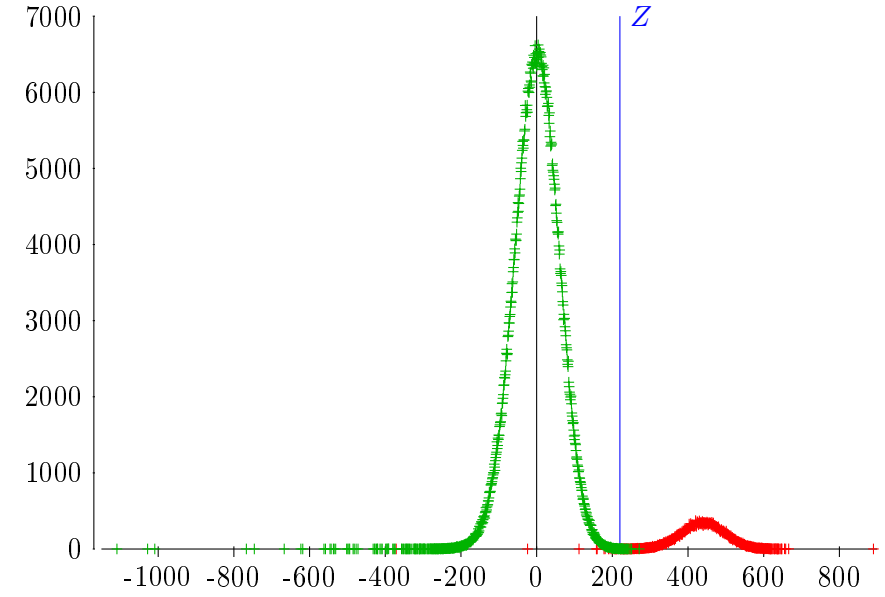
Les fonctions de densité et d'accusation initialement proposées par Tardos [Tar03, Tar08] (qui diffèrent de celles présentées ci-dessus car Tardos avait pris $g(0,*,*)=0$), puis celles proposées par Škorić *et al.* [ŠKC08] assurent que les performances de l'accusation sont constantes, indépendantes de la stratégie d'attaque de la coalition (tant que la *marking assumption* (a) est respectée). Tardos a prouvé que le code fonctionne bien, mais n'a pas apporté de justification quant au choix des fonctions, ni de certaines constantes (intervenant dans la détermination de la longueur, du seuil, et du *cut-off* que je n'ai pas mentionné ici mais qui permet de gérer la limite infinie de $f(p)$ près de 0 et de 1). Blayer *et al.* ont par la suite analysé en détail le choix des constantes [BT08], tandis que Furon *et al.* se sont penchés sur le choix des fonctions [FGC08].

La stratégie de la collusion n'étant pas connue à l'accusation, c'est un paramètre de nuisance. L'idée de Tardos a été de créer un test reposant sur une quantité pivot indépendante du paramètre de nuisance. Les fonctions choisies assurent cette indépendance uniquement pour les moments d'ordre 1 et 2. Grâce à la borne de Markov-Chebyshev, les probabilités d'accuser à tort et de rater un coupable sont majorées et inférieures aux niveaux donnés dans le cahier des charges pour une longueur de code $m = O(c^2 \log(n/\epsilon))$. Furon *et al.* ont tout d'abord montré dans le cas binaire que si l'on cherche à assurer l'indépendance de la distribution des scores par rapport à la stratégie de la coalition, alors on obtient les relations suivantes :

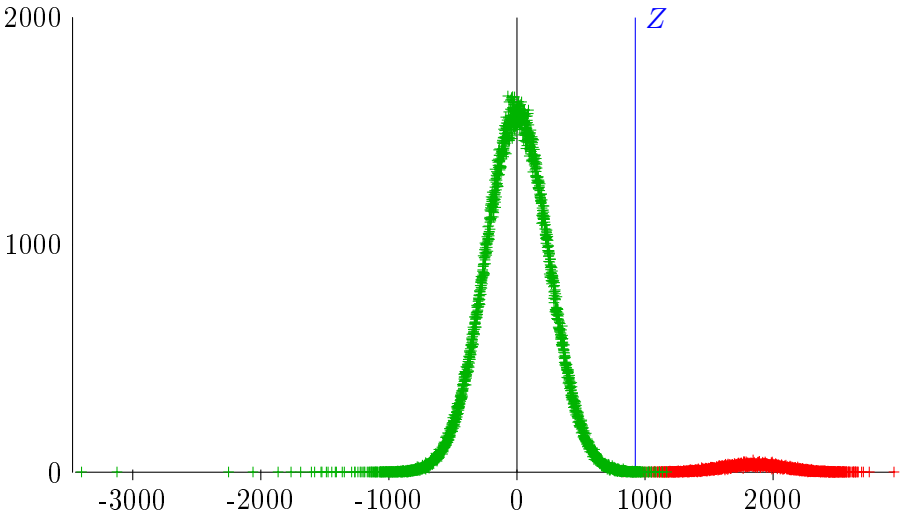
$$\begin{array}{ccccc} p & g(1,0,1-p) & = & -(1-p) & g(1,1,1-p) \\ (1-p) & g(0,0,p) & = & p & g(1,1,p) \end{array}$$

La fonction $g(1,1,p)$ détermine donc toutes les autres. Furon *et al.* ont alors montré qu'à variance contrainte, la fonction qui donne la plus grande espérance au score des coupables est précisément $g(1,1,p) = \sqrt{(1-p)/p}$, et que la fonction de densité optimale est bien $f(p) = 1/(\pi\sqrt{p(1-p)})$ [FGC08].

Ils se sont ensuite posé la question de l'optimisation des fonctions $g(*,*,p)$, toujours dans le cas binaire, pour maximiser l'espérance des scores des coupables lorsque l'on connaît, au moment du traçage, la taille de la coalition et la stratégie qu'elle a utilisée pour produire la copie pirate. Sous l'hypothèse que la stratégie utilisée par la coalition pour produire la copie pirate est la même pour toutes les composantes de l'identifiant (ou de manière équivalente pour tous les blocs), ils ont alors obtenu des fonctions différentes, effectivement plus efficaces que celles que Tardos et Škorić *et al.*. Avec Ana Charpentier et Teddy Furon, nous avons rebondi sur ces travaux pour pousser l'optimisation encore plus loin, et proposer en même temps une estimation dynamique de la stratégie [CXFF09, CFF10]. Ces travaux sont présentés dans le paragraphe 3.3. Ainsi, l'accusation peut s'opérer sans *a priori* de la part des personnes sur la stratégie utilisée, c'est l'estimateur qui la fournit et alimente l'optimisation des fonctions d'accusation. Un meilleur estimateur de stratégie a ensuite été proposé par Furon et Pérez-Freire dans [FPF09a]. Le cas q-aire est à ce jour non



(a) $n = 100, c = 5, 10000$ expériences



(b) $n = 1000, c = 21, 1000$ expériences

FIGURE 3.9: Distribution des scores des utilisateurs (coupables en rouge, innocents en vert) pour un code de Tardos binaire tel que présenté dans [ŠKC08] avec les paramètres $m = 2\pi^2 c^2 \lceil \ln(1/\varepsilon_1) \rceil$ et $Z = 2\pi c \lceil \ln(1/\varepsilon_1) \rceil$ [ŠVCT08]. Les expériences sont menées avec $\varepsilon_1 = 10^{-3}$, et des attaques de type « Echange de blocs ». Les scores sont distribués suivant des gaussiennes.

traité dans la littérature. Nous nous y sommes essayés, mais sans parvenir au bout de l'optimisation.

Avantages et inconvénients. Les codes de Tardos offrent de très nombreux avantages. Ils sont simples à comprendre, faciles à implémenter, flexibles et efficaces. Les identifiants peuvent être générés à la demande, et le nombre total d'utilisateur n'a pas besoin d'être fixé strictement au départ. S'il s'avère que l'on dépasse le nombre d'utlisateurs envisagés initialement, on peut continuer à générer des identifiants sans remettre tout le système en question (dans cette situation, si on utilise des codes anti-collusion issus par exemple de codes correcteurs, on est obligés de générer un nouveau code) ; la probabilité d'accuser un utilisateur à tort augmentera alors légèrement, mais pas dramatiquement. Le traçage est très peu coûteux. Par ailleurs, si l'on applique un traçage avec seuil, on peut en cas de suspicions se contenter de calculer les scores des suspects, sans calculer ceux de tous les utilisateurs.

Ils peuvent être appliqués dans des contextes réalistes, non seulement par leur faible complexité, mais également parce qu'ils reposent sur des hypothèses réalistes. La seule contrainte pour qu'ils se comportent bien est en effet que la *marking assumption* (a) soit respectée. Ceci nous place dans le modèle étendu réaliste (a)+(b′), qui couvre l'ensemble des attaques respectant la *marking assumption*. Comme nous l'avons vu, on peut s'arranger grâce au système de découpage en blocs et aiguillage pour que ce modèle soit effectif. Ces codes résistent par ailleurs très bien aux erreurs et effacements. Ceci est dû au fait que chaque score est calculé en sommant composante après composante les « corrélations » entre le mot extrait de la copie pirate et l'identifiant de l'utilisateur dont on cherche à établir l'innocence ou la culpabilité. Ces composantes étant indépendantes, les effacements ont pour effet de supprimer certains termes dans la somme (le seuil étant bien sûr adapté en conséquence), rendant l'interprétation du score légèrement moins fiable. Les erreurs ont, elles, pour effet de modifier certains termes de la somme. Si on considère les fonctions d'accusation génériques de Tardos ou de Škorić *et al.*, conçues pour que la distribution des scores ne dépende pas de la stratégie de la coalition, alors ces erreurs n'ont aucun impact sur la distribution, tant qu'elles ne mettent pas en défaut la *marking assumption* (a). Si elles la mettent en défaut, alors la fiabilité se dégrade d'autant que le nombre de composantes où elle devrait être vérifiée et où elle ne l'est plus est grand.

Leur seul véritable inconvénient qu'ils peuvent présenter, par rapport à certains codes anti-collusions reposant sur des codes correcteurs d'erreur (*e.g.* **??**), est la complexité que peut représenter la phase d'accusation, qui est ici linéaire en le nombre d'utilisateurs, et donc exponentielle en la longueur des identifiants (alors que celle de **??** est polynomiale en la longueur des identifiants). Elle nécessite en effet de calculer un score pour chaque utilisateur que l'on souhaite tester, donc dans le pire des cas de tous les utilisateurs (par exemple si on s'appuie sur le seuil et que l'on n'a aucun utilisateur suspect *a priori*, ou encore si l'on souhaite effectuer un classement complet des scores pour accuser le plus grand). Néanmoins, en moyenne, et surtout dans le cas où certains utilisateurs sont suspectés, cet inconvénient s'efface devant les très nombreux avantages que présentent ces codes.

Dans les pas de Tardos. Plus d'une cinquantaine de publications ont proposé des études et améliorations de ces codes depuis 2008. J'en mentionne ici quelques unes, qui illustrent la plupart des questions abordées et des approches suivies. J'attire l'attention sur le fait que toutes ne reposent pas exactement sur les mêmes hypothèses. Ceci est du non seulement aux différentes variantes des modèles d'attaques que l'on peut considérer et explorer, mais aussi au grand nombre de paramètres du code sur lesquels on peut jouer, que l'on peut contraindre ou laisser libre. Ces différences ne sont pas mentionnées ici, car elles nécessiteraient de poser les choses beaucoup plus formellement. Elles ne nuisent cependant pas à une présentation synthétique et allégée de ces approches et résultats.

Škorić *et al.* a tout d'abord constaté qu'en relâchant certaines hypothèses sur les paramètres (notamment sur les valeurs des probabilités d'erreur) et en autorisant certaines constantes du schéma initial à varier, on peut réduire la longueur du code [ŠVCT08], de $m = 100c^2\lceil\ln(1/\varepsilon_1)\rceil$ [Tar03, Tar08] à $m = 4\pi^2c^2\lceil\ln(1/\varepsilon_1)\rceil$, et même à $m = 2\pi^2c^2\lceil\ln(1/\varepsilon_1)\rceil$ sous l'hypothèse que les scores suivent des distributions gaussiennes. Ils ont dans le même article montré que les distributions des scores peuvent être considérées comme gaussiennes dès que c dépasse 20. Dans [ŠKC08], ils ont proposé une variante des codes de Tardos qui est aujourd'hui celle de référence, avec un calcul de score symétrique. Poussant alors la réduction de la longueur du code encore un peu plus loin pour cette version symétrique, ils arrivent dans le cas binaire à $m = \pi^2c^2\lceil\ln(1/\varepsilon_1)\rceil$ lorsque c est grand. Ils introduisent par ailleurs dans cet article une version q-aire de ces codes de Tardos avec calcul de score symétrique, et montrent qu'avec cette version q-aire on peut encore réduire la longueur du code.

Nuida *et al.* ont de leur côté étudié comment on peut réduire la longueur du code dans le cas de petites coalitions [HHI06, NHWI07, NFH$^+$07] ; ils montrent également comment modifier légèrement la fonction de densité f pour diminuer la quantité de mémoire nécessaire lors de l'implémentation.

Plusieurs travaux se sont ensuite penchés sur l'amélioration du traçage, parfois appelé aussi *décodage*. Par l'optimisation des fonctions d'accusation [FGC08, CXFF09, CFF10] et l'estimation de la stratégie de la coalition [CXFF09, CFF10, FPF09a]. En n'accusant que le plus grand score [NFH$^+$09]. Ou encore par l'utilisation théorique [HM08, AT09] puis pratique [Nui10, MF11a] de *joint decoders*. Cette dernière série d'études s'appuie sur les travaux visant à optimiser la capacité des schémas, ce qui revient à minimiser la longueur. Ces travaux reposent sur la modélisation en termes de *jeu minmax* du bras de fer du concepteur, qui essaie de maximiser son aptitude à analyser les traces laissées par les attaquants, pendant que ces derniers tentent de les minimise. On peut ainsi chercher à optimiser les paramètres et fonctions internes du schéma (ici $f()$, $g()$, etc) pour estimer ce que le concepteur peut espérer de mieux, lorsque les attaquants opèrent de manière optimale, *e.g.* [HM08], ce qui permet de mieux cerner quelle peut être la capacité optimale, et comment l'atteindre [AT09, HM09, HM10, SŠ11].

Améliorer ainsi le décodage permet de gérer, à longueur fixée des coalitions plus grandes, ou bien à coalitions de taille fixée d'utiliser des codes plus courts. C'est donc également une manière de raccourcir le code.

Une quatrième approche a donné lieu à une diminution de la longueur du code.

S'appuyant sur un algorithme efficace permettant de mesurer expérimentalement la probabilité d'événements rares [CFG08], et sur l'estimation de la pire attaque possible [FPF09b], Furon *et al.* ont proposé une estimation expérimentale de la longueur minimale requise pour garantir les probabilités d'erreur initialement fixées, y compris dans le pire cas. Les longueurs expérimentales obtenues sont nettement plus courtes que les longueurs théoriques [FPFGC09].

Certains articles, enfin, ciblent des applications spécifiques [KvCS07,BHP08], ou étudient le couplage de ces codes avec des techniques d'insertion [XFF08,MBCM10b, MBCM10a,DLGP11]. Cela débouche parfois sur des améliorations telles que la prise en compte de l'extraction par la technique de tatouage de plusieurs symboles pour une même composante de l'identifiant [XFF08,ŠKSC09,ŠKSC11].

3.2 Association (et amélioration) des codes de Tardos avec le schéma de tatouage `Broken Arrows` pour contrer les attaques de type fusion : thèse de F. Xie (2007-2010)

Avec Fuchun Xie et Teddy Furon, nous avons proposé un schéma de fingerprinting complet, reposant sur les codes de Tardos q-aires et la technique d'insertion zéro-bit `Broken Arrows` [FB08] mentionnée au chapitre précédent. Notre objectif était non seulement de proposer un schéma complet, mais aussi de montrer comment il peut contrer une des pires attaques, la fusion de documents. Car les codes de Tardos, comme la plupart des codes anti-collusion, sont avant tout conçus pour contrer des mélanges reposant sur l'échange de blocs. Mais si les pirates décident de fusionner leurs blocs, c'est-à-dire d'en produire de nouveaux, par exemple en opérant des moyennes pixel à pixel, le traçage devient vraiment difficile, car on ne peut pas prédire ce que donnera la détection sur le bloc fusionné. Pourtant c'est une attaque très facile à réaliser, et que l'on ne peut pas se permettre d'écarter.

`Broken Arrows` offre l'avantage d'être particulièrement robuste. Cette robustesse s'obtient au détriment de la capacité : dans sa version originelle, elle ne permet pas d'insérer un message complet, mais seulement de dire si une marque est présente ou non. Nous verrons cependant que ceci n'est pas un problème pour nous, voire même que c'est un atout. Nous avons tout d'abord montré comment l'adapter pour insérer un message, en nous appuyant sur l'utilisation de clés différentes pour chaque symbole de l'alphabet à insérer, chaque symbole étant caché dans un bloc différent du document (par exemple une image ou séquence d'images d'une vidéo). L'insertion est alors si robuste, que si on procède à la fusion d'images contenant des symboles différents, on est à même de retrouver lors de la détection plusieurs de ces symboles, en testant les clés associées aux symboles de l'alphabet. Pour chaque composante i, on note L_i le nombre de symboles détectés $\mathcal{Y}_i = \{Y_i^1, \ldots, Y_i^{L_i}\}$. Nous disposons ainsi d'une information plus riche qu'avec les techniques classiques qui extraient un seul symbole.

Reste ensuite à exploiter cette information au mieux. Les codes de Tardos n'ont

pas été conçus pour prendre en compte cette détection multiple. Nous avons proposé deux directions dans la prise en compte de ces détections multiples au niveau du calcul du score.

Rappelons que la fonction du calcul de score dans le cas q-aire est habituellement

$$S_j = \sum_{i=1}^{m} g(Y_i, X_{ji}, \mathbf{p}_i)$$
en prenant $g(Y_i, X_{ji}, \mathbf{p}_i) = \delta_{Y_i = X_{ji}} g_1(p_i^{Y_i}) + (1 - \delta_{Y_i \neq X_{ji}}) g_0(p_i^{Y_i})$
avec les fonctions d'accusation $g_1(p) = \sqrt{(1-p)/p}$ et $g_0(p) = -\sqrt{p/(1-p)}$.

Notre première proposition a été de conserver la fonction g et de modifier le calcul du score :

$$S_j = \sum_{i=1}^{m} \sum_{\ell=1}^{L_i} g(Y_i^{\ell}, X_{ji}, \mathbf{p}_i). \tag{3.1}$$

C'est un peu comme si la longueur du code était augmentée de m à $m\bar{L} = \sum_{i=1}^{m} L_i$, ce qui rend l'accusation plus fiable.

Notre deuxième proposition a été de conserver le calcul du score, en écrivant

$$S_j = \sum_{i=1}^{m} g(\mathcal{Y}_i, X_{ji}, \mathbf{p}_i),$$

et de modifier la fonction g :

$$g(\mathcal{Y}_i, X_{ji}, \mathbf{p}_i) = \delta_{X_{ji} \in \mathcal{Y}_i} g_1(p_i^{\mathcal{Y}_i}) + (1 - \delta_{X_{ji} \notin \mathcal{Y}_i}) g_0(p_i^{\mathcal{Y}_i}), \tag{3.2}$$

avec $p_i^{\mathcal{Y}_i} = \sum_{\ell=1}^{L_i} p_i^{Y_i^{\ell}}$. On a alors l'avantage de diminuer la variance des scores des coupables : quelque soit leur symbole $X_{ji} \in \mathcal{Y}_i$, ils reçoivent la même pénalisation $g_1(p_{\mathcal{Y}_i})$.

Nous avons étudié ces deux propositions expérimentalement, sur des images fixes. Les paramètres utilisés lors des tests sont $m = 300$, $q = 4$ et $c = 20$; le paramètre de forme de la distribution de Dirichlet κ varie quant à lui de 0.1 à 0.5. Nos statistiques sont établies à partir de 32000 scores d'innocents et 8000 scores de coupables. Nous comparons nos résultats face à une fusion par moyennage des pixels à ceux obtenus avec les codes classiques de [ŠKC08] face à un simple échange de blocs. Les résultats montrent que les espérances μ_I des scores des innocents sont nulles pour les trois méthodes. Les espérances μ_C des scores des coupables de nos deux variantes sont assez similaires, et beaucoup plus élevées que celle des codes classiques. Pour les innocents comme pour les coupables, les variances des scores (σ_I^2 et σ_C^2) calculés avec nos variantes sont plus petites qu'avec les codes classiques. Mais la mesure la plus importante est la distance de Kullback-Leibler entre les distributions des scores des innocents et des coupables. Plus elle est élevée, plus les innocents et les coupables sont faciles à distinguer, et donc plus fiable sera le verdict. En effet, à seuil fixé la probabilité d'accuser un utilisateur à tort décroît lorsque la distance de Kullback-Leibler augmente. Nos résultats sont présentés dans la figure 3.10. Nos deux variantes permettent une meilleure distinction des innocents et des coupables.

Ainsi, si les pirates procèdent à une attaque par moyennage, ils sont encore plus efficacement identifiés que s'ils avaient échangé des blocs ! On les incite donc à se rabattre sur cette attaque simple pour laquelle les codes traçant ont été initialement conçus et sont donc naturellement efficaces.

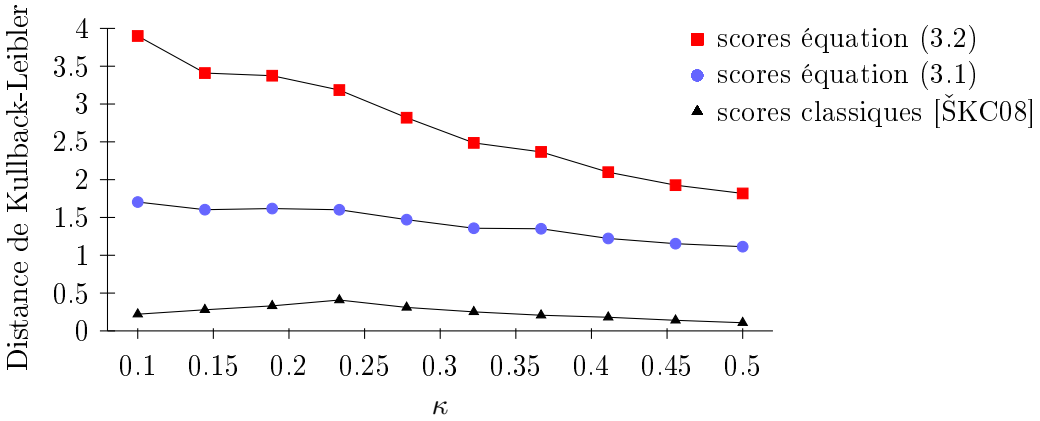


FIGURE 3.10: Distance de Kullback-Leibler entre les distributions des scores des innocents et des coupables, en fonction du paramètre de forme de la distribution de Dirichlet κ pour : l'échange de blocs avec le calcul de scores classique de Škorić *et al.* [ŠKC08], la fusion par moyennage avec le calcul de scores de l'équation (3.1), et la fusion par moyennage avec le calcul de scores de l'équation (3.2). Les paramètres du système sont $m = 300$, $q = 4$ et $c = 20$.

Ces résultats ont été présentés lors de la conférence internationale *ACM Multimedia & Security, MM&SEC'08* [XFF08], ainsi qu'au GRETSI 2009 [XFF09]. Ils ont été valorisés dans le cadre du réseau d'excellence européen ECRYPT, ainsi que du projet ANR-RIAM ESTIVALE.

Ce travail a par ailleurs en quelque sorte montré que le modèle classique de Boneh et Shaw étendu gagne en pratique à être étendu au cas de la détection multiple. Il a inspiré Škorić *et al.* pour pousser plus loin ce raisonnement et aboutir à la formalisation du *Combined Digit Model* [ŠKSC11], modèle d'attaque plus complet, pour lequel ils ont montré que nos améliorations sont performantes.

Ces résultats ont été intégrés dans la plate-forme `FANTOMAS`[5], développée par Mathieu Desoubeaux entre 2007 et 2009. La technique d'insertion `Broken Arrows` a pour l'occasion été adaptée à la vidéo, et la plate-forme permet de personnaliser des films complets, et de rechercher les membres d'une coalition. Nous avons utilisé les résultats de [CFG08] pour évaluer précisément les probabilités d'accuser à tort les utilisateurs de plus haut score. Nous avons pu observer que ces probabilités d'erreur sont largement en-dessous de ε_1 pour les trois plus grands scores (on obtient même

5. `http://www.irisa.fr/temics/demos/BrokenArrows/index.php`

une probabilité d'accuser à tort le plus grand score de l'ordre de 10^{-25}, pour une borne d'erreur $\varepsilon_1 = 10^{-3}$), comme le montre la figure 3.11.

Avec Fuchun Xie et Teddy Furon, nous avons en parallèle mené des travaux complémentaires, présentés dans la section 2.4 et publiés dans [CXFF09, XFF10a, XFF10b], pour améliorer la robustesse et la sécurité de `Broken Arrows`. Afin d'effectuer une synthèse de l'ensemble de ces travaux, Fuchun Xie a combiné dans sa thèse [Xie10] les améliorations de la technique d'insertion avec celles du calcul des scores. La figure 3.12 illustre nos résultats. On voit que l'équation (3.2) donne bien lieu en pratique à une meilleure distinction des innocents et des coupables que l'équation (3.1). On constate également que toutes deux donnent lieu à de meilleurs résultats que le calcul classique des scores. Concernant les techniques d'insertion, la variante `BA-AWC+` (avec les paramètres $N_v = 1024$ et $N_c = 256$) est celle qui donne les meilleurs résultats en termes de traçage, tout en offrant une excellente résistance aux attaques connues (attaques de Westfeld et de Bas [BW09]). Avec la version `BA-AWC++`, nos formules de calcul de scores donnent lieu à un gain beaucoup moins important. Néanmoins, comme nous l'avons souligné dans la section 2.4, le test de décision *a contrario* peut alors nous fournir une probabilité de la présence d'un symbole donné dans le tatouage, et donc nous fournir une information non plus dure (présence ou absence), mais souple, apportant plus d'information pour l'étape d'accusation du code traçant.

3.3 Optimisation dynamique de l'accusation des codes de Tardos : thèse de A. Charpentier (2008-2011)

Comme mentionné dans le paragraphe intitulé « Pourquoi ces choix ? » de la section 3.1.3, Furon *et al.* ont cherché à comprendre les choix effectués par Tardos. Ils ont tout d'abord montré dans le cas binaire que si l'on cherche à assurer l'indépendance de la distribution des scores par rapport à la stratégie de la coalition, alors on obtient les relations suivantes :

$$\begin{array}{cccccc} p & g(1,0,1-p) & = & -(1-p) & g(1,1,1-p) \\ (1-p) & g(0,0,p) & = & p & g(1,1,p) \end{array}$$

La fonction $g(1,1,p)$ détermine donc toutes les autres. Furon *et al.* ont alors montré qu'à variance contrainte, la fonction qui donne la plus grande espérance au score des coupables est précisément $g(1,1,p) = \sqrt{(1-p)/p}$, et que la fonction de densité optimale est bien $f(p) = 1/(\pi\sqrt{p(1-p)})$ [FGC08].

Ils se sont ensuite posé la question de l'optimisation des fonctions $g(*,*,p)$, toujours dans le cas binaire, pour maximiser l'espérance des scores des coupables lorsque l'on connaît, au moment du traçage, la taille de la coalition et la stratégie qu'elle a utilisée pour produire la copie pirate. Sous l'hypothèse que la stratégie utilisée par la coalition pour produire la copie pirate est la même pour toutes les composantes de l'identifiant (ou de manière équivalente pour tous les blocs), ils ont alors obtenu des fonctions différentes, effectivement plus efficaces que celles que Tardos et Škorić *et al.*.

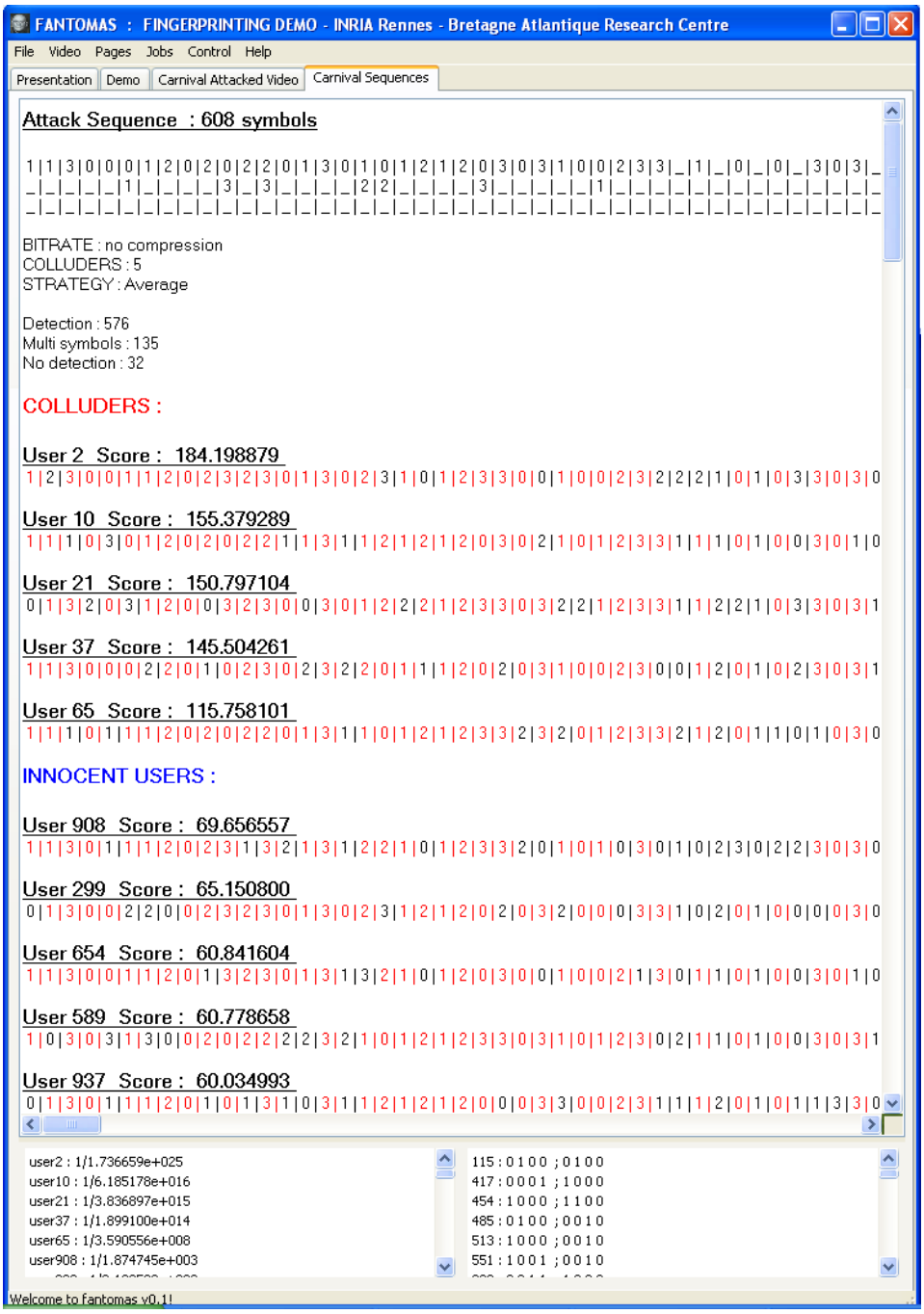


FIGURE 3.11: Plate-forme FANTOMAS : attaque par moyennage (de type fusion). On voit en haut les symboles détectés pour chaque bloc (on détecte parfois plusieurs symboles simultanément). On affiche les scores triés par ordre décroissant. Les cinq coupables arrivent bien en tête, avec des scores nettement supérieurs à ceux des innocents. On trouve, dans la fenêtre en bas à gauche les probabilités précises d'accuser chaque utilisateur à tort [CFG08]. On voit que le risque que l'on prend en accusant le plus grand score est ici de l'ordre de 10^{-25}. Pour le deuxième plus grand score, elle est de 10^{-16}, pour le troisième de 10^{-15}, pour le quatrième de 10^{-14}, et pour le cinquième de 10^{-8}. L'innocent ayant le plus grand score serait, lui, accusé à tort avec une probabililté de 10^{-3}.

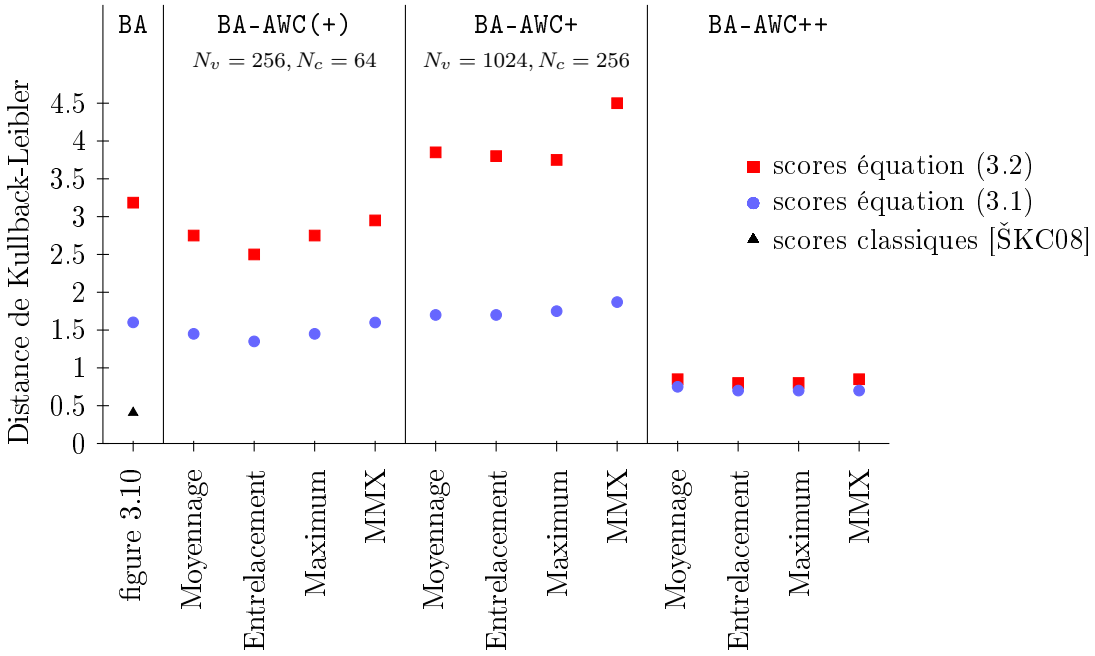


FIGURE 3.12: Evaluation des performances de nos deux fonctions de calcul de scores (3.1) et (3.2), avec les différentes variantes de `Broken Arrows` présentées à la section 2.4. Nous avons choisi pour cette comparaison de prendre $\kappa = 0.23$, car comme le montre la figure 3.10 c'est pour cette valeur que les scores classiques de [ŠKC08] sont les plus performants. Les paramètres du système sont les mêmes que précédemment, soient $m = 300$, $q = 4$ et $c = 20$. On retrouve dans la première colonne les résultats de la figure 3.10, s'appuyant sur le masque originel de `Broken Arrows`, et les paramètres $N_v = 256$ et $N_c = 30$, avec une attaque par échange de blocs pour le calcul classique des scores [ŠKC08], et une fusion par moyennage pour les calculs des scores avec les équations (3.1) et (3.2). Pour les trois variantes de `Broken Arrows`, nous avons testé quatre attaques non-linéaires de type fusion : moyennage des pixels, entrelacement des pixels, maximum des pixels, et enfin l'attaque MMX (*Moderated Minority eXtreme*) due à Schaathun [Sch08a].

Avec Ana Charpentier et Teddy Furon, nous avons continué dans cette direction, proposant pour le cas binaire d'une part une optimisation plus poussée des fonctions d'accusation relativement à la stratégie suivie par la coalition pour produire la copie pirate, et d'autre part des mécanismes d'estimation de cette stratégie ainsi que de la taille de la coalition. Ceci nous a permis de mettre un place un processus complet, itératif, qui tire profit à chaque itération d'une nouvelle estimation de l'ensemble des pirates par une technique d'*Expectation-Maximization* (EM), elle-même ayant tiré profit des scores calculés avec les fonctions d'accusation précédentes. Plus précisément, et tel qu'illustré par la figure 3.3, notre algorithme fonctionne comme suit.

1. Initialisation : nous calculons les scores avec les fonctions d'accusation classiques de [ŠKC08].
2. Un algorithme EM prend en entrées ces scores ainsi que les moyennes et variances théoriques des scores des innocents et des coupables. Il fournit pour chaque score S_j une probabilité $\hat{T}_j$ qu'il corresponde à un coupable.
3. On estime alors le nombre et la stratégie des pirates. On estime la taille de la collusion par $\hat{c} = \lceil \Sigma_{j=1}^{n} \hat{T}_j \rceil$. On considère ensuite les $\hat{c}$ utilisateurs correspondant aux plus grandes probabilités $\hat{T}_j$, et on utilise leurs séquences pour estimer leur stratégie. On modélise la stratégie de la collusion par l'ensemble des probabilités $\theta = \{\mathbb{P}(Y_i = 1 | \Sigma_i = \sigma_i), \sigma_i = 0..c\}_{i=1..m}$, la variable aléatoire $\Sigma_i = \sum_{j \in \mathcal{C}} X_{ji}$ correspondant aux nombre d'identifiants des pirates ayant un 1 à la position i.
4. On optimise les fonctions d'accusation. Cette optimisation vise à maximiser, *via* un Lagrangien, la distance de Kullback-Leibler $D_{KL}(\mathcal{N}_C, \mathcal{N}_I)$ entre la distribution des scores des coupables et celle des innocents, qui suivent des lois normales respectivement de paramètres $\mathcal{N}_C = \mathcal{N}(\mu_C, \sigma_C^2) = \mathcal{N}(m\tilde{\mu}_C, m\tilde{\sigma}_C^2)$ pour les coupables et $\mathcal{N}_I = \mathcal{N}(\mu_I, \sigma_I^2) = \mathcal{N}(m\tilde{\mu}_I, m\tilde{\sigma}_I^2)$ pour les innocents ; la notation « tilde » s'applique individuellement aux composantes de l'identifiant, qui sont i.i.d., d'où les relations $\mu = m\tilde{\mu}$ et $\sigma^2 = m\tilde{\sigma}^2$. Cette optimisation est effectuée sous les contraintes suivantes : $\tilde{\mu}_I = 0$, $\tilde{\sigma}_I = 1$, et indépendance (covariance nulle) des scores des innocents entre eux. La distance de Kullback-Leibler s'écrit alors

$$D_{KL}\left(\mathcal{N}_C, \mathcal{N}_I\right) = \frac{1}{2}\left(m\tilde{\mu}_C^2 - \log(\tilde{\sigma}_C^2) + \tilde{\sigma}_C^2 - 1\right).$$

   Ces contraintes sont les mêmes que dans [FGC08], excepté qu'ici on ne contraint pas la variance des scores des coupables (qui était fixée à $\tilde{\sigma}_C^2 = 1$ dans [FGC08]), et surtout, que l'on relâche l'hypothèse d'indépendance entre la variance des scores des coupables et la stratégie de la coalition. On arrive à des expressions explicites des fonctions d'accusation, données en figure 3.14, et de la valeur de l'espérance du score des coupables.
5. On calcule alors de nouveaux scores avec ces fonctions optimisées, et on itère en bouclant à l'étape 2.

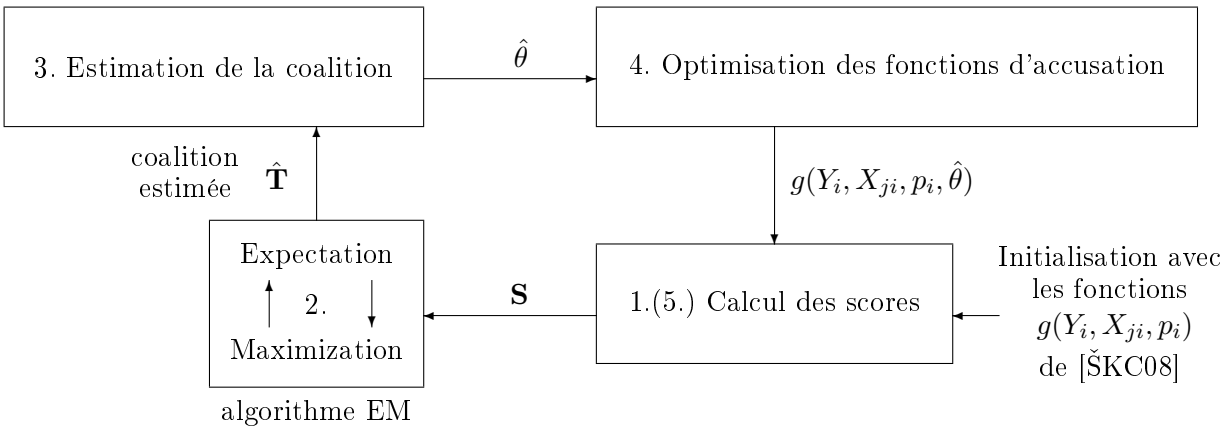


FIGURE 3.13: Estimation de la taille de la coalition ainsi que de la stratégie qu'elle a adoptée pour produire la copie pirate, et optimisation dynamique des fonctions d'accusation.

Pour pouvoir comparer notre optimisation avec celle de [FGC08], nous avon procédé aux même expériences. Celles-ci ont ainsi porté uniquement sur le code anti-collusion, sans prendre en compte l'insertion. Nous avons considéré les mêmes stratégies que dans [FGC08]. Toutes sont de type « échange de blocs ». Elles se nomment *Uniforme, Majorité, Minorité, All1, All0*. Pour estimer la qualité de l'opitimisation des fonctions d'accusation, nous avons calculé le ratio $cm\tilde{\mu}_C/\tilde{\sigma}_I$ obtenu avec les fonctions optimisées données plus haut, pour le comparer aux précédents résultats de [FGC08]. Les résultats sont donnés dans le tableau 3.1. Nos fonctions d'accusation sont, comme espéré, plus efficaces pour la stratégie des pirates pour laquelle elles ont été calculées. Les calculs montrent également qu'elles sont plus efficaces que celles de [FGC08] dans tous les cas, y compris lorsque les stratégies ne coïcident pas.

Pour estimer l'impact de notre processus itératif, ainsi que la qualité de l'estimation de la taille de la coalition c et de sa stratégie θ, nous nous sommes placés dans une situation critique pour les codes de Tardos classiques, choisissant une longueur de code plus courte que celle préconnisée. Nos résultats pour $c = 8$ sont illustrés dans la figure 3.15. On y retrouve en figure (a) ce qui se passe pour un code de Tardos classique [ŠKC08], qui ici n'arrive pas bien à identifier tous les coupables car la longueur de code choisie est plus courte que celle préconisée. Les figures (b), (c) et (d) montrent nos résultats, pour des valeurs de c et de θ connues ou estimées à la volée. On voit que le processus itératif joue bien son rôle, et que l'estimation est tout-à-fait satisfaisante : ils nous permettent d'avoir une accusation globalement bien plus efficace que les codes classiques. On arrive même pour certaines stratégies à retrouver tous les coupables, malgré la faible longueur du code. On note cependant que le comportement de l'accusation diffère largement d'une stratégie à l'autre, certaines étant plus difficiles à affronter. Cette disparité n'est pour l'heure pas expliquée. On

$$g(1,1,p,\hat{\theta}) = \frac{1}{2\lambda}\frac{1-p}{q(p,\hat{\theta})}A(p,\hat{\theta}) \qquad g(0,0,p,\hat{\theta}) = \frac{1}{2\lambda}\frac{p}{1-q(p,\hat{\theta})}A(p,\hat{\theta})$$

$$g(1,0,p,\hat{\theta}) = -\frac{p}{1-p}g(1,1,p,\hat{\theta}) \qquad g(0,1,p,\hat{\theta}) = -\frac{1-p}{p}g(0,0,p,\hat{\theta})$$

avec

$$\begin{aligned}\lambda &= \frac{1}{2}\sqrt{\mathbb{E}_p\left[A^2(p,\hat{\theta})\frac{p}{q(p,\hat{\theta})}\frac{1-p}{1-q(p,\hat{\theta})}\right]}\\ q(p,\hat{\theta}) &= \mathbb{P}(Y=1|P=p,\hat{\theta})\\ A(p,\hat{\theta}) &= \mathbb{P}(Y=1|X=1,P=p,\hat{\theta}) - \mathbb{P}(Y=1|X=0,P=p,\hat{\theta})\end{aligned}$$

On peut alors calculer l'espérance des coupables (maximale) :

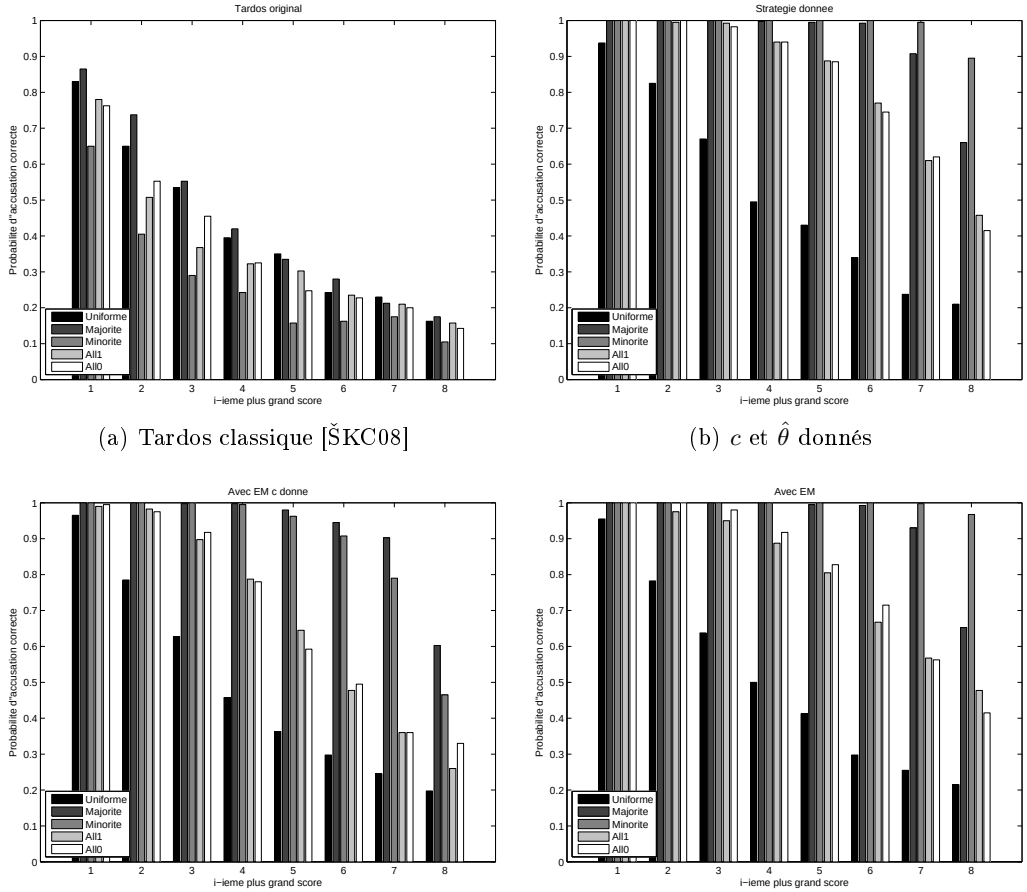
$$\tilde{\mu}_C = \sqrt{\mathbb{E}_p\left[A^2(p,\hat{\theta})\frac{p}{q(p,\hat{\theta})}\frac{1-p}{1-q(p,\hat{\theta})}\right]}.$$

FIGURE 3.14: Fonctions d'accusation qui maximisent l'espérance du score des coupables.

		Stratégie de la coalition				
c	optimisation	Uniforme	Majorité	Minorité	All1	All0
3	Uniforme	**98 (71)**	106 (80)	100 (53)	97 (66)	97 (66)
	Majorité	96 (67)	**110 (84)**	100 (34)	95 (59)	95 (59)
	Minorité	81 (50)	59 (38)	**112 (75)**	89 (56)	89 (56)
	All1	83 (69)	88 (73)	88 (62)	**114 (68)**	84 (68)
	All0	83 (69)	88 (73)	88 (62)	84 (68)	**114 (68)**
4	Uniforme	**98 (71)**	106 (80)	105 (44)	99 (62)	99 (62)
	Majorité	96 (67)	**110 (84)**	105 (17)	97 (50)	97 (50)
	Minorité	61 (34)	25 (15)	**128 (91)**	88 (53)	88 (53)
	All1	79 (65)	83 (63)	88 (72)	**121 (67)**	87 (67)
	All0	79 (65)	83 (63)	88 (72)	87 (67)	**121 (67)**
5	Uniforme	**98 (71)**	110 (83)	110 (33)	100 (58)	100 (58)
	Majorité	94 (63)	**120 (93)**	113 (-22)	98 (35)	98 (35)
	Minorité	37 (19)	-20 (-17)	**155 (121)**	82 (52)	82 (52)
	All1	77 (59)	83 (47)	90 (90)	**128 (69)**	90 (69)
	All0	77 (59)	83 (47)	90 (90)	90 (69)	**128 (69)**

TABLE 3.1: Valeurs de $cm\tilde{\mu}_C/\tilde{\sigma}_I$ obtenues après optimisation des fonctions d'accusation pour $m = 100$, $c = 3, 4, 5$. Entre parenthèses sont données celles obtenues par [FGC08]. Rappelons qu'avec les fonctions de [ŠKC08] on a 64 dans tous les cas.

note également que notre estimation de la taille de la coalition par l'algorithme EM n'est pas très précise, la valeur obtenue étant souvant trop grande. Accuser le plus grand score est alors plus sûr.



(a) Tardos classique [ŠKC08]

(b) c et $\hat{\theta}$ donnés

(c) c donné, $\hat{\theta}$ estimé

(d) avec EM, c et $\hat{\theta}$ estimés

FIGURE 3.15: Probabilité d'accuser correctement le k-ième plus grand score, pour $k \in \{1, \dots, 8\}$. $m = 1000$, $c = 8$, $n = 5000$. 400 expériences réalisées. On compare l'efficacité d'une accusation classique (a) ici mise en défaut par une longueur trop faible par rapport à celle préconisée, avec une accusation optimisée (b)(c)(d).

Ces résultats ont été présentés lors de la conférence internationale *IS&T/SPIE International Symposium on Electronic Imaging'2009* [CXFF09], ainsi qu'au GRETSI 2009 [CFF09] et lors des journées C2 du GDR IM en 2009. Ils ont fait l'objet d'un article publié par la revue française *Traitement du Signal* en 2010 [CFF10]. Ces travaux ont étés valorisés dans le cadre des projets ANR-RIAM ESTIVALE et MEDIEVALS.

3.4 Intégration des codes de Tardos dans un protocole de traçage dit asymétrique, thèse de A. Charpentier (2008-2011)

Avec Ana Charpentier, Teddy Furon et Ingemar Cox, nous nous sommes intéressés ces deux dernières années à la conception d'un protocole de fingerprinting asymétrique, tel ceux présentés dans le paragraphe « Comment intégrer ces techniques dans un protocole complet de distribution de documents ? » page 66. Un tel protocole permet aux utilisateurs de participer eux-mêmes à la génération de leur identifiant, et de s'assurer que le fournisseur de contenus n'a jamais en main l'exemplaire du document destiné à un utilisateur donné, ni l'identifiant qui lui est associé. Ainsi, un fournisseur de contenus malhonnête n'est plus en mesure d'accuser sciemment un utilisateur innocent, puisqu'il ne connaît pas son identifiant, et n'est pas en possession de la version du document qui lui a été délivrée. Un utilisateur qui se retrouve accusé ne peut donc plus prétendre être victime d'un fournisseur de contenus malhonnête.

Notre objectif était d'explorer comment les codes de Tardos, considérés comme les meilleurs codes anti-collusion actuels, pouvaient s'intégrer dans un tel protocole. Nous avions aussi comme préoccupation de spécifier le protocole autant que possible, pour faciliter sa mise en œuvre. Nous avons finalement proposé un protocole asymétrique pour les codes de Tardos binaires, sans nous pencher pour l'instant sur les problèmes d'anonymat, ou les spécificités des codes q-aires.

Nous avons tout d'abord montré qu'il était possible au fournisseur de contenus de tricher lors du processus d'accusation en modifiant (même légèrement) les valeurs du vecteur $\mathbf{p}$. Ces modifications ont en effet comme conséquence d'augmenter la valeur de tous les scores, comme le montre la figure 3.16. Ainsi, le score d'un innocent peut très bien se trouver finalement dépasser le seuil et donner lieu à une accusation à tort. Cette étude montre qu'il faut prendre beaucoup de précautions lors de l'utilisation du vecteur $\mathbf{p}$, et mettre en œuvre des moyens pour garantir son intégrité. Elle montre donc aussi que les codes de Tardos imposent par leur nature des contraintes sur le protocoles, et qu'on ne peut donc pas simplement les intégrer à un protocole existant. Dans notre protocole, nous suggérons d'obliger le fournisseur à déposer un engagement (*commitment*) inintelligible mais public sur les composantes de ce vecteur dans une mémoire *WORM* (Write Once Read Many) dans laquelle on ne peut écrire qu'une seule fois. Ainsi, lors d'une transaction, l'utilisateur peut s'assurer que les informations qui lui sont fournies sur le vecteur $\mathbf{p}$ sont cohérentes avec ce qui a été déposé, et donc que l'identifiant qui sera généré correspondra au bon vecteur. De même, lors d'un calcul de score, on peut s'assurer de la cohérence du vecteur $\mathbf{p}$ utilisé pour calculer le score avec ce qui a été déposé, et s'assurer ainsi que le score correspondra au bon vecteur.

Comme nous l'avons mentionné dans l'introduction, si on se limite à la notion d'asymétrie sans se préoccuper de l'anonymat, la conception de tels protocoles se heurte à trois étapes délicates.

1. L'utilisateur génère son identifiant, avec l'aide du fournisseur. Il y a alors deux

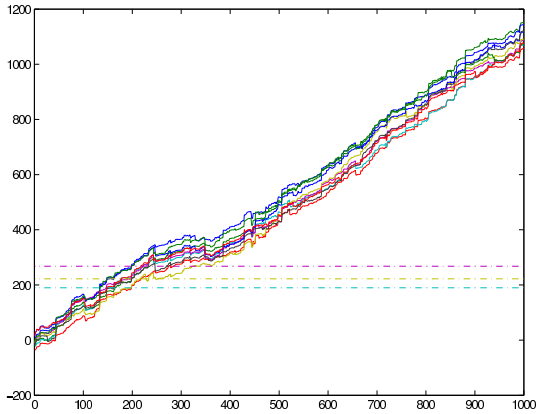


FIGURE 3.16: Impact de la modification du vecteur $\mathbf{p}$ sur le calcul des scores. Expériences menées pour $m = 1000$ avec $c = 3$. L'axe des abscisses indique le nombre de composantes modifiées dans le vecteur $\mathbf{p}$. Les lignes colorées en plein montrent comment les scores de dix utilisateurs pris au hasard augmentent. Les lignes en pointillés montrent les scores des coupables qui seraient calculés avec le bon vecteur $\mathbf{p}$.

points délicats : l'utilisateur doit générer son identifiant suivant les recommandations du fournisseur, et parfois même certains paramètres secrets connus seulement du fournisseur, sans tricher ; le fournisseur doit ensuite pouvoir s'assurer que l'utilisateur a suivi les recommandations, sans pour autant connaître l'identifiant dans son intégralité.

2. Le fournisseur doit fournir à l'utilisateur le contenu demandé, tatoué avec son identifiant. Le point délicat est que l'utilisateur ne doit pas rentrer en possession du contenu non tatoué, et le fournisseur ne connaît pas l'identifiant en entier.
3. Lorsque le fournisseur suspecte une fraude, il procède au traçage pour retrouver au moins un des coupables. Le point délicat est ici de s'assurer que le fournisseur ne triche pas pour biaiser le processus et accuser un innocent.

Notre protocole est trop technique pour être détaillé ici, et je n'en présente ci-dessous que les principes.

La principale difficulté pour utiliser des codes de Tardos dans un protocole asymétrique vient ensuite du fait que l'identifiant (ou une partie de l'identifiant) doit être généré(e) par l'utilisateur, conformément au vecteur $\mathbf{p}$ qu'il ne doit pas connaître (point 1.). Ainsi, l'utilisateur doit générer un bit qui suit une certaine distribution ($\mathbb{P}(X_{ji} = 1) = p_i$) sans connaitre la valeur de p_i. Nous avons résolu ce problème par l'utilisation d'un protocole d'*Oblivious Transfer*. L'*Oblivious Transfer* est une primitive cryptographique [Rab81] qui permet à Bob de choisir k éléments au hasard dans une liste de N éléments possédée par Alice, de telle sorte que : 1) Bob obtient que des éléments qui sont réellement dans la liste ; 2) Bob n'obtient pas d'information sur les éléments qu'il n'a pas choisis ; 3) Alice ne sait pas quels sont les éléments choisis par Bob. Nous avons transformé chaque réel p_i en une liste de N bits contenant L_i '1', avec p_i le plus proche possible de L_i/N. C'est cette liste, permutée différemment

pour chaque utilisateur, qui sera utilisée par le protocole d'*Oblivious Transfer*, le fournisseur de contenus jouant ici le rôle d'Alice, tandis que l'utilisateur joue le rôle de Bob. En tirant un élément de la liste au hasard, l'utilisateur tirera donc un bit, avec une probabilité très proche de p_i que ce soit un '1'. Cette opération est répétée pour chacune des m composantes de l'identifiant à générer. Nous avons montré que pour des raisons de sécurité la longueur de la liste N doit au moins égale à 100, et que l'impact de cette quantification de p_i sur le calcul des scores est d'autant plus faible que la taille de la coalition est grande. Nous avons discuté l'adéquation des protocoles d'Oblivious Tranfer existants, et les avons adaptés à nos contraintes, en les améliorant au besoin. La plupart sont issus de la communauté cryptographique [NP99, CT05, GH07] (et garantissent une sécurité plus formelle), tandis que quelques autres relèvent approches réellement différentes, comme le *Commutative Encryption Scheme* [BDF01, HC05, WZW03]. Ce dernier a été beaucoup moins étudié, et sa sécurité est moins formalisée, mais s'adapte mieux à nos contraintes. Nous avons donc décliné techniquement ces deux solutions, qui offrent des avantages différents. Nous travaillons par ailleurs actuellement au renforcement de la sécurité des *Commutative Encryption Schemes*, afin d'obtenir des schémas atteignant un niveau de sécurité appelé *sécurité sémantique*, essentiel dans notre cas, et qui n'était pas atteint par les schémas présents dans la littérature.

Une fois cette étape franchie, le fournisseur de contenus doit s'assurer que l'identifiant généré par l'utilisateur est conforme au vecteur $\mathbf{p}$, sans pour autant que l'utilisateur lui révèle son identifiant dans son intégralité (point 1.). Par ailleurs, le fournisseur de contenus doit également posséder suffisamment d'information sur l'identifiant de l'utilisateur pour pouvoir lancer son protocole d'accusation en cas de découverte d'une copie falsifiée. C'est pourquoi nous proposons une deuxième phase dans laquelle l'utilisateur révèle une partie (qu'il ne choisit pas lui-même) de son identifiant au fournisseur de contenus, en utilisant à nouveau un protocole d'*Oblivious Transfer*. Lors de cet échange, nous nous assurons que les protagonistes n'obtiennent bien que ce qu'ils doivent apprendre. Avec cet identifiant partiel, le fournisseur est alors en mesure de calculer un score partiel, qui peut lui permettre de décider s'il suspecte ou non l'utilisateur. S'il le suspecte, il peut alors enclencher une procédure judiciaire, plus lourde, au cours de laquelle le juge sommera l'utilisateur de révéler l'intégralité de son identifiant, pour procéder à un calcul de score complet. Cette étape clot le point 1.

Le point 2. concerne l'insertion dans le medium (image, vidéo) de l'identifiant, par un procédé de tatouage qui permet au fournisseur de contenus d'insérer dans le document un identifiant sans le connaitre. Nous utilisons à cette étape des schémas existants reposant sur du chiffrement homomorphique [Kur10, DBPP09].

Le point 3. est assuré par l'utilisation de la mémoire WORM, qui conserve des engagements (*commitment*) sur les listes représentant les p_i. Par leur nature-même, ces engagements sont « chiffrés » et ne révèlent donc pas la valeur des p_i. Le lien entre les p_i et les éléments engagés peut néanmoins être vérifié pour en assurer l'intégrité.

Tous les comportements malicieux auxquels nous avons pensé au cours de l'élaboration des différentes phases du protocole ont été analysés et pris en compte pour être contrés par le protocole.

Ces résultats ont été publiés cette année à la conférence internationale *International Workshop on Information Hiding, IH'11* [CFFC11] [6]. Un article de journal est en cours de rédaction. Ce travail a été valorisé dans le cadre du projet ANR-RIAM MEDIEVALS.

Ce travail ouvre de nombreuses perspectives quant à l'amélioration de la sécurité des *Commutative Encryption Schemes*, à la formalisation de la sécurité générale du protocole, et à son extension pour le transformer en un protocole de fingerprinting asymétrique anonyme.

3.5 Conclusion et perspectives

Les codes de Tardos constituent aujourd'hui les meilleurs codes anti-collusion génériques. Suite à l'ensemble des travaux menés depuis 2003, ils sont aujourd'hui bien compris, maîtrisés et optimisés, surtout dans le cas binaire. Les contributions présentées dans ce chapitre ont porté sur : l'estimation dynamique de la stratégie d'attaque de la coalition et l'optimisation du processus d'accusation (section 3.3) ; l'articulation de ces codes avec la technique d'insertion `Broken Arrows` étudiée par ailleurs au chapitre précédent (section 3.2) ; la conception d'un protocole complet de distribution asymétrique utilisant ces codes (section 3.4).

L'optimisation de la longueur et du processus d'accusation ayant probablement été poussés à leur maximum dans le cas binaire, les grandes avancées à venir porteront principalement sur l'optimisation du processus d'accusation dans le cas q-aire, et l'intégration de ces codes dans des protocoles complets de distribution. Avancées sur l'articulation entre le code lui-même et la technique d'insertion, mais aussi avancées dans la conception des protocoles eux-mêmes.

C'est ce dernier point qui m'intéresse le plus, et le travail que nous avons mené avec Ana Charpentier, Teddy Furon et Ingemar Cox, ouvre en ce sens plusieurs perspectives.

La première concerne l'amélioration de la sécurité des *Commutative Encryption Schemes*, qui se plient particulièrement bien aux contraintes propres à l'utilisation des codes de Tardos. Ils présentent aujourd'hui un inconvénient majeur : on ne trouve dans la littérature aucun exemple de tel schéma atteignant la *sécurité sémantique*, alors que ce niveau de sécurité s'avère essentiel ici car le schéma est amené à chiffrer des bits. Cet inconvénient peut être relativement vite supprimé, car il est clair par exemple que l'on peut aisément obtenir un schéma IND-CPA (qui est le niveau de sécurité sémantique le plus bas) en s'appuyant sur le système de chiffrement El Gamal. On peut peut-être même aller plus loin et obtenir des schémas IND-CCA1, voire IND-CCA2 (qui sont des niveaux supérieurs de sécurité sémantique).

La deuxième concerne l'analyse de la sécurité globale du protocole. Nous l'avons pour l'instant analysée au regard d'un certain nombre de comportements malicieux, considérés séparément. Une autre approche, à développer, serait de s'appuyer sur les modèles *semi-honnête* et *malicieux* pour voir s'il est possible d'obtenir une preuve

6. Article joint en annexe D.

globale de la sécurité du protocole. L'idéal serait bien sûr une preuve dans le modèle malicieux. Mais une telle preuve sera probablement très difficile à mener, pas nécessairement à cause de faiblesses du protocole, mais parce qu'une telle preuve est par nature difficile à appréhender, d'autant que les primitives utilisées ici sont de natures très différentes (*Oblivious Transfer*, chiffrement homomorphique, tatouage, etc) et donc difficiles à analyser simultanément. Une preuve dans le modèle semi-honnête me parait plus abordable, mais assez frustrante. Car si elle permet de montrer que tout se passe bien quand chacun tient son rôle, ce qui est déjà important, il me semble peu réaliste de penser que chacun tiendra effectivement son rôle.

La troisième concerne l'intégration de la notion d'anonymat ou de *privacy* dans le protocole. Usuellement, les protocoles asymétriques anonymes s'appuient pour cela sur une tiers partie de confiance. J'aimerais beaucoup travailler à l'extension du protocole que nous avons développé pour le rendre anonyme, voire *private*. J'aimerais par ailleurs étudier dans quelle mesure on pourrait alors se passer d'un tiers de confiance en s'appuyant sur des techniques de *Multi-Party Computiation*.

Enfin, une pespective plus générale porte sur l'utilisation des codes anti-collusion, et en particulier des codes de Tardos, dans d'autres contextes où l'on souhaite pouvoir exploiter des calculs ou des tests effectués sur des groupes d'individus. L'application des codes de Tardos en *Group Testing* en est un exemple. Mais il y a très probablement d'autres contextes qui pourraient bénéficier de leur efficacité.

Chapitre 4

Conclusion et perspectives générales

J'ai présenté dans ce manuscrit l'ensemble de mes travaux portant sur la protection de contenus multimédia, de leur création à leur diffusion.

Le chapitre 1 a présenté les problématiques liées à la stéganographie, et plus particulièrement à l'usage détourné des codes correcteurs dans la conception de tels systèmes. J'y ai présenté quelques pistes explorées avec Fabien Galand, Daniel Augot et Morgan Barbier pour garantir le succès de l'insertion, tout en préservant autant que possible les critères habituels de furtivité et d'efficacité d'insertion.

Le chapitre 2, consacré aux techniques de tatouage robuste, a montré à quel point la robustesse est insuffisante pour garantir la pérennité de la marque. J'y ai présenté une méthodologie et des outils, développés avec Teddy Furon et François Cayre, qui permettent de formaliser et analyser la sécurité de ces schémas. Nous nous sommes appuyés sur ces nouveaux outils pour étudier la sécurité des techniques de tatouage substitutives, ainsi que des techniques additives reposant sur l'étalement de spectre. Avec Teddy Furon et Fuchun Xie, nous avons ensuite étudié comment renforcer plus spécifiquement la robustesse et la sécurité de la technique de tatouage `Broken Arrows`.

En nous appuyant sur cette technique d'insertion robuste et ses variantes améliorées, nous avons proposé dans le chapitre 3 un schéma d'identification de copies (*fingerprinting*) complet. Ce schéma repose sur les codes anti-collusion de Tardos, qui sont les plus efficaces aujourd'hui. Nous avons tout d'abord montré comment ces codes peuvent être modifiés pour contrer, grâce à la très forte robustesse de `Broken Arrows`, des attaques de type fusion, considérées jusqu'alors comme parmi les plus dangereuses. Avec Teddy Furon et Ana Charpentier, nous avons également montré comment le processus d'accusation des codes de Tardos peut être amélioré lorsque l'on connaît la stratégie d'attaque des pirates. Ce résultat a débouché sur la proposition un schéma qui estime à la volée cette stratégie, pour ensuite optimiser l'identification des pirates. Nous avons enfin montré, avec Teddy Furon, Ana Charpentier et Ingemar Cox, comment ces codes peuvent être intégrés dans un protocole de distribution asymétrique, qui garantit qu'aucune des parties, fournisseur de contenu ou utilisateur, ne pourra tricher pour léser l'autre.

Ces travaux ouvrent tous de nombreuses perspectives. Voici celles qui retiennent le plus mon attention.

L'utilisation de codes correcteurs en stéganographie donne lieu à de nombreuses perspectives, comme par exemple l'étude de l'utilisation de codes non-linéaires. Celles qui m'intéressent le plus actuellement concernent d'une part l'exploitation de codes q-aires dans des schémas complets, et d'autre part la construction que nous avons proposée avec Daniel Augot et Morgan Barbier pour assurer l'insertion du message. Cette construction récente est très prometteuse et doit être étudiée plus en profondeur. Nous avons plusieurs idées à développer dans les prochains mois pour en développer tout le potentiel.

Les codes de Tardos sont maintenant bien établis et ont été beaucoup étudiés, surtout dans le cas binaire. Il reste cependant encore des points à développer. Lors de la conception de notre protocole de fingerprinting asymétrique, nous avons découvert une voie parallèle à la voie cryptographique traditionnelle pour la conception de protocoles d'*Oblivious Transfer* : les *Commutative Encryption Schemes*. Cette voie alternative offre des propriétés très intéressantes dans notre contexte. Elle n'a pas été beaucoup étudiée par la communauté et il est certain que l'on pourrait proposer des protocoles de ce type offrant de meilleurs niveaux de sécurité que ceux actuellement publiés. Une autre perspective importante de ce travail est de voir dans quelle mesure la sécurité de ce protocole peut être prouvée, et pour quel modèle d'attaquant. Une autre, enfin, retient particulièrement mon attention. Car si j'ai beaucoup travaillé sur l'identification de contenus, je me préoccupe aussi de la préservation de la vie privée. Et si ces deux questions semblent antagonistes à première vue, elles ne sont pas incompatibles. L'extension du protocole de *fingerprinting* asymétrique présenté ici, pour en faire un protocole anonyme, voire *private*, en serait une première étape.

Annexe A

Autres contributions

Cette annexe présente brièvement quelques autres contributions.

A.1 Tatouage

A.1.1 Élaboration d'une plate-forme d'évaluation pour les algorithmes de tatouage (1999-2001)

En novembre 1997, Fabien Petitcolas publiait la première version du logiciel `StirMark`. Ce logiciel permettait de tester facilement la robustesse des techniques de tatouage d'image face à des distorsions locales de l'image. Cet outil a connu un grand succès, et a ensuite été enrichi d'autres attaques. En janvier 1999, avec Fabien Petitcolas et Frédéric Raynal, nous avons décidé d'aller plus loin, et de développer une plate-forme d'évaluation automatique de techniques de tatouage. Ce projet était motivé par le fait que dans la littérature chacun effectuait ses propres tests, avec des paramètres différents, et qu'il était difficile de comparer réellement les techniques entre elles, à moins de les reprogrammer pour les tester soit-même. Ce projet a été baptisé `StirMark Benchmark`.

D'autres chercheurs nous ont alors rejoints : Nazim Fates (Microsoft Research Lab, Angleterre), Jana Dittmann et Martin Steinebach (German National Research Center for Information Technology, Allemagne).

Ce projet a finalement donné lieu au développement de bibliothèques de test *open source*[1]. Les tests que nous avons développés dans ces bibliothèques concernent les images, mais aussi les vidéos et les fichiers audio.

En parallèle se sont développés d'autres projets similaires, comme le projet européen `Certimark`[2], ou encore le projet `Checkmark`[3]. Ces trois projets se sont avérés complémentaires, car ils ne proposent pas exactement les mêmes tests, ne traitent pas tous des mêmes types de documents, et sont développés dans des langages différents.

1. `http://www.petitcolas.net/fabien/watermarking/stirmark/` et `http://wwwiti.cs.uni-magdeburg.de/~alang/smba.php`.
2. `http://www.certimark.org/`
3. `http://cvml.unige.ch/ResearchProjects/Watermarking/Checkmark/`

Notre travail a fait l'objet de deux articles présentés aux conférences *IS&T/SPIE International Symposium on Electronic Imaging'2001* [PSR+01] et *International Conference on Information Technology : Coding and Computing, ITCC 2001* [DFF+01], ainsi que d'un article dans le numéro spécial "tatouage et sécurité de l'information" de la revue *Traitement du signal* [RPF01]. Il a également fait l'objet d'un article d'ouvrage [PF04].

A.1.2 Tatouage et compression conjoints (2006-2008)

L'objet du *tatouage robuste* est d'insérer dans le document une information imperceptible mais néanmoins capable de persister malgré un certain nombre d'attaques/traitements potentiels (compression, conversion analogique-numérique, filtrage, etc). Les traitements à prendre en compte dépendent fortement du contexte applicatif. Néanmoins, étant donné que les documents sont presque toujours sujets à compression pour stockage ou diffusion, on peut considérer que le minimum est de résister à une compression. Par ailleurs, puisque cette étape de compression est systématique dans la plupart des applications et que tatouage et compression peuvent tous deux être modélisés comme une transmission *via* un canal, il est intéressant de les envisager conjointement.

Avec Çağatay Dikici, Christine Guillemot, Khalid Idrissi et Attila Baskurt, nous avons adopté cette approche pour étudier le cas d'un schéma conjoint reposant sur une technique de tatouage et un codeur de type Wyner-Ziv. Ce travail a été entamé à l'automne 2006 lors du séjour de Çağatay Dikici, alors en thèse au LIRIS à Lyon, dans notre équipe. Nous avons donc considéré un schéma de tatouage robuste avec information adjacente (voir page 37) au niveau de l'émetteur, comme modélisé par Gel'fand et Pinsker [GP80]. Costa a montré en 1983 qu'alors, dans le cas d'un bruit blanc gaussien additif, la capacité du canal reste la même que si l'état du canal était connu à la fois de l'émetteur et du récepteur [Cos83]. L'élaboration de codes permettant d'atteindre cette capacité est connue sous le nom de *dirty paper coding*. Moulin *et al.* se sont intéressés dans [MW07] au problème de codage soulevé par Gel'fand-Pinsker, dans le cas d'un alphabet source discret. Ils ont établi la capacité du canal et les paramètres des codes effectifs dans les cas où l'émetteur, l'attaquant et/ou le récepteur disposent d'informations adjacentes partielles.

Nous nous sommes intéressés au *dirty paper coding* dans le cas où le récepteur dispose d'une information adjacente (sur le document hôte) partielle (*i.e.* bruitée). Les distorsions introduites par les opérations de tatouage, puis de compression sont maîtrisées, *i.e.* contraintes par des bornes supérieures. Le module de compression ne dispose d'aucune information sur le document original et travaille uniquement avec sa version tatouée. Cette étape de compression peut être assimilée au problème de codage de Wyner-Ziv [WZ76], à savoir une compression avec perte et information adjacente disponible au décodeur.

Nous avons tout d'abord déterminé la fonction de taux-distorsion du module Wyner-Ziv. Contrairement aux précédentes dérivations établies dans le cas d'une information adjacente partielle, nous avions à tenir compte des contraintes posées par le module de tatouage en termes de puissance de signal et de distorsion. Nous avons

modélisé les opérations de compression et décompressions comme un canal de transmission, pour calculer la capacité du module de tatouage, dans le cas de signaux gaussiens. Nous avons ainsi pu analyser les gains en termes de taux et de capacité, lorsque le récepteur dispose de cette information adjacente partielle. Car si Costa a montré qu'aucun gain n'est à espérer quand l'émetteur a une connaissance parfaite du document hôte, nous avons montré que dans le cas de notre schéma la connaissance d'une information partielle lors du décodage Wyner-Ziv permet de diminuer la distorsion liée à la compression, pour un taux de compression fixé. Indirectement cela nous permet d'augmenter la capacité du canal global.

Ce travail théorique a été accompagné par une étude pratique. De nombreux articles ont présenté des codes effectifs répondant aux problèmes posés par Costa et Wyner-Ziv. Le schéma effectif que nous proposons repose sur les *superposition codes* introduits par [BBCS06]. Notre module de tatouage utilise une quantification TCQ (Trellis Coded Quantization) suivie d'un code LDPC (Low Density Parity Check). Le module de compression s'appuie quant à lui sur un quantificateur scalaire suivi de codes LDPC. Nos simulations ont montré que ce schéma permet d'approcher les bornes dérivées pour la fonction de taux-distorsion et la capacité.

Ces résultats ont été présentés lors de la conférence internationale *IEEE International Symposium on Image/Video Communications over fixed and mobile networks, ISIVC 2008* [DGF+08].

A.2 Conception et attaque de systèmes de chiffrement

Dans la lignée des travaux menés durant ma thèse, je me suis intéressée à la confidentialité des données, *via* la conception et l'attaque de *systèmes de chiffrement à clé secrète* (symétriques), qu'ils opèrent *par blocs* [les blocs de message étant traités séquentiellement par des opérations de mélange] ou *par flot* [une suite chiffrante étant alors ajoutée au texte clair, symbole après symbole, pour produire le chiffré].

Je me suis dans un premier temps intéressée aux fonctions booléennes (vectorielles dans le cas du chiffrement par blocs) qui interviennent comme primitives dans ces systèmes, car la sécurité des systèmes repose en grande partie sur leurs qualités. Plus précisément, elles interviennent dans les opérations de mélange pour le chiffrement par bloc, ou dans la génération de la suite chiffrante pour le chiffrement à flot. Ces travaux, débutés pendant ma thèse, ont été poursuivis dans le but, non seulement d'exhiber de bonnes primitives, mais aussi de formaliser les liens entre les différents critères que ces primitives doivent vérifier afin de mieux cerner le compromis à réaliser, et de proposer un système complet et opérationnel de chiffrement à clé secrète particulièrement rapide, appelé `COS`. Ils ont été menés en collaboration avec les chercheurs que j'ai cotoyés lors de ma thèse au projet CODES de l'INRIA.

J'ai développé en parallèle une étude plus personnelle sur la génération de suites pseudo-aléatoires non linéaires comme suites chiffrantes pour le chiffrement à flot. Il existe plusieurs stratégies pour les générer, toutes reposant sur des automates déterministes faisant intervenir des fonctions de transition, linéaires ou non. Traditionnellement, ces suites sont produites par des registres à décalage à rétroac-

tion linéaire (LFSR), qui sont filtrés ou combinés par des fonctions booléennes non linéaires. Les LFSRs ont été très étudiés, et offrent les avantages d'une implémentation très efficace et d'une bonne maîtrise des propriétés de la suite produite. Parmi les attaques proposées contre de tels systèmes, on peut citer l'attaque de Berlekamp-Massey en 1969 [Mas69], les attaques par corrélation (rapides) depuis 1985 [Sie85, MS88, CT00, JJ02], et depuis 2002 les attaques algébriques [Cou02, CM03b, Cou03, Cou05, HR04, AFI+04]. Au fil de la publication de ces attaques, la conception de systèmes robustes utilisant des LFSRs est devenue de plus en plus ardue, et il est impératif de trouver d'autres manières de générer efficacement des suites non linéaires. Des appels à proposition ont d'ailleurs été lancés à plusieurs reprises par les projets européens NESSIE et ECRYPT. Certains travaux se sont orientés vers l'utilisation de registres à décalage différents comme les registres à décalage avec retenue (FCSR), ou de manière plus générale les registres à décalage à rétroaction non linéaire (NLFSR) ; d'autres travaux ont exploré des fonctions de transition plus exotiques, comme par exemple les T-fonctions [KS02, KS04a, KS04b]. J'ai travaillé à la conception de ces suites pseudo-aléatoires non linéaires en utilisant des registres dont la rétroaction est non linéaire (NLFSR), d'abord avec Eric Filiol avec qui j'avais collaboré pendant ma thèse, puis avec Vincent Bénony dont j'ai encadré la thèse à Lille de 2002 à 2006. En parallèle, nous avons mené avec Vincent Bénony une cryptanalyse d'un schéma utilisant des T-fonctions.

A.2.1 Formalisation des compromis et construction de bonnes primitives (1998-2001)

J'ai travaillé pendant ma thèse à la construction de fonctions booléennes (vectorielles) présentant les critères nécessaires à leur utilisation comme primitives dans des systèmes de chiffrement à clé secrète. Pour résister aux attaques connues en 2000, ces fonctions devaient vérifier un certain nombre de critères, en partie incompatibles, et les compromis à réaliser n'étaient pas entièrement établis. Je me suis donc attachée, en collaboration avec Anne Canteaut, Pascale Charpin et Claude Carlet à formaliser une partie de ces compromis pour le cas de fonctions sous-optimales au regard d'un seul critère, mais globalement plus satisfaisantes sur leur ensemble. Nous nous sommes particulièrement focalisés sur le cas des fonctions booléennes vectorielles utilisées en chiffrement par blocs et sur le compromis à réaliser entre une haute *non-linéarité* et un *degré de propagation* élevé.

Ce travail a été présenté lors des colloques *EUROCRYPT'2000* [CCCF00b] et *IEEE International Symposium on Information Theory'2000* [CCCF00a], et a fait l'objet d'un article long (*regular paper*) dans la revue *IEEE Trans. on Information Theory* [CCCF01] [4].

4. Article joint en annexe D.

A.2.2 Alternative aux LFSRs : conception du système de chiffrement COS[vd] (1999-2004)

En parallèle, nous avons utilisé avec Eric Filiol les fonctions construites pendant ma thèse pour concevoir un nouveau système de chiffrement à clé secrète, appelé COS. L'idée générale de ce schéma est de combiner les avantages des chiffrements à clé secrète *par blocs* avec ceux du chiffrement par clé secrète *à flot*, en évitant leurs inconvénients respectifs. Les éléments de base du système sont des registres à décalage dont la rétroaction n'est pas linéaire (NLFSRs). Chaque rétroaction est définie par une fonction booléenne fortement non linéaire. Nous avons choisi d'utiliser des fonctions générées pendant ma thèse, car elles satisfont toutes les propriétés requises, et leur spectre de Fourier est particulièrement plat, ce qui rend les attaques plus difficiles (il n'y a pas de biais à exploiter). Nous avons élaboré une construction originale, nommée *crossing over*, qui permet d'utiliser les états internes des registres comme parties pour les blocs de sortie du système de chiffrement. Une des spécificités de COS est sa rapidité, puisqu'il permet d'attendre une vitesse de chiffrement de 978 Mbits/s en version logicielle. Nous en avons initialement décliné deux modes : le premier (I) est dédié aux documents quelconques, et le deuxième aux données sans redondance (typiquement des données compressées). Le mode I est particulièrement destiné aux utilisateurs néophytes qui ne savent pas qu'il est très imprudent de chiffrer des données structurées (cela peut donner lieu à des attaques car on peut modéliser la structure statistique du texte clair) et n'aura pas pris la précaution de briser cette structure avant chiffrement. Le chiffrement s'effectue de la même manière dans les deux cas, mais le mode II donne lieu à un pré-traitement des données avant le chiffrement lui-même. Cette première version de COS a été publiée à *IEEE International Symposium on Information Theory'2001* [FFV01] ainsi qu'à la conférence internationale *IMA Conference on Cryptography and Coding 2001* [FF01]. Depuis sa publication, ce système a subi plusieurs attaques, mais uniquement pour la version I. La première attaque n'a été diffusée que sous la forme de preprints [5] [6], et n'a de valeur que théorique : elle nécessite un facteur de travail d'au moins 2^{384} , ou 2^{3072} selon les paramètres choisis [7]. Une deuxième attaque, en revanche, a été publiée [WB02] ; cette attaque est très efficace lorsque la taille des blocs est de 128 bits, et que l'on utilise seulement deux registres internes. Elle reste inefficace dans le cas de blocs plus gros, et lorsque l'on utilise trois registres internes. Notons que l'attaque porte sur le module de compression, et non sur le module de chiffrement lui-même. La version II reste, à l'heure actuelle, non attaquée. Eric Filiol, militaire, a conçu pour la défense une variante du mode II, dans laquelle un module chaotique renforçe la non linéarité du chiffrement. Après accord des autorités militaires, nous avons décidé de publier cette version, appelée COSvd, qui résiste à toutes les attaques connues et conserve un débit extrêmement rapide. Nous avons réalisé de nouveaux tests (tests statistiques,

5. *The COS Stream Ciphers are Extremely Weak*, S. Babbage, http://eprint.iacr.org/2001/078

6. Cryptanalysis of the COS (2, 128) stream ciphers, S. Babbage, http://eprint.iacr.org/2001/106

7. *COS Ciphers are not "extremely weak" ! - The Design Rationale of COS Ciphers*, E. Filiol et C. Fontaine, http://eprint.iacr.org/2001/080

mesures de critères cryptographiques) afin de renforcer notre argumentation quant à sa sécurité, et avons publié cette nouvelle version dans [FFJ04]. Aucune attaque n'a été publiée depuis sur cette version du système.

A.2.3 Alternatives aux LFSRs : thèse de V. Bénony (2002-2006)

Nous avons étudié, avec Vincent Bénony, François Recher et Eric Wegrzynowski la production de suites de de Bruijn à l'aide de NLFSRs. Rappelons brièvement qu'une suite de de Bruijn binaire d'ordre n est une suite de période 2^n contenant tous les motifs de longueur n une et une seule fois. Ces suites, utilisées dans divers domaines, sont de bons candidats comme suites chiffrantes dans le cadre des systèmes de chiffrement à flot : elles présentent en effet par nature des propriétés statistiques irréprochables (pour tout $k \leq n$, tous les motifs de longueur k sont équiprobables), et leur *complexité linéaire* est potentiellement grande (on pourrait produire ces suites avec des LFSRs, mais très longs : cela les rend résistantes à l'attaque de Berlekamp-Massey). Les techniques de construction connues présentent souvent des inconvénients, soit cryptographiques [8], soit algorithmiques [9]. De plus, les techniques de construction pré-citées ne permettent pas de générer une grande variété de suites. Or, si on souhaite les utiliser dans un système de chiffrement à flot, il ne faut pas qu'un attaquant puisse prédire quelle suite a servi à chiffrer le message, sinon il pourrait cryptanalyser le système sans difficulté. Lempel a publié en 1970 un procédé efficace et qui semble intéressant en cryptographie, construisant une suite d'ordre k à partir d'une suite d'ordre $k-1$, en gardant une vision locale de la suite, donc sans la stocker en totalité en mémoire. On peut, avec ce procédé, disposer d'un générateur de 2^n suites de de Bruijn, toutes distinctes, efficace en temps et en mémoire. De plus, un tel générateur ne sera pas exposé aux attaques algébriques, mentionnées plus haut, car par nature le système à résoudre lors de l'attaque sera trop complexe (fonction de transition non linéaire) et changeant. Nous avons apporté quelques améliorations à cette génération afin de réduire la mémoire nécessaire, et avons mené une étude sur la complexité linéaire précise des suites produites. Vincent Bénony a présenté dans sa thèse [Bén06] un générateur complet de cette nature, dont l'analyse semble tout à fait satisfaisante au regard des attaques connues. Ce générateur de suites de de Bruijn est entièrement décrit dans le manuscrit de thèse, et nous avions commencé à réfléchir à la construction d'un système de chiffrement à flot complet. Ces travaux sont restés en suspens, suite à l'interruption des travaux de Vincent Bénony (qui a quitté le milieu académique en 2006 à l'issue de sa thèse) et à mon changement de laboratoire (qui a donné lieu à l'interruption de mes travaux sur ce thème, l'IRISA ne souhaitant pas me voir poursuivre mes recherches en cryptographie). Il resterait donc à concevoir le reste du système, notamment l'intégration du vecteur d'initialisation.

8. Extension d'une suite de période 2^n-1 produite par un registre linéaire de longueur n, produisant une suite dont la grande partie possède une complexité linéaire faible par nature ; complétion d'une suite de n zéros de manière à produire tous les motifs, en utilisant la stratégie du « préfère 1 », ce qui donne lieu à un regroupement des 1 en début de suite [Fre82] ; fusion de petits cycles en partant de registres circulants [Fre82, Hua90] .

9. Les suites étant alors très lourdes à gérer, en temps ou en mémoire [Ann97, GG05] .

Nous avons également mené, en parallèle, une étude des T-fonctions, telles que proposées par Klimov et Shamir comme autre alternative aux LFSRs pour la génération de suites chiffrantes [KS02,KS04a,KS04b]. Ces fonctions sont utilisées comme fonctions de transition, et les auteurs proposent ensuite une fonction de filtrage qui permet de générer la suite pseudo-aléatoire à partir des états internes de l'automate. On notera n la longueur des états internes binaires de l'automate. Les auteurs présentent une attaque légère sur la fonction de filtrage telle que décrite dans leur premier article [10]. Nous avons réussi à proposer une attaque plus efficace que la leur, montrant que ce choix pour la fonction de filtrage présente de grands risquent en termes de sécurité. Plus précisément, notre attaque permet à un adversaire qui observe $O(2^{n/4})$ sorties consécutives du générateur de retrouver l'état interne complet associé à l'une de ces sorties, et donc de reconstituer tous les états internes puisqu'il connaît la fonction de transition. Cette attaque ne peut fonctionner que si l'un des états internes en question présente une structure particulière ; mais comme le générateur ne présente qu'un seul cycle pour les états internes (ce qui est d'ailleurs un très bon point pour le générateur, puisque la période est alors maximale), on sait qu'on va à un moment donné passer par un état interne favorable à l'attaque (après $2^{n/4}$ états générés, en moyenne). Notre algorithme présente une complexité en temps de $O(2^{n/4})$, et une complexité en mémoire de $O(n \log n)$. Une réalisation en MAPLE pour un générateur de taille $n = 64$ bits nous a permis de retrouver l'état interne en 30 secondes sur un Pentium cadencé à 2.4 Ghz. Pour $n = 128$ bits, on peut aboutir au même résultat en connaissant 2^{32} sorties du générateur.

Ce travail a été présenté à la conférence internationale *SETA : SEquences and Their Applications* en 2004 [BRWF05].

10. Notons que d'autres fonctions de filtrage sont par ailleurs proposées dans ces publications

Annexe B

Soutenance

Les transparents présentés lors de la soutenance sont donnés en à-plat (sans le détail des animations) dans les pages suivantes.

Assurer la sécurité des contenus multimédia, de leur création à leur diffusion
How to protect multimedia pieces of content, from their creation to their distribution

HDR defense

Caroline Fontaine

November, 28th 2011

Some security issues

Confidentiality
Integrity
Authentication
Cryptology

Errors
Error Correcting Codes

Alice

Subjective Integrity
Fragile watermarking

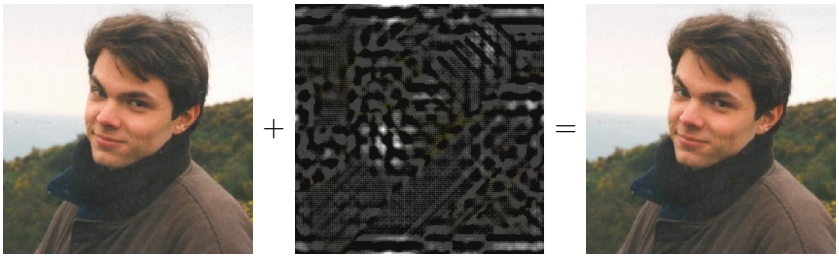
Stealthiness
Steganography

Bob

Copyright
Cryptology (DRM)
Robust watermarking (identify the copyright owner)
Fingerprinting (identify the legitimate end user)

Cryptology [1917-] Error Correcting Codes [1947-] Information Hiding [1990-]

Information Hiding in a nutshell

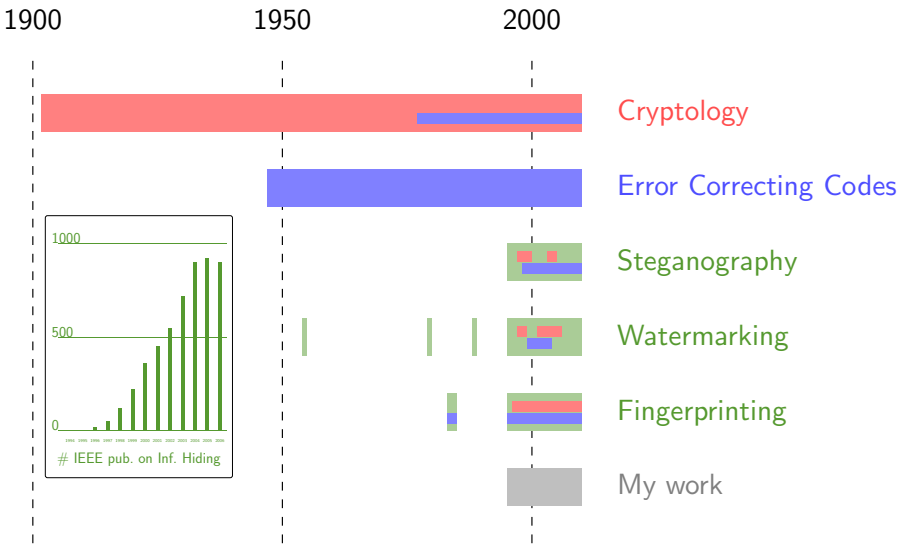


Trade-offs between : capacity, imperceptibility, robustness, security

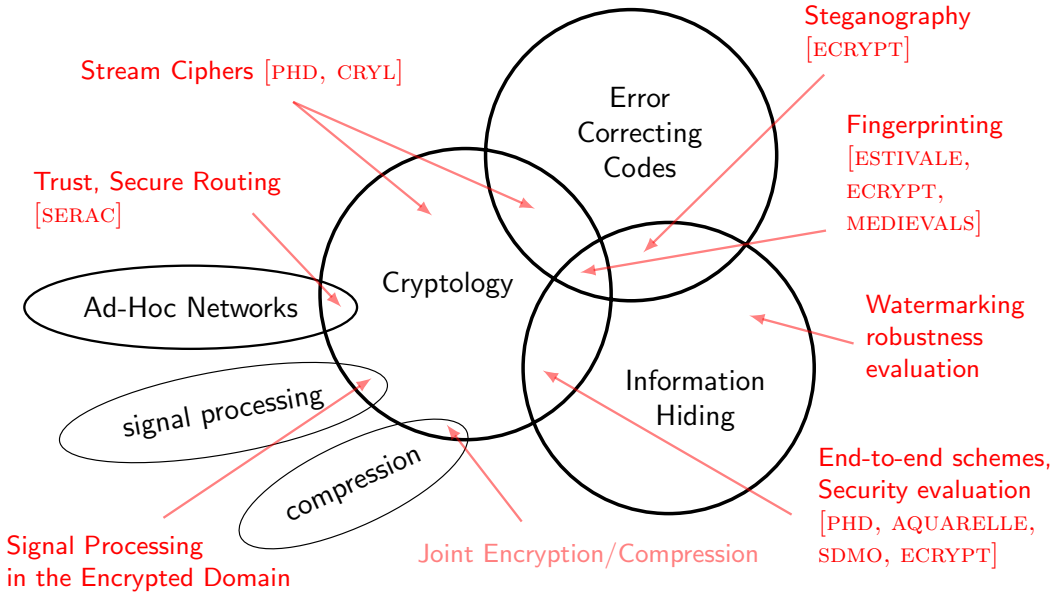
Steganography : (stealth com.)	+	++	-	++
Robust Watermarking : (id. cop. owner)	+	+	++	++
Fingerprinting : (id. end user)	+	+	++	++

+ important, - not important

Modern Evolutions and Crossings



Crossings and Contributions



Some contributions

Design (and attack) of stream ciphers, based on Highly Nonlinear Boolean Functions obtained with the help of Error Correcting Codes [FF98,Fon99,CCCF00,CCCF01,FFJ04,BRWF05] PhD V. Bénony [02-06]

Design of steganographic schemes based on Error Correcting Codes [FG07,FG09,ABF11] PhD M. Barbier [08-11]

Design of content protection architectures mixing cryptographic and watermarking primitives [AFD98,ABD+99,ABTD+06,FDD+08]

Transposed cryptanalysis methodology to the study of the security of watermarking schemes [FR02,CFF05b,CFF05e]

Improvement of the robustness and security of `Broken Arrows` watermarking technique [CXFF09,XFF10a,XFF10b] PhD F. Xie [07-10]

Design of fingerprinting schemes based on a watermarking layer and an anti-collusion code [XFF08,CXFF09,CFF10,CFFC11] PhD F. Xie [07-10], PhD A. Charpentier [08-11]

Outline

1 Context

2 **Contributions in Steganography**
- Interest of Reed-Solomon codes (IH 2007)
- A randomized scheme to ensure embedding (IMACC 2011)

3 Contributions in Fingerprinting

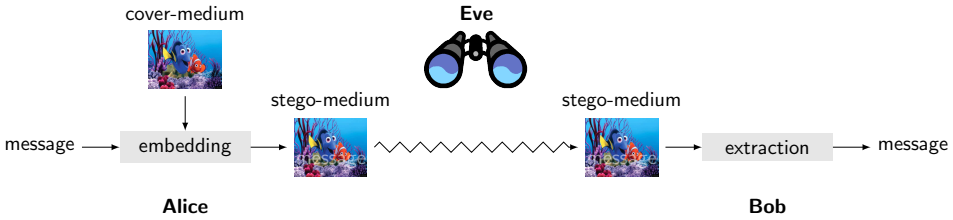
4 General Conclusion and further work

The Warden is watching . . .

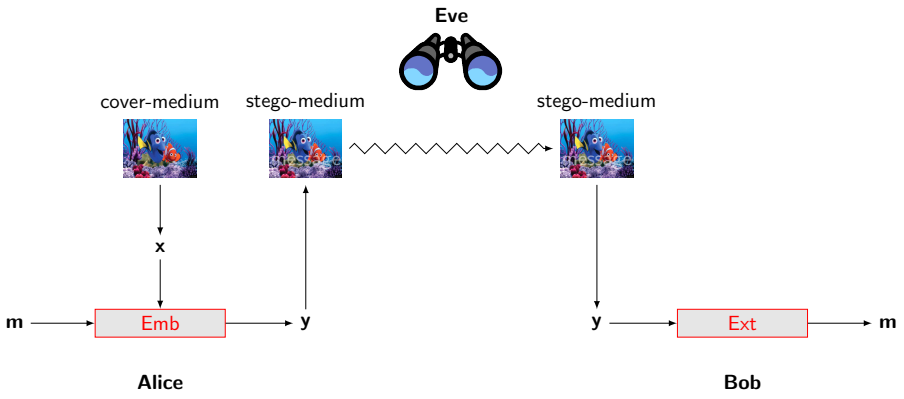
The Prisoners and the Passive Warden [Sim83] :

Alice and Bob want to send each other some important secret messages. Eve keeps a watch on. If she suspects something is going wrong, she interrupts the communication.

⇒ Alice and Bob must exchange only innocuous looking documents! They cannot rely only on cryptograhy, they need a steganographic scheme.



Steganographic schemes : design issues

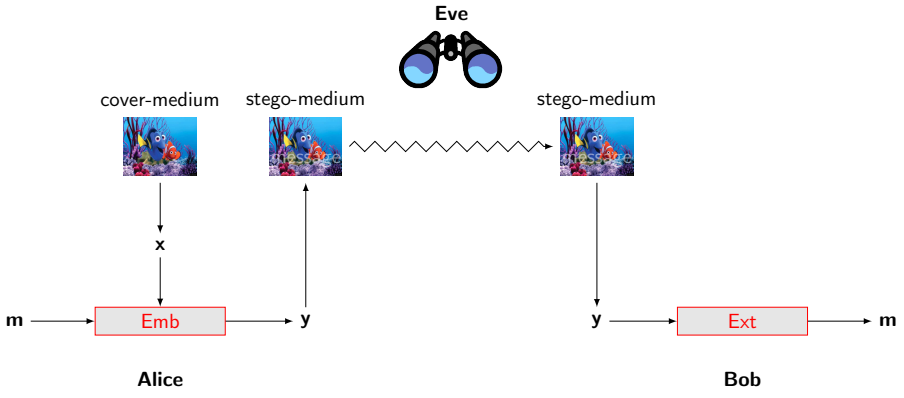


Critical choices to prevent steganalysis (no perfect security currently achievable) :

- Which vectors to derive from the medium ?
- How to process them with Emb and Ext ?

$\Rightarrow$ One strategy : to minimize distorsion, one way : syndrome coding

Minimizing distorsion with syndrome coding



One strategy : to minimize distorsion , one way : syndrome coding (e.g. F5 [Wes01])

$\mathrm{Ext}(\mathrm{Emb}(\mathbf{x}, \mathbf{m})) = \mathbf{m}$

$d_H(\mathbf{x}, \mathrm{Emb}(\mathbf{x}, \mathbf{m})) \leq T$

/* $\mathrm{Emb}(\mathbf{x}, \mathbf{m})_i = \mathbf{x}_i \ \forall i \in \mathcal{W}$ wet paper [FGLS05] */

$\mathrm{Emb}(\mathbf{x}, \mathbf{m})$ of syndrome $\mathbf{m}$
$\mathrm{Ext}(\mathbf{y})$ = syndrome of $\mathbf{y}$

Syndrome coding, more formally

Syndrome coding has been introduced and discussed in [Cra98,Bie01], and properly formalized in [GK03,GK09]. It has been widely studied.

Let $\mathcal{C}$ be a q-ary linear code of length n, dimension k and parity check matrix $\mathbf{H}$: $\mathcal{C} = \{\mathbf{c} \mid \mathbf{c} \cdot \mathbf{H}^t = 0\}$ is a vector subspace of $\mathbb{F}_q^n$ of dimension k.

$$\begin{array}{rrcll} \mathrm{Emb}(\mathbf{x},\mathbf{m}) & = & \mathbf{x} + D(\mathbf{m} - \mathbf{x}\cdot\mathbf{H}^t) & \leftarrow \mathrm{Ext}(\mathrm{Emb}(\mathbf{x},\mathbf{m})) = \mathbf{m}, d_H() \leq T \\ \mathrm{Ext}(\mathbf{y}) & = & \mathbf{y}\cdot\mathbf{H}^t & \leftarrow \text{syndrome of } \mathbf{y} \\ /* \quad \mathbf{y}_i & = & \mathbf{x}_i \qquad \forall i \in \mathcal{W} & \text{wet paper } */ \quad \text{even harder !} \end{array}$$

$D()$ must return $\mathbf{e}$, $d_H(\mathbf{e}, 0) \leq T$, of syndrome $\mathbf{e}\cdot\mathbf{H}^t = \mathbf{m} - \mathbf{x}\cdot\mathbf{H}^t$.

- Does such a vector $\mathbf{e}$ exist ?	Not always (depends on $\mathcal{C}$)
- How to find it ?	A matter of decoding
- May we choose between several vectors $\mathbf{e}$?	If list decoding is possible

$\Rightarrow \mathcal{C}$ must be chosen really carefully

Which code should we use ?

A lot of codes have been studied : Hamming, BCH, Convolutional, etc

Which criteria have been addressed ?

- Embedding efficiency (heavily studied, optimal codes)
- Probability of success (almost never addressed)
  - Dry paper : success is ensured only for perfect codes (Hamming and Golay, but their embedding efficiency is not good)
  - Wet paper : success is ensured only for MDS [q-ary] codes

  When success is not ensured, the probability of success decreases exponentially with the message length !

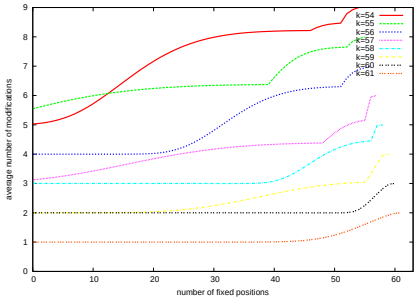
Contributions :

- Reed-Solomon codes (MDS, list decoding) [FG07,FG09]
- A variant of syndrome coding, that ensures embedding success [ABF11] PhD M. Barbier [08-11]

How Reed-Solomon codes can help

With F. Galand [FG07,FG09] (IH 2007)

- ✓ Good parameters (*e.g.* covering radius)
- ✓ MDS (embedding is ensured in the wet paper context)
- ✓ Unique decoding (Lagrange) + List decoding (Guruswami-Sudan)



Estimated Gain of List decoding.
$q = 64$, $n = 63$, Plot only $\Delta\omega \geq 0.3$

- ✓ List decoding → gain in average embedding efficiency
- ☹ Guruswami-Sudan is hard to implement
- ↯ implementation of Guruswami-Sudan
- ↯ must derive q-ary vectors $\mathbf{x}$ from the media

Randomized Syndrome Coding

With D. Augot and M. Barbier [ABF11] (IMACC 2011)

"How can we design a scheme that ensures embedding ?"

Our idea : randomize a part of the syndrome,

replacing $\mathbf{y} \cdot \mathbf{H}^t = \mathbf{m}$

by $\mathbf{y} \cdot \mathbf{H}^t = (\mathbf{m}||\mathbf{R})$

- ✓ Embedding success, even in the wet paper context
- ✓ We provided a way to send the length of $\mathbf{R}$ to the recipient
- ☹ Loss in embedding efficiency (vs. traditional synd. coding)
- ✓ $[\frac{q^p-1}{q-1}, n-p, 3]$ Hamming codes :
  the relative loss in embedding efficiency is only $\frac{\lceil \log_q((q-1)\#\mathcal{W}+1)\rceil}{p}$
- ↯ Must be studied further

Steganography : conclusion and further work

We focused on the success on the embedding,
while preserving a good embedding efficiency.

Reed-Solomon codes :

- ✓ RS could (should ?) be used in practical schemes
- ↯ Native q-ary steganography should be studied
- ↯ Implementation of Guruswami-Sudan list decoding

Randomized Syndrome Coding :

- ↯ Needs to be further studied

Other tracks :

- ↯ Active Warden

Outline

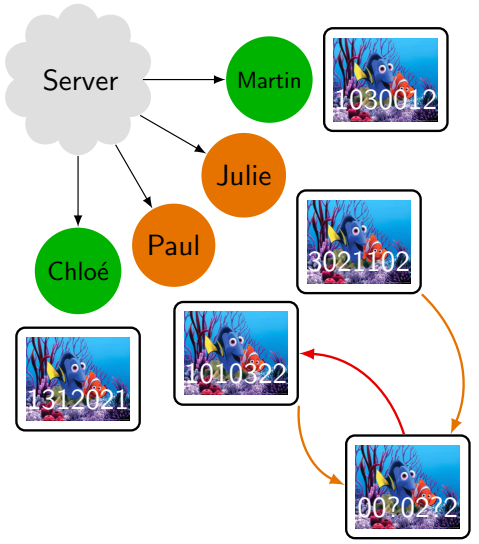
1 Context

2 Contributions in Steganography

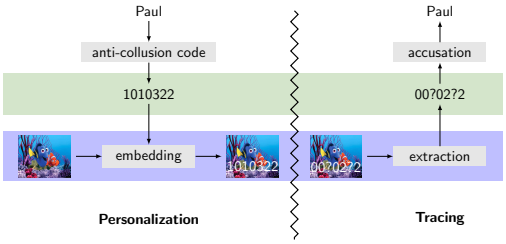
3 **Contributions in Fingerprinting**
• How to provoque multiple detections with `Broken Arrows`, and manage them with Tardos codes (MM&Sec 2008)
• Estimation of the pirates' strategy and optimization of Tardos' score computation (EI 2009 + TS 2010)
• Design of an asymmetric fingerprinting protocol dedicated to Tardos codes (IH 2011)

4 General Conclusion and further work

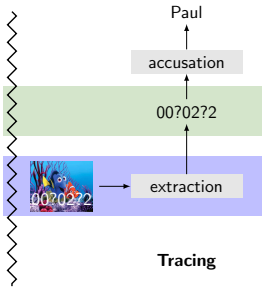
How to prevent illegal redistribution ?



- Cryptography is not sufficient
- A need for watermarking
- A need for an anti-collusion code with a structure enabling tracing



How to link models with reality



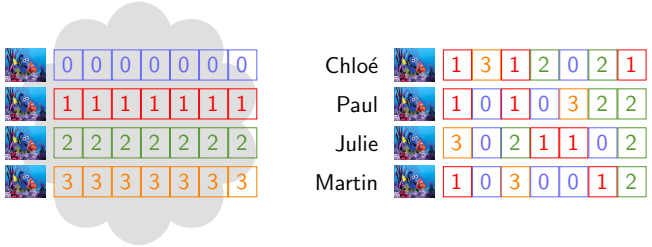
message layer

signal layer

Boneh & Shaw introduced in 1995 a model which remains the most used today (with its extensions)

- ✓ Simple to express
- ✓ Has been intensively studied
- ☹ Not so realistic
- ✓ BUT we can force reality to fit the model :

off-line block-based watermarking and on-line switching



Attacks and assumptions

 1 0 1 0 3 2 2 $\mathbf{X}_{\text{Paul}}$

 3 0 2 1 1 0 2 $\mathbf{X}_{\text{Julie}}$

 1 0 3 0 0 1 2 $\mathbf{X}_{\text{Martin}}$

Collusion (c users among n)

 1 0 3 1 0 2 2 $\mathbf{Y}$ Copy/paste blocks (*e.g.* random, maj., etc)

 0 0 ? 0 2 ? 2 $\mathbf{Y}$ Fusion of blocks (*e.g.* averaging)

Boneh & Shaw : Marking Assumption $X_{j_1 i} = \ldots = X_{j_c i} = a \Rightarrow Y_i = a$.

 ? 0 1 2 ? 1 ? $\mathbf{Y}$ Individual signal processing (*e.g.* compression)

To prevent errors and erasures, watermarking must be as robust as possible.

In the steps of Boneh & Shaw

Strong traceability : $\mathbb{P}$(**accuse an innocent user**) $= 0$

- ⊙ error correcting codes
- ☹ $n \geq 3, c \geq 2$: only copy/paste attacks
- ☹ codes too long, on huge alphabets [HvLLT98,BCE+01,SSW01]

Weak traceability : $\mathbb{P}$(**accuse an innocent user**) $< \varepsilon$

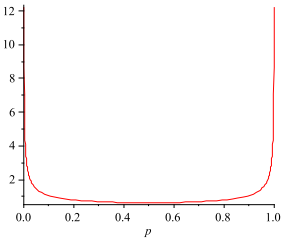
- ⊙ error correcting codes + probabilistic codes
- ✓ copy/paste + fusion attacks
- ⊙ Peikert's bound [PSS03][Tar03,Tar08] : $m \geq \mathcal{O}(c^2 \ln(n/\varepsilon))$
- ✓ first codes to meet the bound : Tardos codes

Binary Tardos code [Tar03]+[SKC08]

n, $1 << c$, $\varepsilon_1 << \varepsilon_2$, $m = 2\pi^2 c^2 \lceil \ln(1/\varepsilon_1) \rceil$, $Z = 2\pi c \lceil \ln(1/\varepsilon_1) \rceil$.
m secret probabilities p_i drawn according to the pdf $f(p) = \frac{1}{\pi\sqrt{p(1-p)}}$.

	p_1	p_2	p_3	...	p_m
X_1	1	0	1	...	0
X_2	0	1	0	...	1
X_3	1	1	0	...	1
⋮	⋮	⋮	⋮	⋮	⋮
X_n	0	0	0	...	1

with $\mathbb{P}(X_{ij} = 1) = p_i$

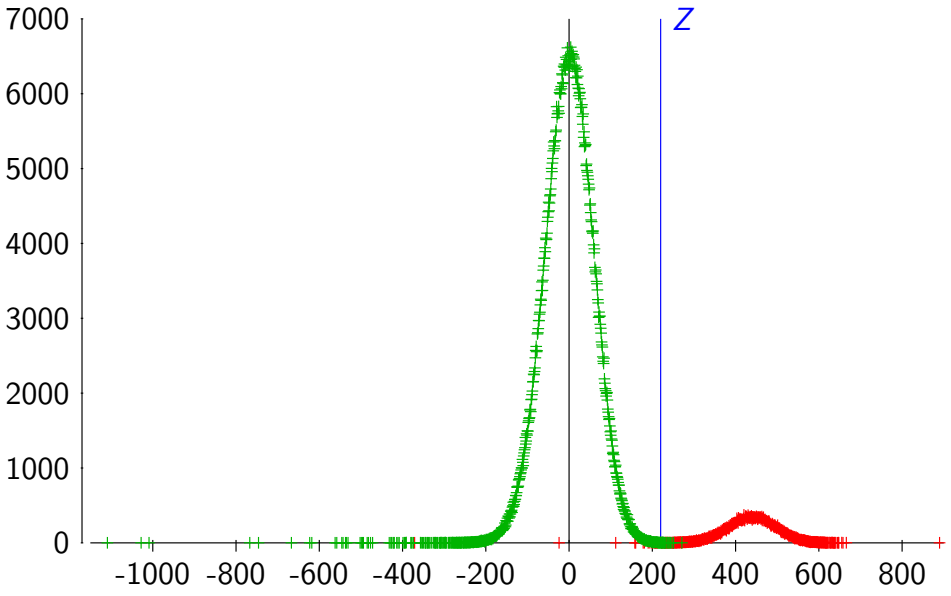


Accusation : User j's score $S_j = \Sigma_{i=1}^{m} g(Y_i, X_{ji}, p_i) >^? Z$

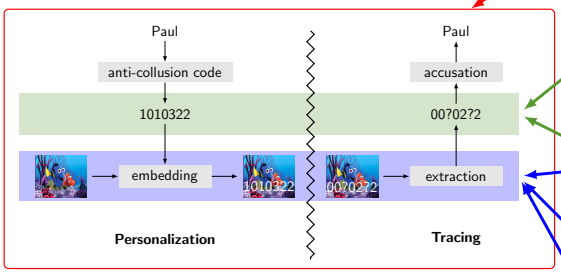
Y		1		1		0	...	1
X_j		1		0		1	...	0
S_j	=	$g(1,1,p_1)$	+	$g(1,0,p_2)$	+	$g(0,1,p_3)$	... +	$g(1,0,p_m)$

$g(1,1,p) = g(0,0,1-p) = \sqrt{(1-p)/p}$ $\quad$ $g(1,0,p) = g(0,1,1-p) = -\sqrt{p/(1-p)}$

Tardos codes have been studied a lot



Contributions on fingerprinting



+ Implementation of FANTOMAS platform
M. Desoubeaux [07-09]

Design of an asymmetric fingerprinting protocol dedicated to Tardos codes (IH 2011)

Estimation of the pirates' strategy, and optimization of Tardos' scores computation (EI 2009 + TS 2010)

How to provoque multiple detections with `Broken Arrows`, and manage them with Tardos codes (MM&Sec 2008)

Improvement of `Broken Arrows` robustness (EI 2009)

Improvement of `Broken Arrows` security (MM&Sec 2010)

PhD A. Charpentier [08-11]

PhD F. Xie [07-10]

`Broken Arrows` + Tardos : a good match

With F. Xie and T. Furon [XFF08] (MM&Sec 2008)

Problem : "Fusion attacks are critical, and easy to perform."

Our idea : if the embedding technique is sufficiently robust, one can be able to detect multiple symbols in case of a fusion attack (e.g. averaging).

- `Broken Arrows` is a very robust zero-bit watermarking technique, designed in 2007 for BOWS-2 contest.
- We adapted it to embed q-ary symbols, and combined it with a q-ary Tardos code ($q = 4$).

$\Rightarrow$ Fusion attacks really lead to multiple symbols detections.

But Tardos codes were not designed to take them into account ...

- We modified the score computation to take them into account.

$\Rightarrow$ It worked really well (and even better than we thought).

Broken Arrows + Tardos : a good match

q-ary "Tardos" codes [SKC08] :

Each $\mathbf{p}_i = (p_i^0, \ldots, p_i^{q-1}) \sim$ Dirichlet distribution of shape parameter κ

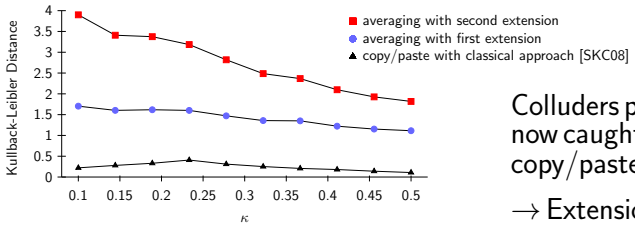
Generation $\mathbb{P}(X_{ji} = a) = p_i^a$

Score $S_j = \sum_{i=1}^{m} \delta_{Y_i = X_{ji}} g_1(p_i^{Y_i}) + (1 - \delta_{Y_i \neq X_{ji}}) g_0(p_i^{Y_i})$

We proposed two different extensions to take advantage of $\mathcal{Y}_i = \{Y_i^1, \ldots, Y_i^{L_i}\}$:

$$S_j = \sum_{i=1}^{m} \sum_{\ell=1}^{L_i} \delta_{Y_i = X_{ji}} g_1(p_i^{Y_i}) + (1 - \delta_{Y_i \neq X_{ji}}) g_0(p_i^{Y_i})$$

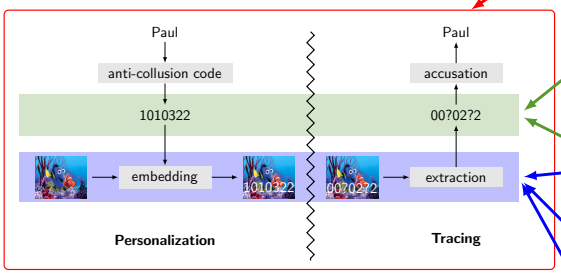
$$S_j = \sum_{i=1}^{m} \delta_{X_{ji} \in \mathcal{Y}_i} g_1\left(\sum_{\ell=1}^{L_i} p_i^{Y_i^\ell}\right) + (1 - \delta_{X_{ji} \notin \mathcal{Y}_i}) g_0\left(\sum_{\ell=1}^{L_i} p_i^{Y_i^\ell}\right)$$



Colluders performing an averaging are now caught more efficiently than for a copy/paste attack!

→ Extension of this work in [SKSC11].

Contributions on fingerprinting



+ Implementation of FANTOMAS platform

M. Desoubeaux [07-09]

Design of an asymmetric fingerprinting protocol dedicated to Tardos codes (IH 2011)

Estimation of the pirates' strategy, and optimization of Tardos' scores computation (EI 2009 + TS 2010)

PhD A. Charpentier [08-11]

How to provoque multiple detections with Broken Arrows, and manage them with Tardos codes (MM&Sec 2008)

Improvement of Broken Arrows robustness (EI 2009)

PhD F. Xie [07-10]

Improvement of Broken Arrows security (MM&Sec 2010)

Dynamical optimization of Tardos' scores

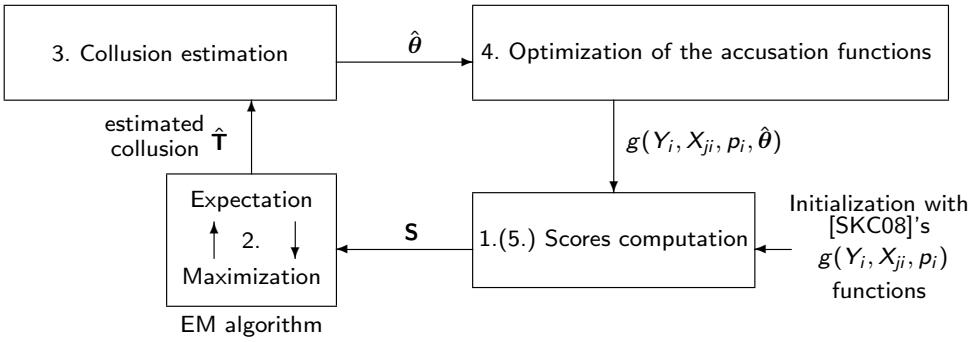
With A. Charpentier and T. Furon [CXFF09] (EI 2009), [CFF10] (TS 2010)

"Are [Tar03,Tar08,SKC08]'s parameters and functions the best ones ?"

- **Tardos [Tar03,Tar08] and Škorić et al. [SKC08]** : for a given c, the scores distributions, $\mathcal{N}_I = \mathcal{N}(0, \sigma_I^2)$ and $\mathcal{N}_C = \mathcal{N}(\mu_C, \sigma_C^2)$, remain the same whatever the colluders' strategy.
- **Furon et al. [FGC08]** :
  - When the colluders' strategy is not known, [Tar03,Tar08,SKC08]'s choices lead to the maximal Kullbach-Leibler Distance between $\mathcal{N}_I$ and $\mathcal{N}_C$. (binary case)
  - BUT if we know the colluders' strategy, we can derive functions $g(Y_i, X_{ji}, p_i)$ leading to a higher Kullbach-Leibler Distance between $\mathcal{N}_I$ and $\mathcal{N}_C$. (binary case)

Dynamical optimization of Tardos' scores

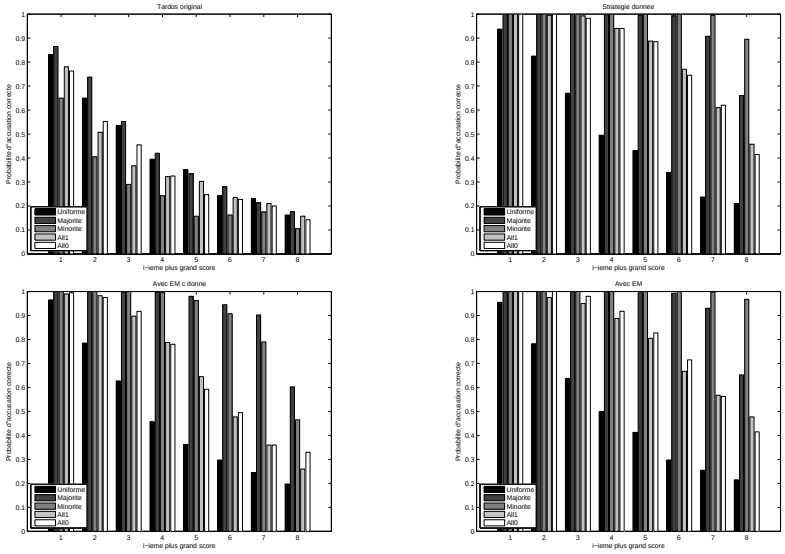
We pushed further, providing a better optimization of the scores, and a way to estimate the colluders' strategy $\theta = \{\mathbb{P}(Y_i = 1 | \Sigma_i = \sigma_i), \sigma_i = 0..c\}_{i=1..m}$.
Assumption : the strategy is the same for all the components



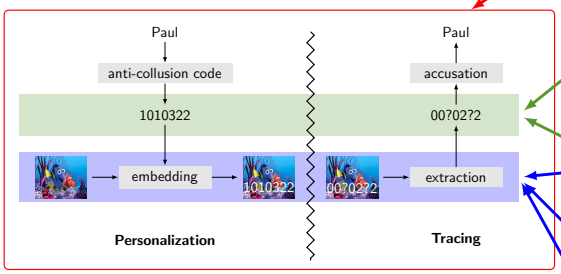
Dynamical optimization of Tardos' scores

$\mathbb{P}(k\text{-th highest score is guilty})$ $k \in \{1,\dots,8\}$, $m = 1000$, $c = 8$, $n = 5000$.

[SKC08] / known c, θ / given c, estimated $\hat{\theta}$ / all are estimated



Contributions on fingerprinting



+ Implementation of FANTOMAS platform

M. Desoubeaux [07-09]

Design of an asymmetric fingerprinting protocol dedicated to Tardos codes (IH 2011)

Estimation of the pirates' strategy, and optimization of Tardos' scores computation (EI 2009 + TS 2010)

How to provoque multiple detections with `Broken Arrows`, and manage them with Tardos codes (MM&Sec 2008)

Improvement of `Broken Arrows` robustness (EI 2009)

Improvement of `Broken Arrows` security (MM&Sec 2010)

PhD A. Charpentier [08-11]

PhD F. Xie [07-10]

How to prevent parties from cheating

With A. Charpentier, T. Furon and I. Cox [CFFC11] (IH 2011)

"Can we trust the provider who delivers the pieces of content ?"

In the usual (symmetric) scenario ...

- An untrustworthy provider may frame an innocent buyer !
- Any accused buyer can argue he/she has been framed by an untrustworthy provider !

Asymmetric fingerprinting protocols have been introduced in [PS96].

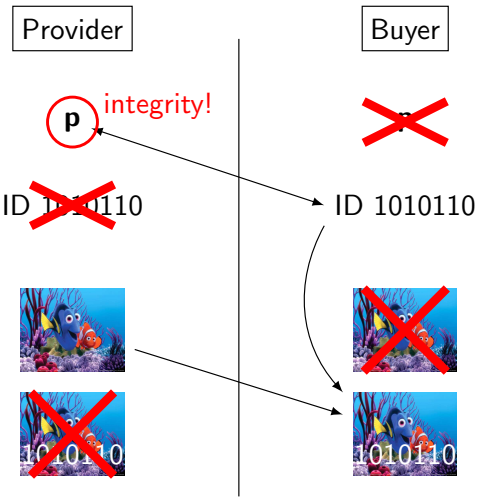
- Most of them (not ours) also provide anonymity of the Buyer.
- Very few also (not ours) provide privacy on the delivered content.

✓ Embedding and tracing techniques are sufficiently mature today to provide complete specifications for such protocols.
☹ No existing protocol is compliant with Tardos codes.

How to prevent parties from cheating

Designing a protocol based on Tardos codes : rules and challenges.

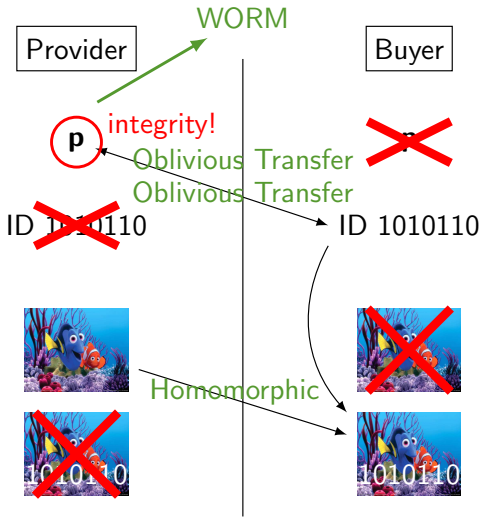
1. Generation of the ID (fingerprint)
2. Embedding



How to prevent parties from cheating

Designing a protocol based on Tardos codes : rules and challenges.

1. Generation of the ID (fingerprint)
2. Embedding



Fingerprinting : conclusion and further work

All aspects addressed : watermarking, anti-collusion code, protocols.

Tardos codes :

- ✓ Multiple detection for q-ary Tardos codes.
- ✓ Estimation of the colluders' strategy, and optimization of the accusation for binary Tardos codes.
- ↯ Optimization should be extended to q-ary Tardos codes (really hard).

Asymmetric fingerprinting protocol :

- ✓ First protocol compliant with binary Tardos codes.
- ↯ Proofs ?
- ↯ Anonymity and Privacy ?

Outline

1 Context

2 Contributions in Steganography

3 Contributions in Fingerprinting

4 General Conclusion and further work

General conclusion and further work

Crossings offer new points of views, new ideas, and a complete overview.

- ✓ Cryptanalysis methodology applied to the definition and study of watermarking security.
- ✓ Syndrome Coding in Steganography.
- ✓ An asymmetric fingerprinting protocol, with all primitives detailed.

My favorite prospects :

- ↯ Syndrome Coding in Steganography.
- ↯ Asymmetric fingerprinting protocol : proofs, commutative encryption, anonymity/privacy.
- ↯ Anonymity issues in general.
- ↯ Implementation of homomorphic encryption schemes.
  PhD S. Fau [11-14]

Annexe C

Curriculum Vitæ, et liste de publications

C.1 Curriculum Vitæ

Parcours académique

depuis 2009	**Chargée de recherche** (CR1) au CNRS, affectée au LABoratoire en Sciences et Technologies de l'Information, de la Communication et de la Connaissance – Lab-STICC (UMR 3192, Brest).
2005-2009	**Chargée de recherche** (CR2 puis CR1) au CNRS, affectée à l'Institut de Recherche en Informatique et Systèmes Aléatoires – IRISA (UMR 6074, Rennes).
2002-2005	**Chargée de recherche** (CR2) au CNRS, affectée au Laboratoire d'Informatique Fondamentale de Lille – LIFL (UMR 8022, Lille).
1999-2002	**Maître de Conférences** à l'UFR IEEA, Laboratoire d'Informatique Fondamentale de Lille – LIFL (UMR 8022, Lille), Université Lille 1.
1999	**Attachée Temporaire d'Enseignement et de Recherche** pendant 3 mois à l'UFR IEEA, Laboratoire d'Informatique Fondamentale de Lille – LIFL (UMR 8022, Lille), Université Lille 1.
1998-1999	**Attachée Temporaire d'Enseignement et de Recherche** à l'UFR d'informatique, Laboratoire de Recherche en Informatique – LRI (UMR 8623, Orsay), Université Paris 11.
1995-1998	**Allocataire de Recherche** (Université de Paris 6 et Institut National de Recherche en Informatique et Automatique - INRIA Rocquencourt) et **Moniteur** (Université de Cergy Pontoise). Thèse effectuée à l'INRIA Rocquencourt, sous la direction de Pascale Charpin. *Contribution à la recherche de fonctions booléennes de haute non-linéarité, et au marquage d'images en vue de la protection des droits d'auteur.*

Diplômes

nov. 1998	**Thèse de doctorat** de l'Université Paris 6, spécialité informatique, algorithmique. Mention Très Honorable.
juin 1995	**DEA** *Informatique, Mathématiques et Applications* (IMA). Université Paris 11. Mention Bien.
juin 1994	**Maîtrise** *Mathématiques et Applications aux Sciences Fondamentales* (MASF). Université Paris 11
juin 1993	**Licence** *Mathématiques et Applications Fondamentales* (MAF). Université Paris 11
juin 1992	**Deug** *Sciences et Structure de la Matière* (Deug A). Université Paris 11. Module « Mathématiques-Informatique ».
juin 1990	**Baccalauréat** *Mathématiques et Sciences Physiques* (Bac C).

Projets et contrats

MEDIEVALS [2007-2011, participant, responsable scientifique pour l'IRISA] *Marquage et Embrouillage pour la Diffusion et les Échanges Vidéos et Audio Légalisés et Sécurisés*, ANR RIAM 2007. Le contexte de ce projet industriel était celui de la vente de vidéo à la demande sur Internet. L'objectif de ce projet était de renforcer la sécurité d'un système de distribution proposé par Medialive, dans lequel les vidéos étaient distribuées sous une forme en partie embrouillée, pour y ajouter des fonctionnalités de traçage. L'utilisateur recevant les moyens de désembrouiller la vidéo une fois l'achat effectué, il doit alors se retrouver en possession d'une vidéo en clair mais traçable. Mes travaux sur le fingerprinting ont trouvé leur application dans le cadre de ce projet, et la thèse d'Ana Charpentier a été financée par ce projet. Les autres partenaires ont été les laboratoires académiques LSS (UMR CNRS), le GET/Télécom Sud-Paris et les industriels Thomson-Civolution, Medialive-Nagra, Amossys.

ESTIVALE [2006-2009, participant, responsable scientifique pour l'IRISA] *Échanges Sécurisés pour le Transfert d'Informations Vidéo, en Accord avec la Législation et l'Économie*, ANR RIAM 2005. Le contexte de ce projet était celui de la vente de vidéo à la demande sur Internet, avec des mécanismes de sécurisation évitant le piratage. Ce projet a donné lieu à une forte réflexion quant aux modèles économiques et juridiques associés à ce contexte, et a permis d'établir une véritable synergie entre économistes-juristes-informaticiens. J'ai été responsable du sous-projet *architecture système du DRM*. Les autres partenaires étaient les laboratoires académiques LIS (UMR CNRS), LSS (UMR CNRS), ADIS (spécialisé dans l'économie industrielle), CERDI (spécialisé dans le droit et la propriété intellectuelle de l'immatériel) et les industriels Nextamp-Thomson, Basic Lead (spécialisé dans l'organisation de marchés de vente de droits audiovisuels), SACD (société de gestion de droits d'auteur).

Dhimyotis [2005, expertise] Le LIFL a effectué, par mon intermédiaire, du conseil pour l'entreprise DHIMYOTIS qui s'est créée début 2005. Ce contrat avait pour objectif d'accompagner scientifiquement cette *start up*, qui développait des outils/produits permettant à l'utilisateur de s'identifier auprès de sa machine de bureau en utilisant son téléphone portable comme terminal d'authentification. L'entreprise prévoyait de développer à long terme des procédés de sécurisation de bases de données. Le stage de DEA que j'ai dirigé en 2004 portait justement sur cette thématique.

SERAC [2004-2007, coordinateur] *modèles et protocoles de SEcurité pour les Réseaux Ad hoC*, ACI Séccurité Informatique 2004. J'ai été coordinatrice de ce projet, dont l'objet était l'élaboration des modèles d'analyse de sécurité pour les réseaux mobiles de type ad hoc, c'est-à-dire sans infrastructure fixe, et la conception de protocoles/procédés de sécurisation. Les autres partenaires étaient le GET (INT et ENST) et l'INRIA (projets CODES, TANC et HIPERCOM).

ECRYPT [2004-2008, participant] réseau d'excellence européen (6ème PCRD IST-2002-507932). Ce réseau, qui regroupait de très nombreux partenaires, visait à dynamiser la recherche en cryptographie et dissimulation d'information en Europe. J'ai participé au laboratoire virtuel WAVILA (tatouage). Les partenaires sont trop nombreux pour être énumérés ici : équipes industrielles et académiques françaises, allemandes, italiennes, ... WAVILA a regroupé la plupart des laboratoires européens travaillant sur le tatouage de documents numériques, et a abordé des sujets tels que : les fondements, la sécurité, les différents contextes et objectifs, les différents média. Mes travaux sur la stéganographie, le fingerprinting, le tatouage ont été soutenus par ce réseau. J'ai par ailleurs organisé en juin 2007 un workshop ECRYPT, Wacha 2007.

SDMO [2003-2006, participant, responsable scientifique pour le LIFL] *Sécurisation de la Diffusion de Musique sur les mObiles*, RNRT 2002. Ce projet précompétitif visait à proposer une architecture permettant la diffusion sécurisée de musique sur les téléphones mobiles de troisième génération. Un démonstrateur a été présenté au Ministère de l'Industrie à la fin du projet, en juin 2006. J'ai coordonné l'activité du sous-projet *sécurité*. J'ai également participé au sous-projet *architecture* ainsi qu'au sous-projet *tatouage*. Les autres partenaires étaient le laboratoire LSS (UMR CNRS), France Télécom R&D, Orange, Oberthur CS, Faith Technologies (anciennement Digiplug).

COM [2002-2005, participant] *Communication haut débit pour Objets Mobiles*, soutenu par l'IRCICA (Institut de Recherche sur les Composants logiciels et matériels pour l'Information et la Communication Avancée – institut fédératif regroupant le LIFL, le CERLA et l'IEMN). Ce projet regroupait des chercheurs du LIFL et de l'IEMN afin de concevoir des moyens de communication adaptés aux réseaux (sans fil) d'objets mobiles.

Nouvelles méthodes de tatouage et de dissimulation de données pour des communications audiovisuelles sécurisées [2002, participant] Action Spécifique du département STIC du CNRS. Cette action a eu pour objectif de stimuler la recherche française en tatouage.

CRYL [2001-2003, coordinateur] *CRYptographie à Lille*, ACI Cryptologie 2000. L'objectif de ce projet était de créer, suite à mon recrutement à Lille, une synergie entre les chercheurs intéressés par la cryptographie. Partenaires : Université Lille 1, au travers des laboratoires d'informatique (LIFL) et de mathématique (AGAT) de Lille, et de l'UFR IEEA.

Aquarelle [1996-1998, participant] *Sharing Cultural Heritage through Multimedia Telematics*, Telematics Application Programme of the European Union. Ce projet avait pour objectif de rendre les bases de données européennes portant sur les patrimoines culturels des pays participants interrogeables comme une seule et

unique base. Les photographies numériques à haute résolution des œuvres d'art ou des monuments diffusées sur les serveurs devaient être protégées contre le piratage, grâce à une articulation entre outils cryptographiques et tatouage.

Directions de thèses et post-doctorats

S. Fau (2011-2014) Je co-dirige avec Guy Gogniat (de l'Université Bretagne Sud, Lab-STICC) et Renaud Sirdey (du CEA Saclay) la thèse de Simon Fau, commencée en octobre 2011. Cette thèse porte sur l'implémentation matérielle des techniques de chiffrement homomorphiques.

M. Barbier (2008-2011) J'ai co-dirigé avec Daniel Augot, de l'équipe TANC de l'INRIA Centre Saclay Ile-de-France, la thèse de Morgan Barbier, intitulée *Décodage en liste et application à la sécurité de l'information* et soutenue le 2 décembre 2011. Ses travaux ont porté sur le décodage des codes correcteurs, et leur lien avec la stéganographie.

A. Charpentier (2008-2011) J'ai dirigé (avec dérogation de l'école doctorale), en co-encadrement avec Teddy Furon de l'équipe TEMICS de l'IRISA/INRIA Centre Rennes Bretagne Atlantique, la thèse d'Ana Charpentier, intitulée *Identification de copies de documents multimédia grâce aux codes de Tardos* et soutenue le 21 octobre 2011. Ses travaux ont porté sur la conception de systèmes d'identification de copies dans des applications de distribution de documents multimédia (vidéo à la demande).

F. Xie (2007-2010) J'ai dirigé (avec dérogation de l'école doctorale), en co-encadrement avec Teddy Furon de l'équipe TEMICS de l'IRISA/INRIA Centre Rennes Bretagne Atlantique, la thèse de Fuchun Xie, initulée *Tatouage sûr et robuste appliqué au traçage de documents multimedia* soutenue le 23 septembre 2010. Ses travaux ont porté sur la conception de systèmes de traçage de traitres (fingerprinting) dans des applications de distribution de documents multimédia (vidéo à la demande), et l'amélioration de la robustesse et de la sécurité de la technique de tatouage Broken Arrows.

F. Galand (2006-2007) J'ai dirigé à l'IRISA le séjour post-doctoral de Fabien Galand, qui a travaillé sur le projet RIAM ESTIVALE du 1er septembre 2006 au 30 avril 2007. Nous avons mené ensemble des recherches sur l'utilisation des codes correcteurs d'erreurs en stéganographie. En parallèle, nous avons rédigé un article présentant les techniques de chiffrement homomorphiques pour les non-spécialistes.

J.C. Hernandez-Castro (2004-2005) J'ai dirigé au LIFL le séjour post-doctoral de Julio César Hernandez Castro, chercheur espagnol qui a travaillé sur le projet RNRT SDMO, du 1er octobre 2004 au 30 juin 2005.

V. Bénony (2002-2006) J'ai co-encadré avec J.-M. Geib (directeur du laboratoire LIFL) la thèse de Vincent Bénony, intitulée *Étude et conception de systèmes de chiffrement à flot dans le contexte d'architectures matérielles fortement contraintes* et soutenue le 29 juin 2006. Vincent Bénony a étudié de nouvelles

primitives pour les systèmes de chiffrement symétriques, abordant tour à tour la conception de nouveaux schémas et l'attaque de schémas existants.

Participation à des jurys de thèse

- *Décodage en liste et application à la sécurité de l'information*, M. Barbier, Ecole Polytechnique, décembre 2011 (co-directrice).
- *Identification de copies de documents multimédia grâce aux codes de Tardos*, A. Charpentier, Université de Rennes 1, octobre 2011 (directrice).
- *Cryptography based Visual Data Protection*, N. Islam, Université de Montpellier II, juillet 2011 (rapporteur).
- *Robust Multichannel Perceptual Color Image Watermarking and Private Anonymous Fingerprinting*, W. Abdul, Université de Poitiers, janvier 2011 (examinatrice).
- *Elaboration de nouveaux algorithmes de crypto-compression basés sur le codage arithmétique*, A. Masmoudi, Université de Montpellier II et Université de Sfax (Tunisie), décembre 2010 (examinatrice).
- *Tatouage sûr et robuste appliqué au traçage de documents multimédia*, F. Xie, Université de Rennes 1, septembre 2010 (directrice).
- *Tatouage de données géographiques et généralisation aux données devant préserver des contraintes*, C. Bazin, Université de Caen, janvier 2010 (rapporteur).
- *Quantization-based blind watermarking of three-dimensional meshes*, K. Wang, INSA Lyon, novembre 2009 (examinateur).
- *Problèmes de sécurité posés par les proxies d'adaptation multimedia : proposition de solutions pour une sécurisation de bout-en-bout*, A.R. Kaced, Télécom ParisTech, mai 2009 (examinatrice).
- *Analyse et détection dynamique de codes viraux dans un contexte cryptographique et application à l'évaluation de logiciels antivirus*, S. Josse, Ecole Polytechnique, avril 2009 (examinatrice).
- *Analyse de train binaire, Stéganographie*, J. Barbier, Ecole Polytechnique, novembre 2007 (examinatrice).
- *Étude et conception de systèmes de chiffrement à flot dans le contexte d'architectures matérielles fortement contraintes*, V. Bénony, Université de Lille 1, juin 2006 (co-encadrante).
- *Construction de codes Z_{p^k}-linéaires de bonne distance minimale, et schémas de dissimulation fondés sur les codes de recouvrement*, F. Galand, Université de Caen, décembre 2004 (examinatrice).
- *Message digests for still images and video contents*, F. Lefèbvre, Université catholique de Louvain, Belgique, mai 2004 (examinatrice).
- *Système d'hordatage dans un environnement de réseau ouvert*, J. Byun, Université de Toulon et du Var, juin 2002 (présidente du jury).
- *Études d'outils pour la dissimulation d'information : approches fractales, protocoles d'évaluation et protocoles cryptographiques*, F. Raynal, Université Paris XI, mars 2002 (examinatrice).

Encadrement de stages de DEA ou Master 2 Recherche

- *Protection de bases de données médicales : application d'algorithmes*, U. Verma, stage de Master 2 Recherche SISEA, Télécom Bretagne, février-août 2010, co-direction avec G. Coatrieux.
- *Fingerprinting : modélisation d'attaques aveugles par collusion*, A. Charpentier, stage de Master 2 Recherche CCC, Université de Limoges, février-août 2007.
- *Etude du chiffrement de bases de données pour utilisateurs mobiles*, S. Labis, stage de DEA d'Informatique, Université Lille 1, février-juin 2004, co-direction avec D. Simplot-Ryl.
- *Propriétés des suites binaires produites par un registre à décalage à rétroaction non linéaire*, V. Bénony, stage de DEA d'Informatique, Université Lille 1, février-juin 2002, co-direction avec V. Cordonnier.
- *Extension du domaine de confiance des données et des traitements*, L. Dabouz, stage de DEA d'Informatique, Université Lille 1, mars-juin 2000, co-direction avec V. Cordonnier.

Rapports pour des conférences, revues, ...

J'ai été rapporteur pour l'évaluation d'articles soumis aux revues : Discrete Applied Mathematics Journal (2001, 2006), Elsevier Signal Processing (2008), EURASIP Journal on Information Security (2007), IEEE Transactions on Circuits and Systems for Video Technology (2006, 2007), IEEE Transactions on Computers (2001, 2005), IEEE Transactions on Image Processing (2001, 2010), IEEE Transactions on Information Theory (2000, 2001, 2003, 2004, 2010), IEEE Transactions on Information Forensics and Security (2007, 2008, 2010), IEEE Transactions on Information Technology in BioMedicine (2010), IEEE Transactions on Multimedia (2003), IEEE Transactions on Parallel and Distributed Systems (2005), IEEE Transactions on Signal Processing (2004, 2005), IEEE Signal Processing Letters (2005), IET Information Security (2007), International Journal of Imaging Systems and Technology (2006), Journal of Communications and Networks (2002), Journal in Computer Virology (2006, 2007), Signal, Image and Video Processing (2007, 2010) SPIE Optical Engineering Letter (2003), Traitement du signal (2001) ; ainsi qu'aux conférences : AAECC (2005), AC (2006), Asiacrypt (2004), Cardis (2006), CORESA (2004, 2007, 2008, 2009, 2010), Eurocrypt (1998), EUSIPCO (2009), ICC Wireless Adhoc and Sensor Networks Symposium (2007), ICC Computer and Communications Network Security Symposium (2007), IEEE/ACM International Conference on Digital Information Management (2007), Indocrypt (2001, 2004, 2005, 2006), ISCAS (2006), ISSPA (2003, 2010), ISIT (2007) PCM (2006), PDS (2010), SAR (2003, 2004), SAR-SSI (2009), SSTIC (2004, 2006, 2007, 2008, 2009, 2010), Stacs (2000), WCC (2003, 2007, 2009), Words (2000).

J'ai été sollicitée pour évaluer des dossiers dans les appels à projets : ACI Sécurité Informatique (2003, 2004), ANR programme Sécurité et Informatique – SetIn (2006), ANR programme Jeunes chercheuses et jeunes chercheurs (2006), ANR programme RIAM (2006), ANR programme SeSur (2007).

Vie de la recherche

- Coordonnation de la préparation de la **fête de la science** 2007, lors de laquelle mon équipe représentait l'IRISA. Cinq démonstrations ont été préparées, illustrant les principes de la compression scalable, de la représentation de signaux vidéo en 3D, des codes correcteurs d'erreurs et du tatouage d'images. Trois stages ont été nécessaires à cette préparation.
- Responsable des **comités de programme et d'organisation** des journées *Codage et Cryptographie* 2005 organisées dans le cadre du GDR ALP, ainsi que du workshop international Wacha'07 organisé dans le cadre du réseau d'excellence européen ECRYPT. Membre des comités de programme des journées *Codes et Stéganographie* (2011), des conférences nationales SSTIC (2004, 2006-2011), C&ESAR (2009), SAR (2004), SAR-SSI (2009), et CORESA (2007, 2009, 2010, 2011), et internationales Indocrypt (2004, 2005), WCC (2007, 2009), Wacha (2007) et IEEE/ACM ICDIM'07 (2007). Membre des comités d'organisations des conférences nationales CORESA (2004), SSTIC (2007-2011), et internationales WCC (1999, 2003, 2007) et Wacha (2007).
- Membre du **comité éditorial** de la revue internationale Journal in Computer Virology, publiée par Springer depuis l'été 2005.
- Membre de 2006 à 2009 du **conseil scientifique du Groupement d'Intérêt scientifique Diwall** de Bretagne, regroupant l'IRISA (UMR 6074), Supelec-Rennes et Télécom Bretagne.
- Membre du **CHS** de l'IRISA UMR 6074 (de 2006 à 2009). Membre élu (de 2002 à 2005) du **conseil de laboratoire** du LIFL UMR 8022, et participation aux differentes commissions internes au laboratoire durant cette période (CHS, Sécurité, Bureaux, Abonnements).
- Membre des **GDRs IM** (depuis 1995) et **ISIS** (depuis 1999). Correspondant local à Rennes du groupe C2 (Codage et Cryptographie) du GDR IM (anciennement ALP) (de 2006 à 2009). Correspondant local à Lille du GDR ISIS (de 2000 à 2005) ainsi que du groupe C2) du GDR ALP (de 1999 à 2005).
- Membre du **jury du prix de thèse SPECIF** en 2005 et 2006.
- Membre de **comités de sélection** des universités Paris 5 et Paris 7 en 2011.
- Membre nommé du **CNU**, section 27, à partir de 2012.
- Lauréate, avec Teddy Furon et François Cayre, du **best paper award** de la conférence internationale IWDW 2004 [CFF05b].

Enseignement

depuis 2011 (Brest)	Cours et Travaux Dirigés de cryptographie à Télécom Bretagne, en 3ème année de formation ingénieur, et en 2ème année de formation professionnelle par alternance.
depuis 2007 (Rennes)	Master Recherche M2RI (Rennes) : co-responsable du parcours « Sécurité », responsable du module « Disponibilité et Protection de Contenus ».
2003-2008	Diverses interventions d'une demi-journée sur le thème du tatouage d'images et de la sécurité DESS CM (Univ. Lille 1), Master SSI (Supelec Rennes et ENST-Bretagne), DESS CCSI (Univ. Bordeaux 1), école doctorale I2S (Univ. Montpellier 2), INSA-Rennes, école chercheurs de l'IRISA
2002-2005 (Lille)	Cours, Travaux Dirigés et Pratiques de cryptographie en maîtrise d'informatique.
1999-2002 (Lille)	Cours de cryptographie dans les DESS TIIR et MICE. Cours, Travaux Dirigés et Pratiques de cryptographie en maîtrise d'informatique. Cours, Travaux Dirigés et Pratiques de programmation C/UNIX en deuxième année d'IUP GMI (niveau licence). Travaux Dirigés de mathématiques de l'information en deuxième année de DEUG MIAS. Cours et Travaux Dirigés sur le codage de l'information en première année de DEUG MIAS.
1998-1999 (Paris 11)	Travaux Dirigés en première année de DEUG MIAS : codage des nombres, algorithmique. Travaux Pratiques de programmation en C en deuxième année d'IUP MIAGE (niveau licence).
1995-1998 (Cergy Pontoise)	Travaux Dirigés en deuxième année de DEUG MIAS : théorie de l'information, bases de données, automates, grammaires non contextuelles, avec des applications programmées en Pascal. Travaux Dirigés d'initiation à l'informatique en première année de DEUG : fonctionnement d'un ordinateur, du DOS, de Windows, d'un traitement de texte (WordPerfect) et d'un tableur (QuattroPro).

Autres encadrements (projets, stages, ingénieurs, ...)

- *Réalisation d'une plate-forme de systèmes de chiffrement*, G. Bonnoron, P. Brunet, F. Cornevaux-Juignet et C. Duchêne, projet S2 de 1ère année, Télécom Bretagne, janvier-juin 2011.
- *Traçabilité des documents en santé*, Y. Lejosne, projet de recherche de 3ème année, Télécom Bretagne, janvier-mars 2011, co-encadré avec G. Coatrieux.
- *Développement de la plate-forme FANTOMAS*, M. Desoubeaux, ingénieur associé INRIA, co-encadré avec T. Furon, 2007-2009. Durant ces deux années, M. Desoubeaux a conçu et développé une plate-forme de démonstration et d'évaluation de techniques de tatouage et fingerprinting.
- *Navigation au sein de scènes 3D*, C. Samson, stage de 4ème année de l'INSA de Rennes, juillet-août 2007.
- *Implémentation du nouveau standard de chiffrement AES dans une carte à puce*, C. Lambert, stage de maîtrise d'informatique, Université de Lille 1, juillet-août 2001, co-direction avec E. Wegrzynowski et S. Jean.
- *Implémentation d'une nouvelle construction de fonctions booléennes possédant de bonnes propriétés cryptographiques*, C. Vermeulen et V. Bénony, stage de maîtrise d'informatique, Université de Lille 1, mars-juin 2001, co-direction avec E. Wegrzynowski.
- *Construction de fonctions booléennes ayant de bonnes propriétés cryptographiques*, C. Cordenier et E. Declercq, stage de 3ème année IUP GMI (niveau maîtrise), Université de Lille 1, septembre 2000.
- *Utilisation des codes de Reed-Muller dans un système de protection des droits d'auteur pour les images numériques*, O. Chermeux et G. Plancke, stage de 3ème année IUP GMI (niveau maîtrise), Université de Lille 1, septembre 2000.
- *Codes de Reed-Muller et leur implication dans les générateurs pseudo-aléatoires*, A. Vermoen, Stage de fin d'étude, Université de Eindhoven, effectué à l'INRIA de novembre 1997 à mai 1998, co-direction avec P. Charpin.

Tutorat de stages en entreprises

- *Développement d'une application client/serveur pour le GAN Capitalisation*, F. Vanderbeke, stage de 3ème année d'IUP MIAGE effectué chez NORSYS, juin-septembre 2000.
- *Montée en charge*, G. Lantoine, stage de 3ème année d'IUP MIAGE effectué chez NORSYS, juin-septembre 2000.
- tutorat de divers stages effectués par des étudiants de DEUG MIAS 2ème année, à l'Université de Cergy Pontoise, 1998

C.2 Liste de publications

Chapitres de livres

[Fon11] C. Fontaine. – 11 entrées sur le chiffrement à flot : *synchronous stream cipher, self-synchronizing stream cipher, non-linear feedback shift register, clock-controlled generator, shrinking generator, self-shrinking generator, linear congruential generator, summation generator, E0 (Bluetooth cipher), RC4, SEAL. Encyclopedia of Cryptography and Security*, mise à jour pour la seconde édition, Springer, à paraître.

[CCF07] F. Cayre, C. Fontaine et T. Furon. – Watermarking Security. *Digital Audio Watermarking Techniques and Technologies : Applications and Benchmarking*, Idea Group Publishing, 2007, ISBN 978-159904513-9.

[Fon05] C. Fontaine. – 11 entrées sur le chiffrement à flot : *synchronous stream cipher, self-synchronizing stream cipher, non-linear feedback shift register, clock-controlled generator, shrinking generator, self-shrinking generator, linear congruential generator, summation generator, E0 (Bluetooth cipher), RC4, SEAL. Encyclopedia of Cryptography and Security*, Springer, 2005, ISBN 0-387-23473-X.

[PF04] F. Petitcolas et C. Fontaine. – Nouveaux outils pour l'évaluation des algorithmes de tatouage. *Tatouage de documents audiovisuels numériques*, chapitre 6, pp. 195–214, Hermès-Lavoisier, 2004, ISBN 2-7462-0816-4.

Articles invités dans des revues internationales (avec comité de lecture)

[ABD+99] D. Augot, J.M. Boucqueau, J.-F. Delaigle, C. Fontaine et E. Goray. – Secure Delivery of Images over Open Networks. *Proc. of the IEEE*, 87(7) :1251-1266. – numéro spécial "Identification and protection of multimedia information", article invité, 1999.

Articles dans des revues internationales (avec comité de lecture)

[FG09] C. Fontaine et F. Galand. – How Reed-Solomon codes can improve steganographic schemes. *EURASIP Journal on Inf. Security.* – Vol. 2009, article ID 274845, numéro spécial "Secure Steganography in Multimedia Content", 2009.

[FDD+08] C. Fontaine, C. Delpha, P. Duhamel, A. Benjelloun-Touimi, M. Milhau, A. Le Guyader, C. Giraud, P. Martin, D. Azemard et J.-B. Fischer. – An end-to-end security architecture for music distribution on mobile phones. *ISAST Trans. on Com. and Networking.* – 2(1) :81–91 – 2008.

[FG07a] C. Fontaine et F. Galand. – A survey of homomorphic encryption for non-specialists. *EURASIP Journal on Inf. Security.* – Vol. 2007, article ID 13801, numéro spécial "Signal Processing in the encrypted domain", 2007.

[CCF05e] F. Cayre, C. Fontaine et T. Furon. – Watermarking security : theory and practice. *IEEE Trans. on Signal Processing.* – 53(10) :3976–3987 numéro spécial "Supplement on Secure Media III", 2005.

[CCCF01] A. Canteaut, C. Carlet, P. Charpin et C. Fontaine. – On cryptographic properties of the cosets of $R(1,m)$. *IEEE Trans. on Inf. Theory.* – 47(4) :1494–1513 – article de fond (regular paper), 2001.

[Fon99] C. Fontaine. – On some cosets of the first-order Reed-Muller code with high minimum weight. *IEEE Trans. on Inf. Theory.* – 45(4) :1237–1243, 1999.

Articles dans des revues nationales (avec comité de lecture)

[CFF10] A. Charpentier, C. Fontaine et T. Furon. – Décodage EM du code de Tardos pour le fingerprinting. *Traitement du Signal.* Vol. 27, num. 2, pp. 127–146, 2010.

[RPF01] F. Raynal, F.A. Petitcolas et C. Fontaine. – Évaluation automatique des méthodes de tatouage. *Traitement du Signal.* Vol. 18, num. 4. – numéro spécial “Tatouage et sécurité de l’information”, 2001.

Articles invités dans des actes de congrès internationaux (avec comité de lecture)

[ABTD+06] D. Azemard, A. Benjelloun-Touimi, C. Delpha, P. Duhamel, J.-B. Fischer, C. Fontaine, C. Giraud, A. Le Guyader, P. Martin et M. Milhau. – Secured diffusion of music on mobile : an end-to-end approach. *Taiwanese-French conference on Information Technology, TFIT’06*, article invité, 2006.

Articles longs dans des actes de congrès internationaux (avec comité de lecture)

[ABF11] D. Augot, M. Barbier et C. Fontaine. Ensuring message embedding in wet paper steganography. *IMA 13th Conference on Cryptography and Coding*, Lecture Notes in Computer Science. – Springer-Verlag, 2011.

[CFFC11] A. Charpentier, C. Fontaine, T. Furon et I. Cox. – An Asymmetric Fingerprinting Scheme based on Tardos Codes. *Information Hiding, IH 2011*, Lecture Notes in Computer Science 6958. pp. 43-58 – Springer-Verlag, 2011.

[XFF10b] F. Xie, T. Furon et C. Fontaine. – Towards Robust and Secure Watermarking. *In : ACM Workshop on Multimedia and Security, MM&SEC 2010.* – ACM, 2010.

[XFF08] F. Xie, T. Furon et C. Fontaine. – On-Off Keying modulation and Tardos fingerprinting. *In : ACM Workshop on Multimedia and Security, MM&SEC 2008.* – ACM, 2008.

[DGF+08] C. Dikici, C. Guillemot, C. Fontaine, K. Idrissi et A. Baskurt. – Dirty Paper Coding with Partial State Information. *IEEE International Symposium*

on Image/Video Communications over fixed and mobile networks, ISIVC 2008. – IEEE, 2008.

[FG07b] C. Fontaine et F. Galand. – How can Reed-Solomon codes improve steganographic schemes. *Information Hiding, IH 2007*, Lecture Notes in Computer Science 4567. pp. 130-144 – Springer-Verlag, 2007.

[BTFF+06] A. Benjelloun-Touimi, J.-B.Fischer, C. Fontaine, C. Giraud et M. Milhau. – Enhanced Security Architecture for Music Distribution on Mobile. *European Symposium on Research in Computer Security, ESORICS 2006*, Lecture Notes in Computer Science 4189. pp. 97-109. – Springer-Verlag, 2006.

[CCF05a] F. Cayre, C. Fontaine et T. Furon. – A theoretical study of watermarking security. *IEEE International Symposium on Information Theory, ISIT 2005*, IEEE, 2005.

[CCF05b] F. Cayre, C. Fontaine et T. Furon. – Watermarking Attack : Security of WSS Techniques. *International Workshop on Digital Watermarking, IWDW 2004*, **Best Paper Award**, Lecture Notes in Computer Science 3304. pp. 171–183.– Springer-Verlag, 2005.

[BRWF05] V. Bénony, F. Recher, E. Wegrzynowski et C. Fontaine. – Cryptanalysis of a particular case of Klimov-Shamir pseudo-random generator. *SEquences and Their Applications, SETA 2004*, Revised Selected Papers, Lecture Notes in Computer Science 3486. pp. 313-322. – Springer-Verlag, 2005.

[FFJ04] E. Filiol, C. Fontaine et S. Josse. – The COSvd ciphers. *SASC : the State of the Art of Stream Ciphers* – NoE ECRYPT Workshop, 2004.

[FF01] E. Filiol et C. Fontaine. – A new fast stream cipher design : COS ciphers. *IMA 8th Conference on Cryptography and Coding*, Lecture Notes in Computer Science 2260. pp. 85–98. – Springer-Verlag, 2001.

[CCCF00b] A. Canteaut, C. Carlet, P. Charpin et C. Fontaine. – Propagation characteristics and correlation-immunity of highly nonlinear Boolean functions. *In : Advances in Cryptology - EUROCRYPT'2000*, Lecture Notes in Computer Science 1807. pp. 507–522. – Springer-Verlag, 2000.

[AFD98] D. Augot, C. Fontaine et J.-F. Delaigle. – DHWM : a scheme for managing watermarking keys in the aquarelle multimedia distributed system. *In : European Symposium on Research in Computer Security - ESORICS 98*, Lecture Notes in Computer Science 1485. pp. 241–255. – Springer-Verlag, 1998.

[FF98] E. Filiol et C. Fontaine. – Highly nonlinear balanced Boolean functions with a good correlation-immunity. *In : Advances in Cryptology - EUROCRYPT'98*, Lecture Notes in Computer Science 1403. pp. 475–488. – Springer-Verlag, 1998.

[Fon96] C. Fontaine. – The nonlinearity of a class of Boolean functions with short representation. *In : PRAGOCRYPT'96*, éd. par Přibyl (J.). pp. 129–144. – CTU Publishing House, 1996.

Articles longs dans des actes de congrès internationaux (avec comité de lecture, acceptation sur résumé)

[XFF10a] F. Xie, T. Furon et C. Fontaine. – Better security levels for 'Broken Arrows'. *In : Proceeedings of the SPIE*. Vol. 7541. IS&T/SPIE International Symposium on Electronic Imaging 2010 : Media Forensics and Security XII. – SPIE, 2010.

[CXFF09] A. Charpentier, F. Xie, C. Fontaine et T. Furon. – Expectation Maximisation decoding of Tardos probabilistic fingerprinting code. *In : Proceeedings of the SPIE*. Vol. 7254. IS&T/SPIE International Symposium on Electronic Imaging 2009 : Media Forensics and Security XI. – SPIE, 2009.

[CCF05c] F. Cayre, C. Fontaine et T. Furon. – Watermarking Security, Part one : theory. *In : Proceeedings of the SPIE*. Vol. 5681. IS&T/SPIE International Symposium on Electronic Imaging 2005 : Security, Steganography, and Watermarking of Multimedia Contents VII. pp. 746–757. – SPIE, 2005.

[CCF05d] F. Cayre, C. Fontaine et T. Furon. – Watermarking Security, Part two : practice. *In : Proceeedings of the SPIE*. Vol. 5681. IS&T/SPIE International Symposium on Electronic Imaging 2005 : Security, Steganography, and Watermarking of Multimedia Contents VII. pp. 758–768. – SPIE, 2005.

[FR02] C. Fontaine et F. Raynal. – About the links between crtyptography and information hiding. *In : Proc. of the SPIE*. Vol. 4675. IS&T/SPIE International Symposium on Electronic Imaging 2002 : Security and watermarking of Multimedia Contents IV. pp. 269–280. – SPIE, 2002.

[SPR+01] M. Steinebach, F.A. Petitcolas, F. Raynal, J. Dittmann, C. Fontaine, C. Seibel et N. Fates. – Stirmark benchmark : audio watermarking attacks. *In : International Conference on Information Technology : Coding and Computing, ITCC 2001*. Special Session in Multimedia Security and Watermarking Applications. IEEE Computer Press, 2001.

[PSR+01] F.A. Petitcolas, M. Steinebach, F. Raynal, J. Dittmann, C. Fontaine et Fates (N.). – Public automated web-based evaluation service for watermarking schemes : Stirmark benchmark. *In : Proc. of the SPIE*. Vol. 4314. IS&T/SPIE International Symposium on Electronic Imaging 2001 : Security and Watermarking of Multimedia Contents III. pp. 575–584. – SPIE, 2001.

[AF98] D. Augot et C. Fontaine. – Key issues for watermarking digital images. *In : Proc. of the SPIE*, Vol. 3409. – EUROPTO Conference on Electronic Imaging : Processing, Printing, and Publishing in Color. Int. Symposium on Electronic Image Capture and Publishing 98. pp. 176–185. – SPIE, 1998.

Articles courts dans des actes de congrès internationaux (avec comité de lecture)

[FFV01] E. Filiol, C. Fontaine et D. Vianne. – A new fast block cipher design : COS ciphers. *IEEE International Symposium on Information Theory, ISIT 2001*, p. 138, IEEE, 2001.

[CCCF00a] A. Canteaut, C. Carlet, P. Charpin et C. Fontaine. – Fourier spectrum of optimal Boolean functions via Kasami's identities. *IEEE International Symposium on Information Theory, ISIT 2000*, p. 183, IEEE, 2000.

[Fon98b] C. Fontaine. – A method to find cosets of the first-order Reed-Muller code with a high minimum weight. *IEEE International Symposium on Information Theory, ISIT 98*, p. 464, IEEE, 1998.

Articles dans des actes de congrès nationaux (avec comité de lecture)

[CFF09] A. Charpentier, C. Fontaine et T. Furon. – Décodage EM du code de Tardos pour le fingerprinting. *GRETSI 2009*, 2009.

[XFF09] F. Xie, C. Fontaine et T. Furon. – Un schéma complet de traçage de documents multimédia reposant sur des versions améliorées des codes de Tardos et de la technique de tatouage Broken Arrows. *GRETSI 2009*, 2009.

Thèse

[Fon98] C. Fontaine. – Contribution à la recherche de fonctions booléennes hautement non linéaires, et au marquage d'images en vue de la protection des droits d'auteurs. – Thèse de doctorat, Université Paris VI, novembre 1998.

Brevets

[GFF+06] C. Giraud, J.-B. Fischer, C. Fontaine, A. Benjelloun-Touimi, M. Milhau et B. Prady. – Dispositif de restitution d'un contenu numérique, entité électronique sécurisée comprenant ces éléments et procédé de restitution d'un contenu numérique. – Demande de brevet 0651089 déposée en France par Oberthur Card System SA et France Telecom le 29 mars 2006.

Articles de vulgarisation

- C. Fontaine. – parcours et fiches sur la dissimulation d'information. – *Portail Internet Cryptologie et Sécurité de l'information - PICSI*, `http://www.picsi.org/accueil.html`, 2005.
- F. Raynal, F. Petitcolas et C. Fontaine. – L'art de dissimuler les informations. – *Pour la Science*, dossier "l'art du secret" , été 2002.
- C. Fontaine. – Le tatouage des images numériques. – *Pour la Science*, dossier "l'art du secret" , été 2002.
- F. Raynal, F. Petitcolas et C. Fontaine. – Introduction à la stéganographie. – *M.I.S.C. le magazine de la sécurité informatique*, num. 1, janvier 2002.
- C. Fontaine. – Le tatouage des images numériques. – *Pour la Science*, num. 270, avril 2000.

Autres

J'ai donné 44 exposés dans des séminaires ou lors de journées thématiques, et participé à la rédaction d'une quarantaine de rapports dans le cadre des contrats ou projets auxquels j'ai participé.

Annexe D

Publications choisies

J'ai choisi de compléter ce mémoire avec le contenu détaillé de certaines publications, peu détaillées dans le manuscrit :

[FG09] How Reed-Solomon Codes Can Improve Steganographic Schemes, *EURASIP Journal on Information Security*, 2009 → page 138

[ABF11] Ensuring message embedding in wet paper steganography, *IMA 13th Conference on Cryptography and Coding*, 2011 → page 148

[CFF05e] Watermarking Security : Theory and Practice, *IEEE Transactions on Signal Processing*, 2005 → page 166

[CFFC11] An Asymmetric Fingerprinting Scheme based on Tardos Codes, *Information Hiding*, 2011 → page 178

[CCCF01] On Cryptographic Properties of the Cosets of $R(1, m)$, *IEEE Transactions on Information Theory*, 2001 → page 194

Hindawi Publishing Corporation
EURASIP Journal on Information Security
Volume 2009, Article ID 274845, 10 pages
doi:10.1155/2009/274845

Research Article

How Reed-Solomon Codes Can Improve Steganographic Schemes

Caroline Fontaine and Fabien Galand

CNRS/IRISA-TEMICS Group, Campus de Beaulieu, 35 042 Rennes Cedex, France

Correspondence should be addressed to Caroline Fontaine, caroline.fontaine@irisa.fr

Received 31 July 2008; Accepted 6 November 2008

Recommended by Miroslav Goljan

The use of syndrome coding in steganographic schemes tends to reduce distortion during embedding. The more complete model comes from the wet papers (J. Fridrich et al., 2005) and allow to lock positions which cannot be modified. Recently, binary BCH codes have been investigated and seem to be good candidates in this context (D. Schönfeld and A. Winkler, 2006). Here, we show that Reed-Solomon codes are twice better with respect to the number of locked positions; in fact, they are optimal. First, a simple and efficient scheme based on Lagrange interpolation is provided to achieve the optimal number of locked positions. We also consider a new and more general problem, mixing wet papers (locked positions) and simple syndrome coding (low number of changes) in order to face not only passive but also active wardens. Using list decoding techniques, we propose an efficient algorithm that enables an adaptive tradeoff between the number of locked positions and the number of changes.

1. Introduction

Steganography aims at sending a message through a cover-medium, in an *undetectable* way. *Undetectable* means that nobody, except the intended receiver of the message, should be able to tell if the medium is carrying a message or not [1]. Hence, if we speak about still images as cover-media, the embedding should work with the smallest possible distortion, not being detectable with the quite powerful analysis tools available [2, 3]. A lot of papers have been published on this topic, and it appears that modeling the embedding and detection/extraction processes with an error correcting code point of view, usually called matrix embedding by the steganographic community, may be helpful to achieve these goals [4–15]. The main interest of this approach is that it decreases the number of components modifications during the embedding process. As a side effect, it was remarked in [8] that matrix embedding could be used to provide an effective answer to the adaptive selection channel problem. The sender can embed the messages adaptively with the cover-medium to minimize the distortion, and the receiver can extract the messages without being aware of the sender choices. A typical steganographic application is the perturbed quantization [16]; during quantization process, for example, JPEG compression, real values v have to be rounded between possible quantized values $x_0, \ldots, x_j$; when v lies close to the middle of an interval $[x_i, x_{i+1}]$, one can choose between x_i and x_{i+1} without adding too much distortion. This allows to embed messages under the condition that the receiver does not need to know which positions were modified.

It has been shown that if random codes may seem interesting for their asymptotic behavior, their use leads to solve really hard problems; syndrome decoding and covering radius computation, which are proved to be NP-complete and Π_2-complete, respectively (the Π_2 complexity class includes the NP class) [17, 18]. Moreover, no efficient decoding algorithm is known, even for a small nontrivial family of codes. From a practical point of view, this implies that the related steganographic schemes are too complex to be considered as acceptable for real-life applications. Hence, it is of great interest to have a deeper look at other kinds of codes, structured codes, which are more accessible and lead to efficient decoding algorithms. In this way, some previous papers studied the Hamming code [4, 6, 9], the Simplex code [11], and binary BCH codes [12]. Here, we focus

on this latter paper, that pointed out the interest in using codes with deep algebraic structures. The authors distinguish two cases, as previously introduced in [8]. The first one is classical: the embedder modifies any position of the cover-data (a vector which is extracted from the cover-medium, and processed by the encoding scheme), the only constraint being the maximum number of modifications allowed. In this case, they showed that binary BCH codes behave well, but pointed out that choosing the most appropriate code among the BCH family is quite hard, we do not know good complete syndrome decoding algorithms for BCH codes. In the second case, some positions are locked and cannot be used for embedding; this is due to the fact that modifying these positions leads to a degradation of the cover-medium that is noticeable. Hence, in order to remain undetectable, the sender restricts himself to keep these positions and lock them. This case is more realistic. The authors showed that there is a tradeoff between the number of elements that can be locked and the efficiency of the code.

This paper is organized as follows. In Section 2, we review the basic setting of coding theory used in steganography. In Section 3, we recall the syndrome coding paradigm, including wet paper codes and active warden. Section 4 presents the classical Reed-Solomon codes and gives details on the necessary tools to use them with syndrome coding, notably the Guruswami-Sudan list decoding algorithm. Section 5 leads to the core of this paper; in Section 5.1, we describe a simple algorithm to use Reed-Solomon codes in an optimal way for wet paper coding, and in Section 5.2 we describe and analyze our proposed algorithm constructed upon the Guruswami-Sudan decoding algorithm.

Before going deeper in the subject, please note that we made the choice to represent vectors as horizontal vectors. For general references to error correcting codes, we orientate the reader toward [19].

2. A Word on Coding Theory

We review here a few concepts relevant to coding theory applications in steganography.

Let $\mathbb{F}_q = \mathrm{GF}(q)$ be the finite field with q elements, q being a power of some prime number. We consider n-tuples over $\mathbb{F}_q$, usually referring to them as *words*. The classical *Hamming weight* $\mathrm{wt}(v)$ of a word v is the number of coordinates that is different from zero, and the *Hamming distance* $d(u, v)$ between two words u, v denotes the weight of their difference, that is, the number of coordinates in which they differ. We denote by $B_a(v)$ the *ball of radius a centered on* v, that is, $B_a(v) = \{u \mid d(u, v) \leq a\}$. Recall that the volume of a ball, that is, the number of its elements does not depend on the center v, and is equal to $V_a = |B_a(v)| = \sum_{i=0}^{a} (q-1)^i \binom{n}{i}$ in dimension n.

A *linear code* $\mathcal{C}$ is a vector subspace of $\mathbb{F}_q^n$ for some integer n, called the *length* of the code. The *dimension* k of $\mathcal{C}$ corresponds to its dimension as a vector space. Hence, a linear code of dimension k contains q^k *codewords*. The two main parameters of codes are their *minimal distance* and *covering radius*. The *minimal distance* of $\mathcal{C}$ is the minimal Hamming distance between two distinct codewords and, since we restrict ourself to linear codes, it is the minimum weight of a nonzero codeword. The minimum distance is closely related to the *error correction capacity* of the code; a code of minimal distance d corrects any error vector of weight at most $t = \lfloor (d-1)/2 \rfloor$; that is, it is possible to recover the original codeword c from any $y = c + e$, with $\mathrm{wt}(e) \leq t$. On the other hand, the *covering radius* ρ is the maximum distance between any word of $\mathbb{F}_q^n$ and the set of all codewords, $\rho = \max d(z, \mathcal{C})$. A linear code of length n, dimension k, minimum distance d, and covering radius ρ is said to be $[n, k, d]\rho$.

An important point about linear codes is their matrix description. Since a linear code is a vector space, it can be described by a set of linear equations, usually in the shape of a single matrix, called the *parity check matrix* . That is, for any $[n, k, d]\rho$ linear code $\mathcal{C}$, there exists an $(n-k) \times n$ matrix H such that

$$c \in \mathcal{C} \iff c \cdot H^t = 0. \tag{1}$$

An important consequence is the notion of *syndrome* of a word, that uniquely identifies the *cosets* of the code. A *coset* of $\mathcal{C}$ is a set $e + \mathcal{C} = \{e + c \mid c \in \mathcal{C}\}$. Two remarks have to be pointed out; first, the cosets of $\mathcal{C}$ form a partition of the ambient space $\mathbb{F}_q^n$; second, for any $y \in e + \mathcal{C}$, we have $y \cdot H^t = e \cdot H^t$, and each coset can be identified by the value of the *syndrome* $z \cdot H^t$ of its elements z denoted here as $E(z)$.

The two main parameters d and ρ have interesting descriptions with respect to syndromes. For any word $e \in \mathbb{F}_q^n$ of weight at most $t = \lfloor (d-1)/2 \rfloor$, the coset $e + \mathcal{C}$ has a unique word of weight at most $\mathrm{wt}(e)$. Stated differently, if the equation $e \cdot H^t = m$ has a solution of weight $\mathrm{wt}(e) \leq t$, then it is unique. Moreover, t is maximal for this property to hold. On the other hand, for m element of $\mathbb{F}_q^n$, the equation $e \cdot H^t = m$ always has a solution e of weight at most ρ. Again, ρ is extremal with respect to this property; it is the smallest possible value for this to be true.

A *decoding mapping*, denoted by D, associates with a syndrome m a vector e of Hamming weight less than or equal to ρ, which syndrome is precisely equal to m, $\mathrm{wt}(D(m)) \leq \rho$ and $E(D(m)) = D(m) \cdot H^t = m$. For our purpose, it is not necessary that D returns the vector e of minimum weight. Please, remark that the effective computation of D corresponds to the complete syndrome decoding problem, which is hard.

Finally, we need to construct a smaller code $\mathcal{C}_{\mathcal{I}}$ from a bigger one $\mathcal{C}$. The operation we need is called *shortening*; for a fixed set of coordinates $\mathcal{I}$, it consists in keeping all codewords of $\mathcal{C}$ that have zeros for all positions in $\mathcal{I}$ and then deleting these positions. Remark that if $\mathcal{C}$ has parameters $[n, k, d]$ with $d > |\mathcal{I}|$, then the resulting code, $\mathcal{C}_{\mathcal{I}}$, has length $n - |\mathcal{I}|$ and dimension $k - |\mathcal{I}|$.

3. Syndrome Coding

The behavior of a steganographic algorithm can be sketched in the following way:

(1) a *cover-medium* is processed to extract a sequence of symbols v, sometimes called *cover-data*;

(2) v is modified into s to embed the message m; s is sometimes called the *stego-data*;

(3) modifications on s are translated on the cover-medium to obtain the *stego-medium*.

Here, we assume that the detectability of the embedding increases with the number of symbols that must be changed to go from v to s (see [6, 20] for some examples of this framework).

Syndrome coding deals with this number of changes. The key idea is to use some syndrome computation to embed the message into the cover-data. In fact, such a scheme uses a linear code $\mathcal{C}$, more precisely its cosets, to hide m. A word s hides the message m if s lies in a particular coset of $\mathcal{C}$, related to m. Since cosets are uniquely identified by the so-called syndromes, embedding/hiding consists exactly in searching s with syndrome m, close enough to v.

3.1. Simple Syndrome Coding. We first set up the notation and describe properly the syndrome coding framework and its inherent problems. Let $v \in \mathbb{F}_q^n$ denote the cover-data and $m \in \mathbb{F}_q^r$ the message. We are looking for two mappings, embedding Emb and extraction Ext, such that

$$\forall (v,m) \in \mathbb{F}_q^n \times \mathbb{F}_q^r, \quad \mathrm{Ext}(\mathrm{Emb}(v,m)) = m, \tag{2}$$

$$\forall (v,m) \in \mathbb{F}_q^n \times \mathbb{F}_q^r, \quad d_H(v, \mathrm{Emb}(v,m)) \leq T. \tag{3}$$

Equation (2) means that we want to recover the message in all cases; (3) means that we authorize the modification of at most T coordinates in the vector v.

From Section 2, it is quite easy to show that the scheme defined by

$$\begin{aligned} \mathrm{Emb}(v,m) &= v + D(m - E(v)), \\ \mathrm{Ext}(y) &= E(y) = y \cdot H^t \end{aligned} \tag{4}$$

enables to embed messages of length $r = n-k$ in a cover-data of length n, while modifying at most $T = \rho$ elements of the cover-data.

The parameter $(n-k)/\rho$ represents the (worst) *embedding efficiency*, that is, the number of embedded symbols per embedding changes in the worst case. In a similar way, one defines the *average embedding efficiency* $(n-k)/\omega$, where ω is the average weight of the output of D for uniformly distributed inputs. Here, both efficiencies are defined with respect to symbols and not bits. Linking symbols with bits is not simple, as naive solutions lead to bad results in terms of efficiency. For example, if elements of $\mathbb{F}_q$ are viewed as blocks of ℓ bits, modifying a symbol roughly leads to $\ell/2$ bit flips on average and ℓ for the worst case.

3.2. Syndrome Coding with Locked Elements. A problem raised by the syndrome coding, as presented above, is that any position in the cover-data v can be changed. In some cases, it is more reasonable to keep some coordinates unchanged because they would produce too big artifacts in the stego-data. This can be achieved in the following way. Let $\mathcal{I}$ be the coordinates that must not be changed, let $H_{\mathcal{I}}$ be the matrix obtained from H by removing the corresponding columns; this matrix defines the shortened code $\mathcal{C}_{\mathcal{I}}$. Let $E_{\mathcal{I}}$ and $D_{\mathcal{I}}$ be the corresponding encoding and decoding mappings, that is, $E_{\mathcal{I}}(y) = y \cdot H_{\mathcal{I}}^t$ for $y \in \mathbb{F}_q^{n-|\mathcal{I}|}$, and $D_{\mathcal{I}}(m) \in \mathbb{F}_q^{n-|\mathcal{I}|}$ is a vector of weight at most $\rho_{\mathcal{I}}$ such that its syndrome, with respect to $H_{\mathcal{I}}$, is m. Here, $\rho_{\mathcal{I}}$ is the covering radius of $\mathcal{C}_{\mathcal{I}}$. Finally, let us define $D_{\mathcal{I}}^*$ as the vector of $\mathbb{F}_q^n$ such that the coordinates in $\mathcal{I}$ are zeros and the vector obtained by removing these coordinates is precisely $D_{\mathcal{I}}$. Now, we have $D_{\mathcal{I}}^*(m) \cdot H = D_{\mathcal{I}}(m) \cdot H_{\mathcal{I}}^t = m$ and, by definition, $D_{\mathcal{I}}^*(m)$ has zeros in coordinates lying in $\mathcal{I}$. Naturally, the scheme defined by

$$\begin{aligned} \mathrm{Emb}(v,m) &= v + D_{\mathcal{I}}^*(m - E(v)), \\ \mathrm{Ext}(y) &= E(y) = y \cdot H^t \end{aligned} \tag{5}$$

performs syndrome coding without disturbing the positions in $\mathcal{I}$. But, it is worth noting that for some sets $\mathcal{I}$, the mapping $D_{\mathcal{I}}$ cannot be defined for all possible values of m because the equation $y \cdot H_{\mathcal{I}}^t = m$ has no solution. This always happens when $|\mathcal{I}| > k$, since $H_{\mathcal{I}}$ has dimension $(n-k) \times (n - |\mathcal{I}|)$, but can also happen for smaller sets.

3.3. Syndrome Coding for an Active Warden. The previous setting focuses on distortion minimization to avoid detection by the entity inspecting the communication channel, the warden. This supposes the warden keeps a passive role, only looking at the channel. But, the warden can, in a preventive way, modify the data exchanged over the channel. To deal with this possibility, we consider that the stego-data may be modified by the warden, who can change up to w of its coordinates. (In fact, we suppose that the action of the warden on the stego-medium translates onto the stego-data in such a way that at most w coordinates are changed.)

This case has been addressed independently with different strategies by [21, 22]. To address it with syndrome coding, we want $\mathrm{Ext}(\mathrm{Emb}(v,m) + e) = m$ with $\mathrm{wt}(e) \leq w$. This requires that the balls $B_e(\mathrm{Emb}(v,m))$ are disjoint for different messages m. In fact, the requirements on Emb lead to a known generalization of error correcting codes, called *centered error correcting codes* (CEC codes). They are defined by an encoding mapping $f : \mathbb{F}_q^n \times \mathbb{F}_q^k \to \mathbb{F}_q^n$ such that $f(v,m) \in B_\rho(v)$ and the balls $B_w(f(v,m))$ do not intersect; f is precisely what we need for Emb in the active warden setting. A decoding mapping for this centered code plays the role of Ext.

Our problem can be reformulated as follows. Let us consider an error correcting code $\mathcal{C}$ of dimension k and length n used for syndrome coding, this code having a $(n-k) \times n$ parity check matrix H; now, let us consider a subcode $\mathcal{C}'$ of $\mathcal{C}$, of dimension k', defined by its $(n-k') \times n$ parity check matrix H', which can be written as

$$H' = \begin{pmatrix} H \\ H_1 \end{pmatrix}. \tag{6}$$

The $k - k'$ additional parity check equations given by H_1 correspond to the restriction from $\mathcal{C}$ to $\mathcal{C}'$. The cosets of $\mathcal{C}'$

in $\mathcal{C}$, that is, the sets $\{c+\mathcal{C}',\ c\in\mathbb{F}_q^n\}\subset\mathcal{C}$, can be indexed in this way

$$C_i=\{c\in\mathbb{F}_q^n,\ c\cdot H^t=0\ ,\ c\cdot H_1^t=i\},\quad 0\le i<k-k'. \tag{7}$$

The equation, $c\cdot H^t=0$, means that the word c belongs to $\mathcal{C}$, and $c\cdot H_1^t$ gives the coset of $\mathcal{C}'$ in which c lies. These cosets are pairwise disjoint and their union is $\mathcal{C}$. The index i may be identified with its binary expansion, and we can identify the embedding step with looking for a word $\mathrm{Emb}(v,m)$ such that

$$\begin{aligned}\mathrm{Emb}(v,m)\cdot\begin{pmatrix}H\\H_1\end{pmatrix}^t&=\begin{pmatrix}\mathrm{Emb}(v,m)\cdot H^t & \mathrm{Emb}(v,m)\cdot H_1^t\end{pmatrix}\\&=(0\,m).\end{aligned} \tag{8}$$

Hence, we can choose $\mathrm{Emb}(v,m)=v+y$, where y is a solution of $y\cdot(H^t\,H_1^t)=(0\,m)$, with $\mathrm{wt}(y)\le T$.

3.4. A Synthetic View of Syndrome Coding for Steganography. The classical problem of syndrome coding presented in Section 3.1 can be extended in several directions, as presented in Sections 3.2 and 3.3. It is possible to merge both in one to get at the same time reduced distortion and active warden resistance. This has some impact on the parity check matrices we have to consider.

Starting from the setting of the active warden, the problem becomes to find solutions of $y\cdot H'^t=(0\,m)$, with the additional restriction that $y_i=0$ for $i\in\mathcal{A}$. This means that we have to solve a particular instance of syndrome coding with locked elements, the syndrome has a special shape $(0\,m)$.

4. What Reed-Solomon Codes Are, and Why They May Be Interesting

Reed-Solomon codes over the finite field $\mathbb{F}_q$ are optimal linear codes. The *narrow-sense RS codes* have length $n=q-1$ and can be defined as a particular subfamily of the BCH codes. But, we prefer the alternative, and larger, definition as an evaluation code, which leads to the *generalized Reed-Solomon codes* (GRS *codes*).

4.1. Reed-Solomon Codes as Evaluation Codes. Roughly speaking, a GRS code of length $n\le q$ and dimension k is a set of words corresponding to polynomials of degree less than k evaluated over a subset of $\mathbb{F}_q$ of size n. More precisely, let $\{\gamma_0,\dots,\gamma_{n-1}\}$ be a subset of $\mathbf{F}_q$ and define $\mathrm{ev}(P)=(P(\gamma_0),P(\gamma_1),\dots,P(\gamma_{n-1}))$, where P is a polynomial over $\mathbb{F}_q$. Then, we define $\mathrm{GRS}(n,k)$ as

$$\mathrm{GRS}(n,k)=\{\mathrm{ev}(P)\mid\deg(P)<k\}. \tag{9}$$

This definition, *a priori*, depends on the choice of the γ_i and the order of evaluation; but, as the code properties do not depend on this choice, we will only focus here on the number n of γ_i and will consider an arbitrary set $\{\gamma_i\}$ and order. Remark that when $\gamma_i=\beta^i$ with β a primitive element of $\mathbb{F}_q$ and $i\in\{0,\dots,q-2\}$, we obtain the *narrow-sense Reed-Solomon codes* .

As we said, GRS codes are optimal since they are maximum distance separable (MDS); the minimal distance of $\mathrm{GRS}(n,k)$ is $d=n-k+1$, which is the largest possible. On the other hand, the covering radius of $\mathrm{GRS}(n,k)$ is known and equal to $\rho=n-k$.

Concerning the evaluation function, recall that if we consider $n\le q$ elements of $\mathbb{F}_q$, then it is known that there is a unique polynomial of degree at most $n-1$ taking particular values on these n elements. This means that for every v in $\mathbb{F}_q^n$, one can find a polynomial V with $\deg(V)\le n-1$, such that $\mathrm{ev}(V)=v$; moreover, V is unique. With a slight abuse of notation, we write $V=\mathrm{ev}^{-1}(v)$. Of course, ev is a linear mapping, $\mathrm{ev}(\alpha\cdot P+\beta\cdot Q)=\alpha\cdot\mathrm{ev}(P)+\beta\cdot\mathrm{ev}(Q)$ for any polynomials P,Q and field elements α,β.

Thus, the evaluation mapping can be represented by the matrix

$$\Gamma=\begin{pmatrix}\mathrm{ev}(X^0)\\\mathrm{ev}(X^1)\\\mathrm{ev}(X^2)\\\cdots\\\mathrm{ev}(X^{n-1})\end{pmatrix}=\begin{pmatrix}\gamma_0^0&\gamma_1^0&\cdots&\gamma_{n-1}^0\\\gamma_0&\gamma_1&\cdots&\gamma_{n-1}\\\gamma_0^2&\gamma_1^2&\cdots&\gamma_{n-1}^2\\&&\vdots&\\\gamma_0^{n-1}&\gamma_1^{n-1}&\cdots&\gamma_{n-1}^{n-1}\end{pmatrix}. \tag{10}$$

If we denote by $\mathrm{Coeff}(V)\in\mathbb{F}_q^n$ the vector consisting of the coefficients of V, then $\mathrm{Coeff}(V)\cdot\Gamma=\mathrm{ev}(V)$. On the other hand, Γ being nonsingular, its inverse Γ^{-1} computes $\mathrm{Coeff}(V)$ from $\mathrm{ev}(V)$. For our purpose, it is noteworthy that the coefficients of monomials of degree at least k can be easily computed from $\mathrm{ev}(V)$, splitting Γ^{-1} in two parts

$$\Gamma^{-1}=\begin{pmatrix}\underbrace{A}_{k\text{ columns}}&\underbrace{B}_{n-k\text{ columns}}\end{pmatrix}, \tag{11}$$

$\mathrm{ev}(V)\cdot B$ is precisely the coefficients vector of the monomials of degree at least k in V. In fact, B is the transpose of a parity check matrix of $\mathrm{GRS}(n,k)$, since a vector c is an element of the code if and only if we have $c\cdot B=0$. So, instead of B, we write H^t, as it is usually done.

4.2. A Polynomial View of Cosets. Now, let us look at the cosets of $\mathrm{GRS}(n,k)$. A coset is a set of the type $y+\mathrm{GRS}(n,k)$, with $y\in\mathbb{F}_q^n$ not in $\mathrm{GRS}(n,k)$. As usual with linear codes, a coset is uniquely identified by the vector $y\cdot H^t$, syndrome of y. In the case of GRS codes, this vector consists of the coefficients of monomials of degree at least k.

4.3. Decoding Reed-Solomon Codes

4.3.1. General Case. Receiving a vector v, the output of the decoding algorithm may be

(i) a single polynomial P, if it exists, such that the vector $\mathrm{ev}(P)$ is at distance at most $\lfloor(n-k+1)/2\rfloor$ from v (remark that if such a P exists, it is unique), and nothing otherwise;

(ii) a list of all polynomials P such that the vectors $\mathrm{ev}(P)$ are at distance at most λ from v, λ being an input parameter.

The second case corresponds to the so-called list decoding; an efficient algorithm for GRS codes was initially provided by [23], and was improved by [24], leading to the Guruswami-Sudan (GS) algorithm.

We just set here the outline of the GS algorithm, providing more details in the appendix. The Guruswami-Sudan algorithm uses a parameter called the interpolation multiplicity μ. For an input vector $(a_0, \ldots, a_{n-1})$, the algorithm computes a special bivariate polynomial $R(X, Y)$ such that each couple (γ_i, a_i) is a root of R with multiplicity μ. The second and last step is to compute the list of factors of R, of the form $Y - P(X)$, with $\deg(P) \leq k - 1$. For a fixed μ, the list contains all the polynomials which are at distance at most $\lambda_\mu \approx n - \sqrt{(1 + (1/\mu))(k-1)n}$. The maximum decoding radius is, thus, $\lambda_{\mathrm{GS}} = n - 1 - \lfloor \sqrt{n \cdot (k-1)} \rfloor$. Moreover, the overall algorithm can be performed in less than $\mathcal{O}(n^2\mu^4)$ arithmetic operations over $\mathbb{F}_q$.

4.3.2. Shortened GRS Case. The Guruswami-Sudan algorithm can be used for decoding shortened GRS codes. For a fixed set $\mathcal{I}$ of indices, we are looking for polynomials P such that $\deg(P) < k$, $P(\gamma_i) = 0$ for $i \in \mathcal{I}$ and $P(\gamma_i) = Q(\gamma_i)$ for as many $i \notin \mathcal{I}$ as possible. Such P can be written as $P(X) = F(X)G(X)$ with $F(X) = \prod_{i \in \mathcal{I}}(X - \gamma_i)$. Hence, decoding the shortened code reduces to obtain G such that $\deg(G) < k - |\mathcal{I}|$ and $G(\gamma_i) = (Q/F)(\gamma_i)$ for as many $i \notin \mathcal{I}$ as possible. Stated differently, it reduces to decode in $\mathrm{GRS}(n - |\mathcal{I}|, k - |\mathcal{I}|)$, which can be done by the GS algorithm.

5. What Can Reed-Solomon Codes Do?

Our problem is the following. We have a vector v of n symbols of $\mathbb{F}_q$, extracted from the cover-medium, and a message m. We want to modify v into s such that m is embedded in s, changing at most T coordinates in v.

The basic principle is to use syndrome coding with a GRS code. We use the cosets of a GRS code to embed the message, finding a vector s in the proper coset, close enough to v. Thus, we suppose that we have fixed $\gamma_0, \ldots, \gamma_{n-1} \in \mathbb{F}_q$, constructed the matrix Γ whose ith row is $\mathrm{ev}(X^i)$, and inverted it. In particular, we denote by H^t the last $n - k$ columns of Γ^{-1}, and therefore, according to section Section 4.1, H is a parity-check matrix. Recall that a word s embeds the message m if $s \cdot H^t = m$.

To construct s, we need a word y such that its syndrome is $m - v \cdot H^t$; thus, we can set $s = y + v$, which leads to $s \cdot H^t = y \cdot H^t + v \cdot H^t = m$. Moreover, the Hamming weight of y is precisely the number of changes we apply to go from v to s; so, we need $w(y) \leq T$.

When T is equal to the covering radius of the code corresponding to H, such a vector y always exists. But, explicit computation of such a vector y, known as the bounded syndrome decoding problem, is proved to be NP-hard for general linear codes. Even for families of deeply structured codes, we usually do not have polynomial time (in the length n) algorithms to solve the bounded syndrome decoding problem up to the covering radius. This is precisely the problem faced by [12].

GRS codes overcome this problem in a nice fashion. It is easy to find a vector with syndrome $m = (m_0, \ldots, m_{n-1-k})$. Let us consider the polynomial $M(X)$ that has coefficient m_i for the monomial X^{k+i}, $i \in \{0, \ldots, n - 1 - k\}$; according to the previous section, we have $\mathrm{ev}(M) \cdot H^t = m$. Now, finding y can be done by computing a polynomial P of degree less than k such that for at least k elements $\gamma \in \{\gamma_0, \ldots, \gamma_{n-1}\}$, we have $P(\gamma) = M(\gamma) - V(\gamma)$. With such a P, the vector $y = \mathrm{ev}(M - V - P)$ has at least k coordinates equal to zero, and the correct syndrome value. Hence, $T = n - k$ and the challenge lies in the construction of P.

It is noteworthy to remark that locking the position i, that is, requiring $s_i = v_i$, is equivalent to require $y_i = 0$ and, thus, to ask for $P(\gamma_i) = M(\gamma_i) - V(\gamma_i)$.

5.1. A Simple Construction of P

5.1.1. Using Lagrange Interpolation. A very simple way to construct P is Lagrange interpolation. We choose k coordinates $\mathcal{I} = \{i_1, \ldots, i_k\}$ and compute

$$P(X) = \sum_{i \in \mathcal{I}} (M(\gamma_i) - V(\gamma_i)) \cdot L_{\mathcal{I}}^{(i)}(X), \tag{12}$$

where $L_{\mathcal{I}}^{(i)}$ is the unique polynomial of degree at most $k - 1$ taking values 0 on γ_j, $j \neq i$ and 1 on γ_i, that is,

$$L_{\mathcal{I}}^{(i)}(X) = \prod_{j \in \mathcal{I} \setminus \{i\}} (\gamma_i - \gamma_j)^{-1}(X - \gamma_j). \tag{13}$$

The polynomial P we obtain by this way clearly satisfies $P(\gamma_i) = M(\gamma_i) - V(\gamma_i)$ for any $i \in \mathcal{I}$ and, thus, can match $y = \mathrm{ev}(M - V - P)$. As pointed out earlier, since, for $i \in \mathcal{I}$, we have $y_i = 0$, we also have $s_i = v_i + y_i = v_i$, *that is,* positions in $\mathcal{I}$ are locked.

The above proposed solution has a nice feature; by choosing $\mathcal{I}$, we can choose the coordinates on which s and v are equal, and this does not require any loss in computational complexity or embedding efficiency. This means that we can perform the syndrome decoding directly with the additional requirement of wet papers keeping unchanged the coordinates whose modifications are detectable.

5.1.2. Optimal Management of Locked Positions. We can embed $r = n - k$ elements of $\mathbb{F}_q$, changing not more than $T = n - k$ coordinates, so the embedding efficiency r/T is equal to 1 in the worst case. But, we can lock *any* k positions to embed our information.

This is to be compared with [12], where binary BCH codes are used. In [12], the maximal number of locked positions, without failing to embed the message m, is experimentally estimated to be $k/2$. To be able to lock up to $k - 1$ positions, it is necessary to allow a nonzero probability of nonembedding. It is also noteworthy that the average embedding efficiency decreases fast.

In fact, embedding $r = n - k$ symbols while locking k symbols amongst n is optimal. We said in Section 3 that locking the positions in $\mathcal{I}$ leads to an equation $y \cdot H_{\mathcal{I}}^t = m$, where $H_{\mathcal{I}}$ has dimension $(n-k)\times(n-|\mathcal{I}|)$. So, when $|\mathcal{I}| > k$, there exist some values m for which there is no solution. On the other hand, let us suppose we have a code with parity check matrix H such that for any $\mathcal{I}$ of size k, and any m, this equation has a solution, that is, $H_{\mathcal{I}}$ is invertible. This means that any $(n-k)\times(n-k)$ submatrix of H is invertible. But, it is known that this is equivalent to require the code to be MDS (see, e.g., [19, Corollary 1.4.14]), which is the case of GRS codes. Hence, GRS codes are optimal in the sense that we can lock as many positions as possible, that is, up to k for a message length of $r = n - k$.

5.2. A More Efficient Construction of P. If the number of locked positions is less than k, Lagrange interpolation is not optimal since it changes $n - k$ positions, almost always. Unfortunately, Lagrange interpolation is unable to use the additional freedom brought by fewer locked positions.

A possible way to address this problem is to use a decoding algorithm in order to construct P, that is, we try to decode $\mathrm{ev}(M - V)$. Locked positions can be dealt with as explained in Section 3.2. If it succeeds, we get a P in the ball centered on $\mathrm{ev}(M - V)$ of radius λ, where λ is the decoding radius of the decoding algorithm. Here, the Guruswami-Sudan algorithm helps; it provides a large λ, that is, greater chances of success, and outputs a list of P which allows to choose the best one with respect to some additional constraints on undetectability. In case of a decoding failure, we can add a new locked position and retry. If we already have k locked positions, we fall back on Lagrange interpolation.

5.2.1. Algorithm Description. We start with the "while loop" of the algorithm. So suppose that we have a set $\mathcal{I}$ of positions to lock. Let $L(X)$ be the Lagrange interpolation polynomial for $\{(\gamma_i, M(\gamma_i) - V(\gamma_i))\}$, that is, $L(\gamma_i) = M(\gamma_i) - V(\gamma_i)$ for all $i \in \mathcal{I}$. Thus, we can write $M(X) - V(X) - L(X) = F(X)G(X)$ with $F(X) = \prod_{i\in\mathcal{I}}(X - \gamma_i)$. We perform a GS decoding on $G(X)$ in $\mathrm{GRS}(n - |\mathcal{I}|, k - |\mathcal{I}|)$, that is, we compute the list of polynomials $U(X)$ such that $\deg(U) < k - |\mathcal{I}|$ and

$$U(\gamma_i) = \left(\frac{M - V - L}{F}\right)(\gamma_i) \tag{14}$$

for at least $n - |\mathcal{I}| - \lambda$ values $i \in 0, \ldots, n-1 \subset \mathcal{I}$, where λ is the decoding radius of the GS algorithm, which depends on $n - |\mathcal{I}|$ and $k - |\mathcal{I}|$. If the decoding is successful, then $\mathrm{ev}(F(X)U(X))$ has zeros on positions in $\mathcal{I}$ and is equal to $\mathrm{ev}(M(X) - V(X) - L(X))$ for at least $n - |\mathcal{I}| - \lambda$ positions $i \in \{0, \ldots, n-1\} \setminus \mathcal{I}$. Pick up U such that the distortion induced by $y = \mathrm{ev}(M - V - L - FU)$ is as low as possible. Remark that here P is equal to $L - FU$.

The full algorithm (see Algorithm 1) is simply a while loop on the previous procedure, at the end of which, in case of a decoding failure, we add a new position to $|\mathcal{I}|$. Before commenting the algorithm, let us describe the three external procedures that we use:

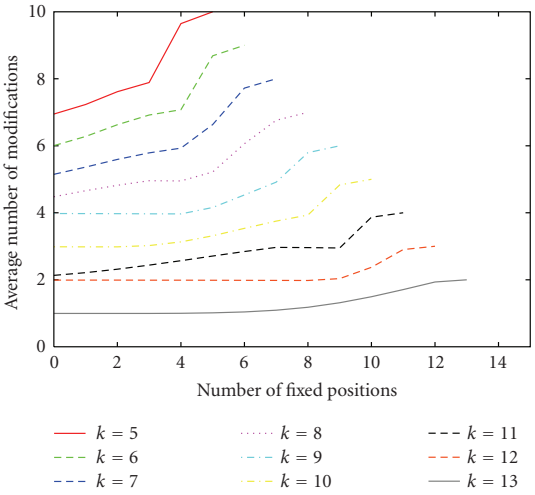


FIGURE 1: Average number of changes with respect to the number of locked positions for $q = 16$. Only curves with $\Delta\omega \geq 0.3$ are plotted.

(i) the `Lagrange`$(Q(X), \mathcal{I})$ procedure outputs a polynomial L such that $L(\gamma_i) = Q(\gamma_i)$ for all $i \in \mathcal{I}$ and $\deg(L) < |\mathcal{I}|$;

(ii) the `GSdecode` procedure refers to the Guruswami-Sudan list decoding (Section 4.3.1). For the sake of simplicity, we just write `GSdecode`$(Q(X), \mathcal{I})$ for the output list of the GS decoding of $(Q(\gamma_{i_0}), \ldots, Q(\gamma_{i_{n'-1}}))$, $i_j \in \{0, \ldots, n-1\} \setminus \mathcal{I}$ with respect to $\mathrm{GRS}(n - |\mathcal{I}|, k - |\mathcal{I}|)$. So, this procedure returns a good approximation $U(X)$ of $Q(X)$, on the evaluation set, of degree less than $k - |\mathcal{I}|$;

(iii) the `selectposition` procedure returns an integer from the set given as a parameter. This procedure is used to choose the new position to lock before retrying list decoding.

Lines 1 to 5 of the algorithm depicted in Algorithm 1 simply do the setup for the while loop. The while loop, Lines 6 to 12, tries to use list decoding to construct a good solution, as described above. Remark that if all GS decodings fail, we have $Y = M - V - L$ with L is equal to polynomial P of Section 5.1, that is, we just fall back on Lagrange interpolation. Lines 13 to 16 use the result of the while loop in case of a decoding success, according to the details given above.

Correctness of this algorithm follows from the fact that through the whole algorithm we have $\mathrm{ev}(Y) \cdot H^t = m - v \cdot H^t$ and $Y(\gamma_i) = 0$ for $i \in \mathcal{I}$. Termination is clear since each iteration of the Loop 6-12 increases $|\mathcal{I}|$.

5.2.2. Algorithm Analysis. The most important property of embedding algorithms is the number of changes introduced during the embedding. Let $\omega(n, k, i)$ be the average number of such changes when GRS (n, k) is used and i positions are locked. For our algorithm, this quantity depends on two parameters related to the Guruswami-Sudan algorithm:

Inputs: $v = (v_0, \ldots, v_{n-1})$, the cover-data
$m = (m_0, \ldots, m_{n-k-1})$, symbols to hide
$\mathcal{I}$, set of coordinates to remain unchanged, $|\mathcal{I}| \leq k$
Output: $s = (s_0, \ldots, s_{n-1})$, the stego-data
$(s \cdot H^t = m; s_i = v_i, i \in \mathcal{I}; d_H(s, v) \leq n - k)$

(1) $V(X) \Leftarrow v_0 X^0 + \cdots + v_{n-1} X^{n-1}$
(2) $M(X) \Leftarrow m_0 X^k + \cdots + m_{n-k-1} X^{n-1}$
(3) $L(X) \Leftarrow \texttt{Lagrange}(M - V, \mathcal{I})$
(4) $Y(X) \Leftarrow M(X) - V(X) - L(X)$
(5) $F(X) \Leftarrow \texttt{Lagrange}(0, \mathcal{I})$
(6) **while** $|\mathcal{I}| < k$ and $\texttt{GSdecode}(\frac{Y}{F}, \mathcal{I}) = \theta$ **do**
(7) $\quad i \Leftarrow \texttt{selectposition}(\{0, \ldots, n-1\} \setminus \mathcal{I})$
(8) $\quad \mathcal{I} \Leftarrow \mathcal{I} \cup \{i\}$
(9) $\quad L(X) \Leftarrow \texttt{Lagrange}(M - V, \mathcal{I})$
(10) $\quad F(X) \Leftarrow \texttt{Lagrange}(0, \mathcal{I})$
(11) $\quad Y(X) \Leftarrow M(X) - V(X) - L(X)$
(12) **end while**
(13) **if** $\texttt{GSdecode}(\frac{Y}{F}, \mathcal{I}) \neq \theta$ **then**
(14) $\quad U(X) \Leftarrow \texttt{GSdecode}(\frac{Y}{F}, \mathcal{I})$
(15) $\quad Y(X) \Leftarrow Y(X) - F(X)U(X)$
(16) **end if**
(17) $s \Leftarrow v + \mathrm{ev}(Y)$
(18) **return** s

Algorithm 1: Algorithm for embedding with locked positions using a GRS(n, k) code $(\gamma_0, \ldots, \gamma_{n-1}$ fixed). It embeds $r = n - k$ $\mathbb{F}_q$ symbols with up to k locked positions and at most $n - k$ changes.

(i) the probability $p(n, k)$ that the list decoding of a word in $\mathbb{F}_q^n$ outputs a nonempty list of codewords in GRS (n, k);

(ii) the average distance $\delta(n, k)$ between the closest codewords in the (nonempty) list and the word to decode.

We denote by $q(n, k)$ the probability of an empty list and for conciseness let $n' = n - |\mathcal{I}|$, $k' = k - |\mathcal{I}|$. Thus, the probability that the first $\ell - 1$ list decodings fail and the ℓth succeeds can be written as $p^*(\ell) \prod_{e=0}^{\ell-1} q^*(e)$ with $p^*(\ell) = p(n' - \ell, k' - \ell)$ and $q^*(e) = q(n' - e, k' - e)$. Remark that in this case, $\delta^*(\ell) = \delta(n' - \ell, k' - \ell)$ coordinates are changed on average.

Now, the average number of changes required to perform the embedding can be expressed by the following formula:

$$\omega(n, k, i) = \left(\sum_{\ell=0}^{k'-1} \delta^*(\ell) \cdot p^*(\ell) \prod_{e=0}^{\ell-1} q^*(e) \right) + (n - k) \prod_{e=0}^{k'-1} q^*(e). \tag{15}$$

(a) Estimating p and δ. To (upper) estimate $p(n, k)$, we proceed as follows. Let Z be the random variable equal to the size of the output list of the decoding algorithm. The Markov inequality yields $\Pr(Z \geq 1) \leq \mathbb{E}(Z)$, where $\mathbb{E}(Z)$ denotes the expectation of Z. But, $\Pr(Z \geq 1)$ is the probability that the list is nonempty and, thus, $\Pr(Z \geq 1) = p(n, k)$. Now, $\mathbb{E}(Z)$ is the average number of elements in the output list, but this is exactly the average number of codewords in a Hamming ball of radius λ_{GS}. Unfortunately, no adequate information can be found in the literature to properly estimate it; the only paper studying a similar quantity is [25], but it cannot be used for our $\mathbb{E}(Z)$.

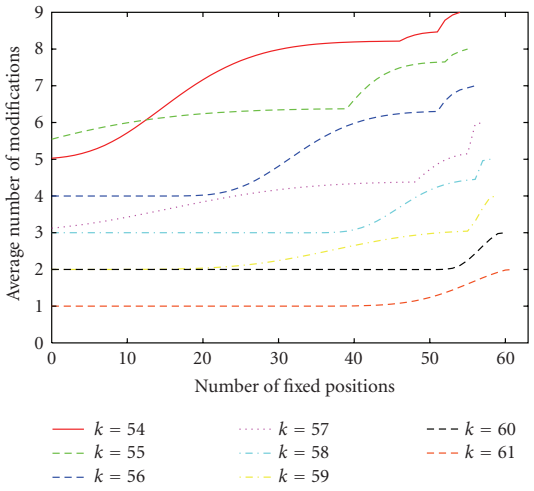


Figure 2: Average number of changes with respect to the number of locked positions for $q = 64$. Only curves with $\Delta\omega \geq 0.3$ are plotted.

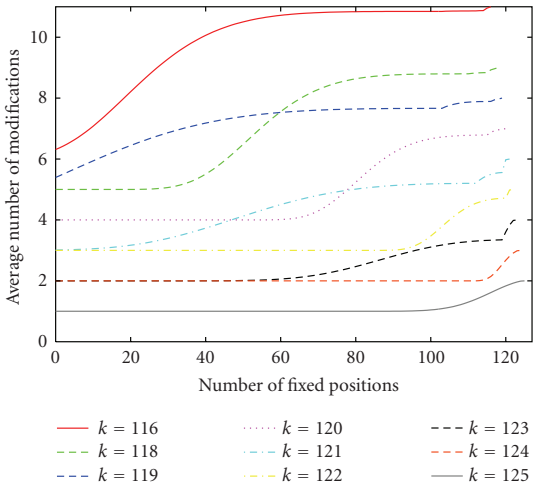


FIGURE 3: Average number of changes with respect to the number of locked positions for $q = 128$. Only curves with $\Delta\omega \geq 0.3$ are plotted.

So, we set

$$\mathbb{E}(Z) = \frac{q^k}{q^n} \cdot V_{\lambda_{GS}} = \frac{\sum_{i=0}^{\lambda_{GS}} (q-1)^i \binom{n}{i}}{q^{n-k}}, \tag{16}$$

where $V_{\lambda_{GS}}$ is the volume of a ball of radius λ_{GS}. This would be the correct value if GRS codes were *random* codes over $\mathbb{F}_q$ of length n, with q^k codewords uniformly drawn from $\mathbb{F}_q^n$. That is, we estimate $\mathbb{E}(Z)$ as if GRS codes were random codes. Thus, we use $\overline{p} = \min(1, q^{k-n} V_{\lambda_{GS}})$ to upper estimate p.

The second parameter we need is $\delta(n,k)$, the average number of changes required when the list is nonempty. We consider that the closest codeword is uniformly distributed over the ball of radius λ_{GS} and, therefore, we have

$$\delta(n,k) = \frac{\sum_{i=0}^{\lambda_{GS}} i \cdot (q-1)^i \binom{n}{i}}{V_{\lambda_{GS}}}. \tag{17}$$

(b) Estimating the Average Number of Changes. Using our previous estimations for $p(n,k)$ and $\delta(n,k)$, we plotted $\omega(n,k,i)$ in Figure 1 ($q = 16$), Figure 2 ($q = 64$), Figure 3 ($q = 128$). For each figure, we set $n = q - 1$ and plotted ω for several values of k.

Remember that $i \leq k$ and that when $i = k$, our algorithm simply uses Lagrange interpolation, which leads to the maximum number of changes, that is, $\omega(n,k,k) = n - k$. On the other side, when $i = 0$, our algorithm tries to use Guruswami-Sudan algorithm as much as possible. Therefore, our algorithm improves upon the simpler Lagrange interpolation when

$$\Delta\omega = \frac{\omega(n,k,k) - \omega(n,k,0)}{n-k} \tag{18}$$

is large. A second criterion to estimate the performance is the slope of the plotted curves, the slighter, the better.

With this in mind, looking at Figure 1, we can see that $k = 13$ provides good performances; $\Delta\omega = 0.5$, which means that list decoding avoids up to 50% of the changes required by Lagrange interpolation, and on the other hand, the slope is nearly 0 when $i \leq 8$. For higher embedding rate, all values of k less than 3 have $\Delta\omega \geq 0.28$.

In Figure 2, $\Delta\omega \geq 0.3$ for $k \geq 54$. In Figure 3, $\Delta\omega \geq 0.3$ for $k \geq 116$, except for $k = 117$. Remark that $k = 120$, the slope is nearly 0 for $i \leq 70$, which means that we can lock about half the coordinates and still have $\Delta\omega = 42\%$ of improvement with respect to Lagrange interpolation.

6. Conclusion

We have shown in this paper that Reed-Solomon codes are good candidates for designing efficient steganographic schemes. They enable to mix wet papers (locked positions) and simple syndrome coding (small number of changes) in order to face not only passive but also active wardens. If we compare them to the previous studied codes, as binary BCH codes, Reed-Solomon codes improve the management of locked positions during embedding, hence ensuring a better management of the distortion; they are able to lock twice the number of positions. Moreover, they are optimal in the sense that they enable to lock the maximal number of positions. We first provide an efficient way to do it through Lagrange interpolation. We then propose a new algorithm based on Guruswami-Sudan list decoding, which is slower but provides an adaptive tradeoff between the number of locked positions and the average number of changes.

In order to use them in real applications, several issues still have to be addressed. First, we need to choose an appropriate measure to properly estimate the distortion induced at the medium level when modifying the symbols at the data level. Second, we need to use a nonbinary, and preferably large, alphabet. A straightforward way to deal with this would be to simply regroup bits to obtain symbols of our alphabet and consider that a symbol should be locked if it contains a bit that should be. Unfortunately, it would lead to a large number of locked symbols (e.g., 5% of locked bits leads to up to 20% of locked symbols if we use GF(16)). A better way would be to use grid coloring [26], keeping a 1-to-1 ratio. But, the price to this 1-to-1 ratio would be a cut in payload. We think a good solution has yet to be figured out. Nevertheless, in some settings, a large alphabet arises naturally; for example, in [14], a (binary) wet paper code is used on the syndromes of a $[2^k - 1, 2^k - k - 1]$ Hamming code, some of these syndromes being locked; here, since whole syndromes are locked, we can view syndromes as elements of the larger field GF(2^k) and use our proposal. Third, no efficient implementation of the Guruswami-Sudan list decoding algorithm is available. And, as the involved mathematical problems are really tricky, only a specialist can perform a real efficient one. Today, these three issues remain open.

Appendix

Guruswami-Sudan Algorithm

We provide here the core of the Guruswami-Sudan algorithm, without deep details on (important) algorithms that are required to achieve a good complexity (the interested reader may refer to [19, 24, 25]).

A.1. Description. Recall we have a vector $\mathrm{ev}(Q) = (Q(\gamma_0), \ldots, Q(\gamma_{n-1}))$ and we want to find all polynomials P such that $\mathrm{ev}(P)$ is at distance at most λ from $\mathrm{ev}(Q)$, and $\deg(P) < k$. We construct a bivariate polynomial R over $\mathbb{F}_q$ such that $R(\gamma_i, P(\gamma_i)) = 0$ for all P at distance at most λ from Q. Then, we compute all P from a factorization of R.

First, let us define what is called the multiplicity of a zero for bivariate polynomial: $R(X, Y)$ has a zero (a, b) of multiplicity μ if and only if the coefficients of the monomials $X^i Y^j$ in $R(X+a, Y+b)$ are equal to zero for all i, j with $i+j < \mu$. This leads to $\binom{\mu+1}{2}$ linear equations in the coefficients of R. Writing $R(X, Y) = \sum_{i,j} r_{i,j} X^i Y^j$, then $R(X+a, Y+b) = \sum_{i,j} r_{i,j}(a, b) X^i Y^j$ with

$$r_{i,j}(a, b) = \sum_{\substack{i' \geq i \\ j' \geq j}} \binom{i'}{i} \binom{j'}{j} r_{i',j'} a^{i'-i} b^{j'-j}. \tag{A.1}$$

Since a multiplicity μ in (a, b) is exactly $r_{i,j}(a, b) = 0$ for $i+j < \mu$, and we have $\binom{\mu+1}{2}$ values of i and j such that $i + j < \mu$, we have the right number of equations.

The principle is to use the $n\binom{\mu+1}{2}$ linear equations in the coefficients of R, obtained by requiring $(\gamma_i, Q(\gamma_i))$ to be a zero of R with multiplicity μ for $i \in \{0, \ldots, n-1\}$. Solving this system leads to the bivariate polynomial R, but, to be sure our system has a solution, we need more unknowns than equations. To address this point, we impose a special shape on R. For a fixed integer ℓ, we set $R(X, Y) = \sum_{j \leq \ell} R_j(X) Y^j$ with the restriction that $\deg(R_j) \leq \mu(n-\lambda) - j(k-1)$. Thus, R has at most

$$\sum_{j \leq l} \deg(R_j) = (\ell+1)\mu(n-\lambda) - \frac{\ell(\ell+1)}{2}(k-1) \tag{A.2}$$

coefficients. Choosing ℓ such that $\sum_{j \leq \ell} \deg(R_j) > n\binom{\mu+1}{2}$ guarantees to have nonzero solutions. Of course, since degrees of R_j must be nonnegative integers, we have $\lambda \leq n - (\ell/\mu)(k-1)$.

On the other hand, under the conditions we imposed on R, one can prove that for all polynomials P of degree less than k and at distance at most λ from Q, $Y - P(X)$ divides $R(X, Y)$. Detailed analysis of the parameters shows it is always possible to take ℓ less than or equal to

$$\ell \leq \sqrt{\frac{k}{(k-1)^2} n(\mu+1)\mu} \tag{A.3}$$

(see [19, Chapter 5]). Thus, we have the formula $\lambda \approx n - 1 - \lfloor \sqrt{n(k-1)(1+(1/\mu))} \rfloor$, which leads to the maximum radius $\lambda_{\mathrm{GS}} = \max_{\mu \geq 1} \lambda = n - 1 - \lfloor \sqrt{n(k-1)} \rfloor$ for μ large enough.

A.2. Complexity. Using $\ell = m\sqrt{n/k}$ in (A.2), there are $n\binom{\mu}{2}$ linear equations with roughly $n\mu^2$ unknowns. Solving these equations with fast general linear algebra can be done in less than $\mathcal{O}(n^{5/2}\mu^5)$ arithmetic operations over $\mathbb{F}_q$ (see [27, Chapter 12]).

Finding the factor $Y - P(X)$ can be achieved in a simple way, considering an extension of $\mathbb{F}_q$ of order k. A (univariate) polynomial P over $\mathbb{F}_q$ of degree less than k can be uniquely represented by an element $\widetilde{P}$ of $\mathbb{F}_{q^k}$ and, under this representation, to find factors $Y - P(X)$ of R is equivalent to find factors $Y - \widetilde{P}$ of $\widetilde{R}(Y) = \sum_{j \leq l} \widetilde{R}_j Y^j$, that is, to compute factorization of a univariate polynomial of degree ℓ over $\mathbb{F}_{q^k}$ which can be done in at most $\mathcal{O}(\mu \cdot \sqrt{n \cdot k^3})$ operations over $\mathbb{F}_q$, neglecting logarithmic factors (see [27, Chapter 14]).

The global cost of this basic approach is heavily dominated by the linear algebra part in $\mathcal{O}(n^{5/2}\mu^5)$ with a particularly large degree in μ. It is possible to perform the Guruswami-Sudan algorithm at a cheaper cost, still in $\mathcal{O}(n^2\mu^4)$, with less naive algorithms. Complete details can be found in [25].

To sum up, Guruswami-Sudan decoding algorithm finds polynomials P of degree at most k and at distance at most $n - 1 - \lfloor \sqrt{n(k-1)} \rfloor$ from Q using simple linear algebra and factorization of univariate polynomial over a finite field for a cost in less than $\mathcal{O}(n^{5/2}\mu^5)$ arithmetic operations in $\mathbb{F}_q$. This can be reduced to $\mathcal{O}(n^2\mu^4)$ with dedicated algorithms.

Acknowledgments

Dr. C. Fontaine is supported (in part) by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT and by the French National Agency for Research under Contract ANR-RIAM ESTIVALE. The authors are in debt to Daniel Augot for numerous comments on this work, in particular for pointing out the adaptation of the Guruswami-Sudan algorithm to shortened GRS used in the embedding algorithm.

References

[1] G. J. Simmons, "The prisoners' problem and the subliminal channel," in *Advances in Cryptology*, pp. 51–67, Plenum Press, New York, NY, USA, 1984.

[2] R. Böhme and A. Westfeld, "Exploiting preserved statistics for steganalysis," in *Proceedings of the 6th International Workshop on Information Hiding (IH '04)*, vol. 3200 of *Lecture Notes in Computer Science*, pp. 82–96, Springer, Toronto, Canada, May 2004.

[3] E. Franz, "Steganography preserving statistical properties," in *Proceedings of the 5th International Workshop on Information Hiding (IH '02)*, vol. 2578 of *Lecture Notes in Computer Science*, pp. 278–294, Noordwijkerhout, The Netherlands, October 2002.

[4] R. Crandall, Some notes on steganography. Posted on steganography mailing list, 1998, http://os.inf.tu-dresden.de/~westfeld/crandall.pdf.

[5] J. Bierbrauer, On Crandall's problem. Personal communication, 1998, http://www.ws.binghamton.edu/fridrich/covcodes.pdf.

[6] A. Westfeld, "F5—a steganographic algorithm: high capacity despite better steganalysis," in *Proceedings of the 4th International Workshop on Information Hiding (IH '01)*, vol. 2137 of *Lecture Notes in Computer Science*, pp. 289–302, Pittsburgh, Pa, USA, April 2001.

[7] F. Galand and G. Kabatiansky, "Information hiding by coverings," in *Proceedings of IEEE Information Theory Workshop (ITW '03)*, pp. 151–154, Paris, France, March-April 2003.

[8] J. Fridrich, M. Goljan, P. Lisonek, and D. Soukal, "Writing on wet paper," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, part 2, pp. 3923–3935, 2005.

[9] J. Fridrich, M. Goljan, and D. Soukal, "Efficient wet paper codes," in *Proceedings of the 7th International Workshop on Information Hiding (IH '05)*, vol. 3727 of *Lecture Notes in Computer Science*, pp. 204–218, Barcelona, Spain, June 2005.

[10] J. Fridrich, M. Goljan, and D. Soukal, "Wet paper codes with improved embedding efficiency," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 1, pp. 102–110, 2006.

[11] J. Fridrich and D. Soukal, "Matrix embedding for large payloads," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 3, pp. 390–395, 2006.

[12] D. Schönfeld and A. Winkler, "Embedding with syndrome coding based on BCH codes," in *Proceedings of the 8th Workshop on Multimedia and Security (MM&Sec '06)*, pp. 214–223, ACM, Geneva, Switzerland, September 2006.

[13] D. Schönfeld and A. Winkler, "Reducing the complexity of syndrome coding for embedding," in *Proceedings of the 9th International Workshop on Information Hiding (IH '07)*, vol. 4567 of *Lecture Notes in Computer Science*, pp. 145–158, Springer, Saint Malo, France, June 2007.

[14] W. Zhang, X. Zhang, and S. Wang, "Maximizing steganographic embedding efficiency by combining Hamming codes and wet paper codes," in *Proceedings of the 10th International Workshop on Information Hiding (IH '08)*, vol. 5284 of *Lecture Notes in Computer Science*, pp. 60–71, Santa Barbara, Calif, USA, May 2008.

[15] J. Bierbrauer and J. Fridrich, "Constructing good covering codes for applications in steganography," in *Transactions on Data Hiding and Multimedia Security III*, vol. 4920 of *Lecture Notes in Computer Science*, pp. 1–22, Springer, Berlin, Germany, 2008.

[16] J. Fridrich, M. Goljan, and D. Soukal, "Perturbed quantization steganography," *ACM Multimedia and Security Journal*, vol. 11, no. 2, pp. 98–107, 2005.

[17] A. Vardy, "The intractability of computing the minimum distance of a code," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1757–1766, 1997.

[18] A. McLoughlin, "The complexity of computing the covering radius of a code," *IEEE Transactions on Information Theory*, vol. 30, no. 6, pp. 800–804, 1984.

[19] W. C. Huffman and V. Pless, *Fundamentals of Error-Correcting Codes*, Cambridge University Press, Cambridge, UK, 2003.

[20] Y. Kim, Z. Duric, and D. Richards, "Modified matrix encoding technique for minimal distortion steganography," in *Proceedings of the 8th International Workshop on Information Hiding (IH '06)*, vol. 4437 of *Lecture Notes in Computer Science*, pp. 314–327, Springe, Alexandria, Va, USA, June 2006.

[21] F. Galand and G. Kabatiansky, "Steganography via covering codes," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '03)*, p. 192, Yokohama, Japan, June-July 2003.

[22] X. Zhang and S. Wang, "Stego-encoding with error correction capability," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E88-A, no. 12, pp. 3663–3667, 2005.

[23] M. Sudan, "Decoding of Reed Solomon codes beyond the error-correction bound," *Journal of Complexity*, vol. 13, no. 1, pp. 180–193, 1997.

[24] V. Guruswami and M. Sudan, "Improved decoding of Reed-Solomon and algebraic-geometry codes," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 1757–1767, 1999.

[25] R. J. McEliece, "The Guruswami-Sudan decoding algorithm for Reed-Solomon codes," IPN Progress Report 42-153, California Institute of Technology, Pasadena, Calif, USA, May 2003, http://tmo.jpl.nasa.gov/progress_report/42-153/153F.pdf.

[26] J. Fridrich and P. Lisonek, "Grid colorings in steganography," *IEEE Transactions on Information Theory*, vol. 53, no. 4, pp. 1547–1549, 2007.

[27] J. von zur Gathen and J. Gerhard, *Modern Computer Algebra*, Cambridge University Press, Cambridge, UK, 2nd edition, 2003.

Ensuring message embedding in wet paper steganography

Daniel Augot[1], Morgan Barbier[1], and Caroline Fontaine[2]

[1] Computer science laboratory of École Polytechnique
INRIA Saclay – Île de France
[2] CNRS/Lab-STICC and Télécom Bretagne, Brest, France

Abstract. Syndrome coding has been proposed by Crandall in 1998 as a method to stealthily embed a message in a cover-medium through the use of bounded decoding. In 2005, Fridrich *et al.* introduced wet paper codes to improve the undetectability of the embedding by enabling the sender to lock some components of the cover-data, according to the nature of the cover-medium and the message. Unfortunately, almost all existing methods solving the bounded decoding syndrome problem with or without locked components have a non-zero probability to fail.In this paper, we introduce a randomized syndrome coding, which guarantees the embedding success with probability one. We analyze the parameters of this new scheme in the case of perfect codes.

Keywords: steganography, syndrome coding problem, wet paper codes.

1 Introduction

Hiding messages in innocuous-looking *cover-media* in a *stealthy* way, steganography is the art of stealth communications. The sender and receiver may proceed by cover selection, cover synthesis, or cover modification to exchange messages. Here, we focus on the cover modification scenario, where the sender chooses some *cover-medium* in his library, and modifies it to carry the message she wants to send. Once the cover-medium is chosen, the sender extracts some of its components to construct a *cover-data* vector. Then, she modifies it to embed the message. This modified vector, called the *stego-data*, leads back to the *stego-medium* that is communicated to the recipient. In the case of digital images, the insertion may for example consist in modifying some of the images components, *e.g.* the luminance of the pixels or the values of some transform (DCT or wavelet) coefficients. For a given transmitted document, only the sender and receiver have to be able to tell if it carries an hidden message or not [33]. This means that the *stego-media*, which carry the messages, have to be

statistically indistinguishable from original media [6,7]. But statistical detectability of most steganographic schemes increases with *embedding distortion* [24], which is often measured with the number of embedding changes. Hence it is of importance for the sender to embed the message while modifying as less components of the cover-data as possible.

In 1998, Crandall proposed to model the embedding and extraction process with the use of linear error correcting codes. He proposed to use Hamming codes, which are covering codes [9]. The key idea of this approach, called *syndrome coding*, or *matrix embedding*, is to modify the cover-data to obtain a stego-data lying in the *right coset* of the code, its *syndrome* being precisely equal to the message to hide. Later on, it has been showed that designing steganographic schemes is precisely equivalent to designing covering codes [3,22,23], meaning that this covering codes approach is not restrictive. Moreover, it has been shown to be really helpful and efficient to minimize the embedding distortion [3,22,23,4]. It has also been made popular due to its use in the famous steganographic algorithm F5 [36]. For all these reasons, this approach is of interest.

The process which states which components of the cover-data can actually be modified is called the *selection channel* [1]. Since the message embedding should introduce as little distortion as possible, the selection channel is of utmost importance. The selection channel may be arbitrary, but a more efficient approach is to select it dynamically during the embedding step, accordingly to the cover-medium and the message. This leads to a better undetectability, and makes attacks on the system harder to run, but in this context the extraction of the hidden message is more difficult as the selection channel is only known to the sender, and not to the recipient. *Wet Paper Codes* were introduced to tackle this non-shared selection channel, through the notions of *dry* and *wet* components [18]. By analogy with a sheet of paper that has been exposed to rain, we can still write easily on dry spots whereas we cannot write on wet spots. The idea is, adaptively to the message and the cover-medium, to *lock* some components of the cover-data — the wet components — to prevent them being modified. The other components — the dry components — of the cover-data remain free to be modified to embed the message.

Algorithmically speaking, syndrome coding provides the recipient an easy way to access the message, through a simple syndrome computation. But to embed the message, the sender has to tackle an harder challenge, linked with bounded syndrome coding. It has been shown that if random codes may seem interesting for their asymptotic behavior, their use leads to solve really hard problems: syndrome decoding

and covering radius computation, which are proved to be NP-complete and Π_2-complete respectively [34,25]. Moreover, no efficient decoding algorithm is known, for generic, or random, codes. Hence, attention has been given on structured codes to design Wet Paper Codes: Hamming codes [9,21], Simplex codes [20], BCH codes [31,32,37,30,27], Reed-Solomon codes [14,15], perfect product codes [29,28], low density generator matrix codes [17,39,38,10], and convolutional codes [13,11,12].

Embedding techniques efficiency is usually evaluated through their relative payload (number of message symbols per cover-data (modifiable) symbol) and average embedding efficiency (average number of message symbols per cover-data modification). Today, we can find in the literature quasi-optimal codes in terms of average embedding efficiency and payload [17,39,38,16,10]. Nevertheless, we are interested here in another criterion, which is usually not discussed: the probability for the embedding to fail. In fact, the only case for which it never fails is when using perfect codes (a), without locking any component of the cover-data (b). But very few codes are perfect (namely the Hamming and Golay codes), and their average embedding efficiency is quite low. Moreover it is really important in practice to be able to lock some components of the cover-data. Hence, efficient practical schemes usually do not satisfy either condition (a) or condition (b), leading to a non-zero probability for the embedding to fail. And this probability increases with the number of locked components. More precisely, syndrome coding usually divides the whole message into fragments, that are separately inserted in different cover-data vectors (coming from one or several cover-medium). Inserting each fragment involves finding a low weight solution of a linear system which may not always have a solution for a given set of locked components. Consequently, the probability that the whole message can be embedded decreases exponentially with the number of fragments to hide and with the number of locked components [21].

Hence, we have to decide what to do when embedding fails. In the common scenario where the sender has to choose a cover-medium in a huge collection of documents, she can drop the cover-medium that leads to a failure and choose another one, iterating the process until finding a cover-medium that is adequate to embed the message. Another solution may be to cut the message into smaller pieces, in order to have shorter messages to embed, and a lower probability of failure. If none of these is possible, for example if the sender only has few pieces of content, she may unlock some locked components [13] to make the probability of failure decrease. But, even doing this modified embedding, and decreasing the

probability of failure, the sender will not be able to drop it to zero, except if she falls back to perfect codes without locked components.

In this paper, we consider the "worst case" scenario, where the sender does not have too much cover documents to hide his message in, and then absolutely needs embedding to succeed. This scenario is not the most studied one, and concerns very constrained situations. Our contribution is to propose an embedding scheme that will never fail, and does not relax the management of locked components of his cover-data to make embedding succeed. It is, to our knowledge, the first bounded syndrome coding scheme that manages locked components while guaranteeing the complete embedding of the message for any code, be it perfect or not. To do so, we modify the classical syndrome coding approach by using some part of the syndrome for randomization. Of course, as the message we can embed is now shorter than the syndrome, there is a loss in terms of embedding efficiency. We analyze this loss in the case of linear perfect codes. Moreover, inspired by the ZZW construction [39], we show how the size of the random part of the syndrome, which is dynamically estimated during embedding, can be transmitted to the recipient without any additional communication.

The paper is organized as follows. Basic definitions and notation on both steganography and syndrome coding are introduced in Section 2. The traditional syndrome coding approach is recalled at the end of this section. In Section 3, we show how to slightly relax the constraints on the linear system to make it always solvable, and also estimate the loss of embedding efficiency. We discuss the behavior of our scheme in the case of the Golay and Hamming perfect codes in Section 4. Finally, as our solution uses a parameter r that is dynamically computed during embedding, we provide in Section 5 a construction that enables to transmit r to the recipient through the stego-data itself, that is, without any parallel or side-channel communication. We finally conclude in Section 6.

2 Steganography and coding theory

2.1 Steganographic schemes

We define a *stego-system* (or a *steganographic scheme*) by a pair of functions, Emb and Ext. Emb embeds the message $\mathbf{m}$ in the cover-data $\mathbf{x}$, producing the stego-data $\mathbf{y}$, while Ext extracts the message $\mathbf{m}$ from the stego-data $\mathbf{y}$. To make the embedding and extraction work properly, these functions have to satisfy the following properties.

Definition 1 (Stego-System). *Let $\mathcal{A}$ a finite alphabet, $r, n \in \mathbb{N}$ such that $r < n$, $\mathbf{x} \in \mathcal{A}^n$ denote the* cover-data*, $\mathbf{m} \in \mathcal{A}^r$ denote the message to embed, and T be a strictly positive integer. A* stego-system *is defined by a pair of functions Ext and Emb such that:*

$$Ext(Emb(\mathbf{x}, \mathbf{m})) = \mathbf{m} \tag{1}$$

$$d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{m})) \leq T \tag{2}$$

where $d(.,.)$ denoting the Hamming distance over $\mathcal{A}^n$.

Two quantities are usually used to compare stego-systems: the embedding efficiency and the relative payload, which are defined as follows.

Definition 2 (Embedding efficiency). *The* average embedding efficiency *of a stego-system, is usually defined by the ratio of the number of message symbols we can embed by the average number of symbols changed. We denote it by e.*

Definition 3 (Relative payload). *The* relative payload *of a stego-system, denoted by α, is the ratio of the number of message symbols we can embed by the number of (modifiable) symbols of covered data.*

For q-ary syndrome coding, the sphere-covering bound gives an upper bound for the embedding efficiency [16]. Note that it is usually stated for binary case, using the binary entropy function.

Proposition 1 (Sphere-covering bound). *For any q-ary stego-system $\mathcal{S}$, the* sphere-covering bound *gives*

$$e \leq \frac{\alpha}{\mathcal{H}_q^{-1}(\alpha)},$$

where $\mathcal{H}_q^{-1}()$ denotes the inverse function of the q-ary entropy $\mathcal{H}_q(x) = x \log_q(q-1) - x \log_q(x) - (1-x) \log_q(1-x)$ on $[0, 1-1/q]$, and α is the relative payload associated with $\mathcal{S}$.

2.2 From coding theory to steganography

This section recalls how coding theory may help embedding the message, and how it tackles the non-shared selection channel paradigm. In the rest of paper, the finite alphabet $\mathcal{A}$ is a finite field of cardinal q, denoted $\mathbb{F}_q$.

Here we focus on the use of linear codes, which is the most studied. Let $\mathcal{C}$ be a $[n, k, d]_q$-linear code, with parity check matrix H and covering

radius ρ — it is the smallest integer such that the balls of radius ρ centered on $\mathcal{C}$'s codewords cover the whole ambient space $\mathbb{F}_q^n$. A syndrome coding scheme based on $\mathcal{C}$ basically modify the cover-data $\mathbf{x}$ in such a way that the syndrome $\mathbf{y}H^t$ of the stego-data $\mathbf{y}$ will precisely be equal to the message $\mathbf{m}$. Determining which symbols of $\mathbf{x}$ to modify leads to finding a solution of a particular linear system that involves the parity check matrix H. This embedding approach has been introduced by Crandall in 1998 [9], and is called *syndrome coding* or *matrix embedding*.

We formulate several embedding problems. The first one addresses only Eq. (1) requirements, whereas the second one also tackles Eq. (2).

Problem 1 (Syndrome coding problem). Let $\mathcal{C}$ be an $[n, k, d]_q$ linear code, H be a parity check matrix of $\mathcal{C}$, $\mathbf{x} \in \mathbb{F}_q^n$ be a cover-data, and $\mathbf{m} \in \mathbb{F}_q^{n-k}$ be the message to be hidden in $\mathbf{x}$. The *syndrome coding problem* consists in finding $\mathbf{y} \in \mathbb{F}_q^n$ such that $\mathbf{y}H^t = \mathbf{m}$.

Problem 2 (Bounded syndrome coding problem). Let $\mathcal{C}$ be an $[n, k, d]_q$ linear code, H be a parity check matrix of $\mathcal{C}$, $\mathbf{x} \in \mathbb{F}_q^n$ be a cover-data, $\mathbf{m} \in \mathbb{F}_q^{n-k}$ be the message to be hidden in $\mathbf{x}$, and $T \in \mathbb{N}^*$ be an upper bound on the number of authorized modifications. The *bounded syndrome coding problem* consists in finding $\mathbf{y} \in \mathbb{F}_q^n$ such that $\mathbf{y}H^t = \mathbf{m}$, and $d(\mathbf{x}, \mathbf{y}) \leq T$.

Let us first focus on Problem 1, which leads to describing the stego-system in terms of syndrome computation:

$$\begin{aligned} \mathbf{y} = Emb(\mathbf{x}, \mathbf{m}) &= \mathbf{x} + D(\mathbf{m} - \mathbf{x}H^t), \\ Ext(\mathbf{y}) &= \mathbf{y}H^t, \end{aligned}$$

where D is the mapping associating to a syndrome $\mathbf{m}$, a vector whose syndrome is precisely equal to $\mathbf{m}$. The mapping D is thus directly linked to a decoding function $f_{\mathcal{C}}$ of $\mathcal{C}$ of arbitrary radius T_f, defined as $f_{\mathcal{C}} : \mathbb{F}_q^n \longrightarrow \mathcal{C} \cup \{?\}$, such that for all $\mathbf{y} \in \mathbb{F}_q^n$, either $f_{\mathcal{C}}(\mathbf{y}) = ?$, or $d(\mathbf{y}, f_{\mathcal{C}}(\mathbf{y})) \leq T_f$.

The Hamming distance between vectors $\mathbf{x}$ and $\mathbf{y}$ is then less than or equal to T_f. Since decoding general codes is NP-Hard [2], finding such a mapping D is not tractable if $\mathcal{C}$ does not belong to a family of codes we can efficiently decode. Moreover, to be sure that the Problem 2 always has a solution, it is necessary and sufficient that $f_{\mathcal{C}}$ can decode up to the covering radius of $\mathcal{C}$. This means that solving Problem 2 with $T = \rho$ is precisely equivalent to designing a stego-system which find solutions to both Eqs. (1) and (2) requirements for any $\mathbf{x}$ and $\mathbf{m}$. In this context,

perfect codes, for which the covering radius is precisely equal to the error-correcting capacity ($\rho = \lfloor \frac{d-1}{2} \rfloor$), are particularly relevant.

Unfortunately, using perfect codes leads to an embedding efficiency which is far from the bound given in Prop. 1 [4]. Hence non-perfect codes have been studied (see the Introduction), even if they can only tackle Problem 2 for some T much lower than ρ. This may enable to force the system to perform only a small number of modifications.

As discussed in the introduction, *Wet paper* codes were introduced to improve embedding undetectability through the management of locked, or *wet*, components [18].

Problem 3 (Bounded syndrome wet paper coding problem). Let $\mathcal{C}$ be an $[n, k, d]_q$ linear code, H be a parity check matrix of $\mathcal{C}$, $\mathbf{x} \in \mathbb{F}_q^n$, $\mathbf{m} \in \mathbb{F}_q^{n-k}$, $T \in \mathbb{N}^*$, and a set of locked, or wet, components $\mathcal{I} \subset \{1, \ldots, n\}$, $\ell = |\mathcal{I}|$. The *Bounded syndrome wet paper coding problem* consists in finding $\mathbf{y} \in \mathbb{F}_q^n$ such that $\mathbf{y}H^t = \mathbf{m}$, $d(\mathbf{x}, \mathbf{y}) \leq T$, and $\mathbf{x}_i = \mathbf{y}_i$ for all $i \in \mathcal{I}$.

Of course, solving Problem 3 is harder and even perfect codes may fail here. More precisely, to deal with locked components, we usually decompose the parity check matrix H of $\mathcal{C}$ in the following way [18,19]:

$$\begin{aligned} \mathbf{y}H^t &= \mathbf{m}, \\ \mathbf{y}_{|\bar{\mathcal{I}}} H^t_{|\bar{\mathcal{I}}} + \mathbf{y}_{|\mathcal{I}} H^t_{|\mathcal{I}} &= \mathbf{m}, \\ \mathbf{y}_{|\bar{\mathcal{I}}} H^t_{|\bar{\mathcal{I}}} &= \mathbf{m} - \mathbf{y}_{|\mathcal{I}} H^t_{|\mathcal{I}}, \end{aligned}$$

where $\bar{\mathcal{I}} = \{1, \ldots, n\} \setminus \mathcal{I}$. The previous equation can only be solved if $rank(H_{\bar{\mathcal{I}}}) = n - k$. Since the potential structure of H does not help to solve the previous problem, we could as well choose H to be also a random matrix, which provides the main advantage to maximize asymptotically the average embedding efficiency [22,19].

Hiding a long message requires to split it and to repeatedly use the basic scheme. Let P_H the success probability for embedding $(n-k)$ symbols, then the global success probability P for a long message of length $L(n-k)$ is P_H^L. This probability decreases exponentially with the message length.

In order to bypass this issue, previous works propose either to take another cover-medium, or to modify some locked components. In this paper, we still keep unmodified the locked components, thus maintaining the same level of undetectability. Moreover, we tackle the particular case where the sender does not have a lot of cover-media available, and needs a successful embedding, even if this leads to a smaller embedding efficiency.

In the original Wet Paper Setting of [18], the embedding efficiency is not dealt with. In that case, we have a much easier problem.

Problem 4 (Unbounded wet paper Syndrome coding problem). Let $\mathcal{C}$ be an $[n,k,d]_q$ linear code, H be a parity check matrix of $\mathcal{C}$, $\mathbf{x} \in \mathbb{F}_q^n$, $\mathbf{m} \in \mathbb{F}_q^{n-k}$, and a set of locked components $\mathcal{I} \subset \{1,\dots,n\}$, $\ell = |\mathcal{I}|$. The *Unbounded wet paper Syndrome coding problem* consists in finding $\mathbf{y} \in \mathbb{F}_q^n$ such that $\mathbf{y}H^t = \mathbf{m}$, and $\mathbf{x}_i = \mathbf{y}_i$, for all $i \in \mathcal{I}$.

In a random case setting, this problem can be discussed using a lower bound on random matrices, provided by [5].

Theorem 1. *Let M be a random $ncol \times nrow$ matrix defined over $\mathbb{F}_q$, such that $ncol \geq nrow$. We have:*

$$P\left(rank(M) = nrow\right) \geq \begin{cases} 0.288, & \text{if } ncol = nrow \text{ and } q = 2, \\ 1 - \frac{1}{q^{ncol-nrow}(q-1)}, & \text{otherwise.} \end{cases}$$

In a worst-case, or infallible, setting, the relevant parameter of the code is its *dual distance*.

Proposition 2. *Consider a q-ary wet channel on length n with at most ℓ wet positions, and that there exists a q-ary code C whose dual code $C^\perp$ has parameters $[n, k^\perp, d^\perp = \ell]_q$ with $k^\perp + d^\perp = n + 1 - g$. Then we can surely embed $n - \ell - g$ symbols using a parity check matrix of C.*

Proof. This can be derived from [26, Theorem 2.3].

This means that if the code is g far from the Singleton bound, then we loose g information symbols with respect to the maximum. In particular, if $n < q$, there exists a q-ary Reed-Solomon code with $g = 0$, and we can always embed $n - \ell$ symbols when there are ℓ wet symbols. Coding theory bounds tells us that the higher q, the smallest g can be achieved, eventually using Algebraic-Geometry codes [35].

3 Randomized (wet paper) syndrome coding

Since embedding a message has a non-zero probability to fail, we propose to relax the constraints in the following way:

Problem 5 (Randomized bounded syndrome coding problem for wet paper). Let $\mathcal{C}$ be an $[n,k,d]_q$ linear code, H be a parity check matrix of $\mathcal{C}$, r and T be two integers, $\mathbf{x} \in \mathbb{F}_q^n$, $\mathbf{m} \in \mathbb{F}_q^{n-k-r}$ be the message to embed, and

$\mathcal{I} \subset \{1, \ldots, n\}$ be the set of locked components, $\ell = |\mathcal{I}|$. Our *randomized syndrome coding problem for wet paper* consists in finding $\mathbf{y} \in \mathbb{F}_q^n$ and $\mathbf{R} \in \mathbb{F}_q^r$ such that (i) $\mathbf{y}H^t = (\mathbf{m}||\mathbf{R})$, and $||$ denotes the concatenation operator, (ii) $d(\mathbf{x}, \mathbf{y}) \leq T$, and (iii) $\mathbf{x}_i = \mathbf{y}_i$, for all $i \in \mathcal{I}$.

We thus randomize one fraction of the syndrome to increase the number of solutions. This gives a degree of freedom which may be large enough to solve the system. The traditional approach can then be applied to find $\mathbf{y}_{|\bar{\mathcal{I}}}$ and consequently $\mathbf{y}$. Using some random symbols in the syndrome was used in the signature scheme of Courtois, Finiasz and Sendrier [8]. While this reformulation allows to solve the bounded syndrome coding problem in the wet paper context without failure, we obviously lose some efficiency compared to the traditional approach.

We now estimate the loss in embedding efficiency for a given number of locked components. Let e denote the embedding efficiency of the traditional approach, and e' denote the efficiency of the randomized one. We obtain a relative loss of:

$$\frac{e - e'}{e} = \frac{r}{n-k},$$

while being assured that any $n - k - r$ message be embedded, as long as $r < n - k$.

Optimizing the parameter r is crucial, to ensure that our reformulated problem always has a solution, while preserving the best possible embedding efficiency. This is the goal on next Section.

4 Case of perfect linear codes

We discuss in this Section a sufficient condition on the size r of randomization, for our reformulated problem to always have a solution.

4.1 General Statement

The *syndrome function* associated with H, noted S_H, is defined by:

$$\begin{aligned} S_H : \mathbb{F}_q^n &\longrightarrow \mathbb{F}_q^{n-k} \\ \mathbf{x} &\longmapsto \mathbf{x}H^t. \end{aligned}$$

This function S_H is linear and surjective, and satisfies the following well-known properties. Let $\mathcal{B}(\mathbf{x}, T)$ denote the Hamming ball of radius T centered on $\mathbf{x}$.

Proposition 3. *Let $\mathcal{C}$ be an $[n,k,d]_q$-linear code, with covering radius ρ, H a parity check matrix of $\mathcal{C}$, and S_H the syndrome function associated with H. For all $\mathbf{x} \in \mathbb{F}_q^n$, the function S_H restricted to $\mathcal{B}(\mathbf{x}, \lfloor \frac{d-1}{2} \rfloor)$ is one-to-one, the function S_H restricted to $\mathcal{B}(\mathbf{x}, \rho)$ is surjective. When $\mathcal{C}$ is perfect, the syndrome function restricted to $\mathcal{B}(\mathbf{x}, \rho)$ is bijective.*

Now, we give a sufficient condition for upper-bounding r in Problem 5.

Proposition 4. *Given a $[n,k,d]$ perfect code with $\rho \frac{d-1}{2}$, if the inequality*

$$q^{n-k} + 1 \leq q^r + \sum_{i=0}^{\rho} (q-1)^i \binom{n-\ell}{i}, \tag{3}$$

is satisfied, then there exists a vector $\mathbf{y} \in \mathbb{F}_q^n$ and a random vector $\mathbf{R}$, which are solution of Problem 5. In this case, Problem 5 always has a solution $\mathbf{y}$.

Proof. Let N_1 —respectively N_2— be the number of different syndromes generated by the subset of $\mathbb{F}_q^n$ satisfying (i) of Problem 5 — respectively (ii) and (iii). If

$$N_1 + N_2 > q^{n-k}. \tag{4}$$

Then there exists $\mathbf{y}$ which fulfills conditions (i), (ii), and (iii). The number of different syndromes satisfying by the first constraint, for all $\mathbf{R}$, is q^r. Keeping in mind that ℓ components are locked and the syndrome function restricted to $\mathcal{B}(\mathbf{x}, \rho)$ is bijective, then

$$N_2 = \sum_{i=0}^{\rho} (q-1)^i \binom{n-\ell}{i}.$$

Combined with the sufficient condition (4) we obtain the result.

Next Section is devoted to the non trivial perfect codes: the Golay codes, and the (q-ary) the Hamming codes.

4.2 Golay codes

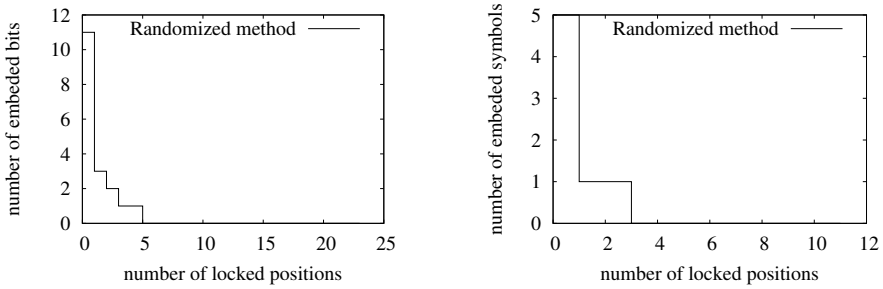
Binary Golay code We start by study the case of the binary $[23, 12, 7]_2$ Golay code, which is perfect. The inequality of the proposition 4 gives

$$r \geq \log_2 \left(1 + \frac{796}{3}\ell - \frac{23}{2}\ell^2 + \frac{1}{6}\ell^3 \right). \tag{5}$$

Ternary perfect Golay code The ternary Golay code has parameters $[11, 6, 5]_3$. Using the Proposition 4, we obtain:

$$r \geq \log_3 \left(1 + 44\ell - 2\ell^2\right). \tag{6}$$

Eqs 5 and 6 does not say much. We have plotted the results in Fig. 1, and we see that the number of available bits for embedding degrades very fast with the number of locked positions.



(a) Binary Golay code. (b) Ternary Golay code.

Fig. 1. Size of the random part for the two Golay codes. The number of remaining bits is plotted, in terms of the number of locked positions.

4.3 Hamming codes

We study the infinite family of Hamming codes. We find r, analyze the found parameters, and study its asymptotic behavior.

Computation of r Let $\mathcal{C}$ be a $[(q^p-1)/(q-1), n-p, 3]_q$ Hamming code over $\mathbb{F}_q$, for some p. Its covering radius is $\rho = 1$, and thus its embedding efficiency if p. We aim to minimize r, the length of the random vector $\mathbf{R}$. Since $q^{n-k} = q^p$, $(q^p - 1)/(q - 1) = n$, Proposition 4 gives:

$$r \geq \log_q \left(1 + (q-1)\ell\right). \tag{7}$$

Analysis of parameters In order to find an extreme case, it we maximize the number of locked components ℓ while still keeping $n - k - r \geq 1$.

A direct computation gives:

$$p - 1 = \log_q((q-1)\ell + 1),$$
$$\ell = \frac{q^{p-1} - 1}{q-1} \approx \frac{n}{q}.$$

Therefore, using Hamming codes, we can embed at least one information symbol if no more than a fraction of $\frac{1}{q}$ of the components are locked. This is of course best for $q = 2$. The minimum r which satisfies inequality (7) is $r = \lceil \log_q((q-1)\ell+1) \rceil$. In other words, for Hamming codes, the minimum number of randomized symbols needed to guarantee that the whole message can be embedded, is logarithmic in the number of locked components. Our randomized approach always solves successfully Problem 5 while traditional syndrome coding (including wet paper) exhibits a non-zero failure rate, when $\frac{\ell}{n} < \frac{1}{q}$.

Asymptotic behavior Now we evaluate the loss in embedding efficiency. Then, for a given ℓ, the relative loss of the embedding efficiency is given by:

$$\frac{\lceil \log_q((q-1)\ell + 1) \rceil}{p}.$$

To conclude this section, we propose to focus on the normalized loss in symbols for the family of Hamming codes. We assume that the rate of ℓ, the number of locked components to compare to n, the length of the cover-data stays constant, i.e. $\ell = \lambda n$, for a given $\lambda \in [0, \frac{1}{q}($. Then the asymptotic of relative loss is

$$\frac{\log_q((q-1)\ell + 1)}{p} \sim \frac{\log_q(n(q-1)\lambda)}{p} \sim 1 + \frac{\log_q \lambda}{p}.$$

This goes to 1 when p goes to infinity, i.e. all the symbols of syndrome are consumed by the randomization. It makes sense, since dealing with a given proportion λ of arbitrarily locked symbols in a long stego-data is much harder than dealing with several smaller stego-data with the same proportion λ of locked positions.

5 Using ZZW construction to embed dynamic parameters

In the approach given in previous Section, the sender and recipient have to fix in advance the value of r. Indeed the recipient has to know which

part of syndrome is random. This is not very compliant with the Wet Paper model, where the recipient does not know the quantity of wet bits. We propose in this Section a variant of ZZW's scheme [39], which enables to convey dynamically the value r, depending on the cover-data.

5.1 The scheme

We consider that we are treating n blocks of $2^p - 1$ bits, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, for instance displayed as in Figure 2. Each block $\mathbf{x}_i$ is a binary vector of length $2^p - 1$, set as column, and we let $\mathbf{v} = (v_1, \ldots, v_n)$ be the binary vector whose i-th coordinate v_i is the parity bit of column $\mathbf{x}_i$. We use the (virtual) vector $\mathbf{v}$ to convey extra information, while at the same time the $\mathbf{x}_i$ are using for syndrome coding.

Our scheme is threefold : syndrome coding on the $\mathbf{x}_i$'s using the parity check H_1 of a first Hamming code, with our randomized method, then (unbounded wet paper) syndrome embedding on the syndromes $\mathbf{s}_i$'s of the $\mathbf{x}_i$'s. This second syndrome embedding see the $\mathbf{s}_i$ as q-ary symbols, and the matrix in use is the parity check matrix H_q of a q-ary Reed-Solomon code. We call the n first embeddings the H_1-embeddings, and the second one the H_q-embedding. Finally, we use $\mathbf{v}$ to embed dynamic information: the number r of random bits, and f the number of failure in the H_1-embeddings. We call this last embedding the H_2-embedding, where H_2 is the parity check matrix of a second, much shorter, binary Hamming code.

We assume that r is bounded by design, say $r \leq r_{\max}$. We shall see, after a discussion on all the parameters, that this is one design parameter of the scheme, together with o, which the precision, in bits, for describing real numbers $\in]\frac{1}{2}, 1]$.

Embedding

Inspect. Each column $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is inspected, to find the number of dry bits in each. This enables to determine the size r of the randomized part, which shall be the same for all columns. This determines the columns $\mathbf{x}_i$'s where the H_1-embeddings are feasible. Let f be the number of $\mathbf{x}_i$'s where the H_1-embeddings fail.

Build the wet channel. For each of the $n - f$ columns $\mathbf{x}_i$'s where the H_1-embedding is possible, there is a syndrome s_i of p bits, where the last r bits are random, thus wet, and the $p - r$ first bits are dry. We consider these blocks of $p - r$ bits as a q-ary symbols, with $q = 2^{p-r}$. Thus we have a q-ary wet channel with $n - f$ dry q-ary symbols, and f wet q-ary symbols

Embed for the wet channel. Then, using a Reed-Solomon over the alphabet $\mathbb{F}_q$, we can embed $(n-f)$ q-ary symbols, using a $n \times (n-f)$ q-ary parity check matrix H_q of the code. Note that the number of rows of this matrix is dynamic since f is dynamic.

Embed dynamic data. We have to embed dynamic parameters r and f which are unknown to the recipient, using ZZW's virtual vector $\mathbf{v}$. For this binary channel, the dry bits v_i correspond to the columns $\mathbf{x}_i$ where the H_1-embedding has failed, and where there is at least one dry bit in $\mathbf{x}_i$. A second Hamming code is used with parity check H_2 for this embedding.

Recovery

H_2-extraction. First compute $\mathbf{v}$, and using the parity check matrix of the Hamming code H_2, extract r and f.

H_1-extraction Extract the syndromes of all the column $\mathbf{x}_i$'s using the parity check matrix H_1, and collect only the first $p-r$ bits in each column, to build q-ary symbols.

H_q-extraction Build the parity check matrix H_q of the q-ary $[n, f]_q$ Reed-Solomon code, with $q = 2^{p-r}$. Using this matrix, get the $(n-f)$ q-ary information symbols, which are the actual payload.

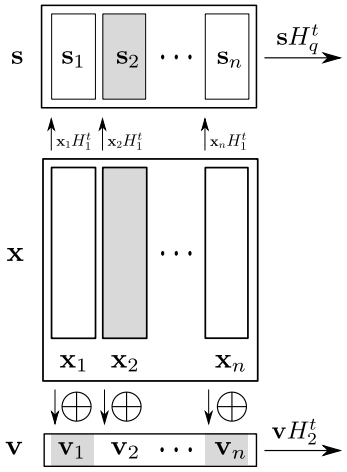


Fig. 2. A graphical view of our scheme inspired from ZZW. A syndrome $\mathbf{s}_i$ is considered wet for the H_q-embedding when the H_1-embedding is not feasible. Then the corresponding bit v_i in the vector $\mathbf{v}$ is dry for the H_2-embedding. Wet data is grey on the Figure.

5.2 Analysis

There are several constraints on the scheme.

First, for a Reed-Solomon code of length n to exist over the alphabet $\mathbb{F}_{2^{p-r}}$, we must have $n \leq 2^{p-r}$, for any r, i.e. $n \leq 2^{p-r_{\max}}$. We fix $n = 2^{p-r_{\max}} - 1$, and let us briefly denote $u = p - r_{\max}$.

Then the binary $[n = 2^u - 1, 2^u - u - 1]_2$ Hamming code, with parity check matrix H_2, is used for embedding in the vector $\mathbf{v}$, with f dry symbols. This a unbounded wet channel. From Proposition 2, we must have

$$f \geq 2^{u-1}, \tag{8}$$

which implies that some columns $\mathbf{x}_i$ may be artificially declared wet, for satisfying Eq. 8. Third, we also must have

$$u = \lceil \log r_{\max} \rceil + \lceil \log f_{\max} \rceil, \tag{9}$$

to be able to embed r and f. Since $f \leq 2^u - 1$, we have $\lceil \log f_{\max} \rceil = u$. Eq. 9 becomes $u = \lceil \log r_{\max} \rceil + u$, this is clearly not feasible. To remedy this, instead of embedding f, we embed its relative value $f_u = \frac{f}{2^u} \in [.5, 1]$, up to a fixed precision, say o bits, with o small. Then Eq. 9 is replaced by

$$u = \lceil \log r_{\max} \rceil + o, \tag{10}$$

$$p = r_{\max} + \lceil \log r_{\max} \rceil + o, \tag{11}$$

which is a condition easy to fulfill. It is also possible, by design, to use the all-one value of f_u as an out-of-range value to declare an embedding failure. The scheme is locally adaptive to the media: for instance, in a given image, r and f may take different values for different areas of the image.

In conclusion, the number of bits that we can embed using that scheme is bounded by $(n - f)(p - r) \leq 2^{u-1}(p - r)$, with dynamic r and f.

6 Conclusion

In this paper, we addressed the "worst-case" scenario, where the sender cannot accept embedding to fail, and does not want relax the management of locked components of his cover-data. As traditional (wet) syndrome coding may fail, and as the failure probability increases exponentially with the message length, we proposed here a different approach, which

never fails. Our solution is based on the randomization of a part of the syndrome, the other part still carrying symbols of the message to transmit. While our method suffers from a lost of embedding efficiency, we showed that this loss remains acceptable for perfect codes. Moreover, we showed how the size of the random part of the syndrome, which is dynamically estimated during embedding, may be transmitted to the recipient without any additional communication.

References

1. Anderson, R., Petitcolas, F.: On the limits of steganography. IEEE Journal on Selected Areas in Communications 16(4), 474 –481 (May 1998)
2. Berlekamp, E., McEliece, R., Van Tilborg, H.: On the inherent intractability of certain coding problems. IEEE Trans. on Information Theory 24(3), 384 – 386 (May 1978)
3. Bierbrauer, J.: On Crandall's problem. Personal communication (2001), `http://www.ws.binghamton.edu/fridrich/covcodes.pdf`
4. Bierbrauer, J., Fridrich, J.: Constructing good covering codes for applications in steganography. In: Shi, Y.Q. (ed.) Transactions on data hiding and multimedia security III. pp. 1–22. Springer Berlin / Heidelberg, Berlin, Heidelberg (2008)
5. Brent, R.P., Gao, S., Lauder, A.G.B.: Random Krylov spaces over finite fields. SIAM J. Discrete Math 16, 276–287 (2001)
6. Cachin, C.: An information-theoretic model for steganography. In: Information Hiding, 2nd International Workshop - IH 1998. Lecture Notes in Computer Science, vol. 1525, pp. 306–318. Springer-Verlag (1998)
7. Cachin, C.: An information-theoretic model for steganography. Information and Computation 192(1), 41–56 (2004)
8. Courtois, N., Finiasz, M., Sendrier, N.: How to achieve a McEliece-based digital signature scheme. In: Boyd, C. (ed.) Advances in Cryptology ASIACRYPT 2001, Lecture Notes in Computer Science, vol. 2248, pp. 157–174. Springer Berlin / Heidelberg (2001)
9. Crandall, R.: Some notes on steganography (1998), `http://os.inf.tu-dresden.de/~westfeld/crandall.pdf`, posted on the steganography mailing list.
10. Filler, T., Fridrich, J.: Wet ZZW construction for steganography. In: IEEE International Workshop on Information Forensics and Security - WIFS 2009. pp. 131–135 (2009)
11. Filler, T., Fridrich, J.: Minimizing additive distortion functions with non-binary embedding operation in steganography. In: IEEE International Workshop on Information Forensics and Security - WIFS 2010 (2010)
12. Filler, T., Judas, J., Fridrich, J.: Minimizing additive distortion in steganography unsing syndrome-trellis codes. IEEE Trans. on Information Forensics and Security (2011)
13. Filler, T., Judas, J., Fridrich, J.: Minimizing embedding impact in steganography using trellis-coded quantization. In: IS&T/SPIE International Symposium on Electronic Imaging 2010 - Media Forensics and Security II. Proceedings of the SPIE, vol. 7541. SPIE (2010)
14. Fontaine, C., Galand, F.: How can Reed-Solomon codes improve steganographic schemes. In: 9th Information Hiding – IH'07. Lecture Notes in Computer Science, vol. 4567, pp. 130–144. Springer-Verlag (2007)

15. Fontaine, C., Galand, F.: How Reed-Solomon codes can improve steganographic schemes. EURASIP J. Inf. Secur. 2009, 1–10 (2009)
16. Fridrich, J.: Asymptotic behavior of the ZZW embedding construction. IEEE Transactions on Information Forensics and Security 4(1), 151–153 (2009)
17. Fridrich, J., Filler, T.: Practical methods for minimizing embedding impact in steganography. In: IS&T/SPIE International Symposium on Electronic Imaging 2007 - Security, Steganography, and Watermarking of Multimedia Contents IX. Proceedings of the SPIE, vol. 6505. SPIE (2007)
18. Fridrich, J., Goljan, M., Lisonek, P., Soukal, D.: Writing on wet paper. IEEE Trans. on Signal Processing 53(10), 3923 – 3935 (October 2005)
19. Fridrich, J., Goljan, M., Soukal, D.: Wet paper codes with improved embedding efficiency. IEEE Trans. on Information Forensics and Security 1(1), 102 – 110 (March 2006)
20. Fridrich, J., Soukal, D.: Matrix embedding for large payloads. IEEE Trans. on Information Forensics and Security 1(3), 390 –395 (Sep 2006)
21. Fridrich, J.J., Goljan, M., Soukal, D.: Efficient wet paper codes. In: Information Hiding. pp. 204–218 (2005)
22. Galand, F., Kabatiansky, G.: Information hiding by coverings. In: Proc. ITW 2003. pp. 151–154 (2003)
23. Galand, F., Kabatiansky, G.: Coverings, centered codes, and combinatorial steganography. Problems of Information Transmission 45(3), 289–297 (2009)
24. Kodovský, J., Fridrich, J., Pevný, T.: Statistically undetectable jpeg steganography: Dead ends, challenges, and opportunities. In: Proc. of the ACM Multimedia and Security Workshop 2007. pp. 3–14. ACM (2007)
25. McLoughlin, A.: The complexity of computing the covering radius of a code. IEEE Trans. on Information Theory 30(6), 800–804 (1984)
26. Munuera, C., Barbier, M.: Wet paper codes and the dual distance in steganography (April 2011), `http://arxiv.org/abs/1104.1970`
27. Ould Medeni, M., Souidi, E.M.: A steganography schema and error-correcting codes. Journal of Theoretical and Applied Information Technology 18(1), 42–47 (2010)
28. Rifà, J., Ronquillo, L.: Product perfect Z2Z4-linear codes in steganography. In: International Symposium on Information Theory and its Applications - ISITA 2010 (2010)
29. Rifà-Pous, H., Rifà, J.: Product perfect codes and steganography. Digital Signal Processing 19(4), 764–769 (2009)
30. Sachnev, V., Kim, H., Zhang, R.: Less detectable jpeg steganography method based on heuristic optimization and BCH syndrom coding. In: ACM Multimedia & Security'09. pp. 131–139. ACM Press (2009)
31. Schönfeld, D., Winkler, A.: Embedding with syndrome coding based on BCH codes. In: Proceedings of the 8th workshop on Multimedia and security. pp. 214–223. ACM (2006)
32. Schönfeld, D., Winkler, A.: Reducing the complexity of syndrome coding for embedding. In: Proc. of the 10th International Worksop on Information Hiding. Lecture Notes in Computer Science, vol. 4567, pp. 145–158. Springer-Verlag (2007)
33. Simmons, G.: The prisoners' problem and the subliminal channel. In: Advances in Cryptology – CRYPTO'83. pp. 51–67. Plenum Press (1984)
34. Vardy, A.: The intractability of computing the minimum distance of a code. IEEE Trans. on Information Theory 43(6), 1757–1766 (1997)

35. Vladut, S., Nogin, D., Tsfasman, M.: Algebraic Geometric Codes: Basic Notions (Mathematical Surveys and Monographs). American Mathematical Society (September 2007)
36. Westfeld, A.: F5 - A steganographic algorithm. In: Moskowitz, I. (ed.) Information Hiding, Lecture Notes in Computer Science, vol. 2137, pp. 289–302. Springer Berlin / Heidelberg (2001)
37. Zhang, R., Sachnev, V., Kim, H.: Fast BCH syndrome coding for steganography. In: Katzenbeisser, S., Sadeghi, A.R. (eds.) Information Hiding. Lecture Notes in Computer Science, vol. 5806, pp. 48–58. Springer-Verlag (2009)
38. Zhang, W., Zhang, X., Wang, S.: Near-optimal codes for information embedding in gray-scale signals. IEEE Trans. on Information Theory 56(3), 1262–1270 (2010)
39. Zhang, W., Zhang, X., Wang, S.: Maximizing steganographic embedding efficiency by combining Hamming codes and wet paper codes. In: Proc. of the 10th International Worksop on Information Hiding. Lecture Notes in Computer Science, vol. 5284, pp. 60–71. Springer-Verlag (2008)

3976 IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 53, NO. 10, OCTOBER 2005

Watermarking Security: Theory and Practice

François Cayre, Caroline Fontaine, and Teddy Furon

***Abstract*—This paper proposes a theory of watermarking security based on a cryptanalysis point of view. The main idea is that information about the secret key leaks from the observations, for instance, watermarked pieces of content, available to the opponent. Tools from information theory (Shannon's mutual information and Fisher's information matrix) can measure this leakage of information. The security level is then defined as the number of observations the attacker needs to successfully estimate the secret key. This theory is applied to two common watermarking methods: the substitutive scheme and the spread spectrum-based techniques. Their security levels are calculated against three kinds of attack. The experimental work illustrates how Blind Source Separation (especially Independent Component Analysis) algorithms help the opponent exploiting this information leakage to disclose the secret carriers in the spread spectrum case. Simulations assess the security levels derived in the theoretical part of the paper.**

***Index Terms*—Blind source separation, equivocation, Fisher information matrix, security, watermarking.**

I. Introduction

DIGITAL watermarking studies have always been driven by the improvement of *robustness*. Most of articles of this field deal with this criterion, presenting increasingly more impressive experimental assessments. Some key events in this quest are the use of spread spectrum [1], the invention of resynchronization schemes [2], [3], the discovery of the side information channel [4], [5], and the formulation of the opponent actions as a game [6].

On the contrary, *security* has received little attention in the watermarking community. The first difficulty is that security and robustness are neighboring concepts, which are hardly perceived as different. The intentionality behind the attack is not enough to make a clear cut between these two concepts. An image compression is clearly an attack related to robustness, but it might happen intentionally, i.e., with the purpose of removing the watermark, or not. *Robust* watermarking is defined in [7] as a communication channel multiplexed into original content in a nonperceptible way and whose "*capacity degrades as a smooth function of the degradation of the marked content.*" We add that the degradation is due to a classical content processing (compression, lowpass filtering, noise addition, geometric attack …).

Manuscript received June 30, 2004; revised November 13, 2004. This work was supported in part by the French Government through the ACI Fabriano, the RNRT project SDMO, and by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The associate editor coordinating the review of this manuscript and approving it for publication was Guest Editor Dr. Stefan Katzenbeisser.

F. Cayre and T. Furon are with IRISA/TEMICS, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France (e-mail: cayre@irisa.fr; teddy.furon@irisa.fr).

C. Fontaine is with CNRS, Laboratoire d'Informatique Fondamentale de Lille, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'ascq Cedex, France (e-mail: caroline.fontaine@lifl.fr).

Digital Object Identifier 10.1109/TSP.2005.855418

The attacker has three known strategies to defeat watermark robustness: to remove enough watermark signal energy, to jam the hidden communication channel, or to desynchronize the watermarked content.

Kalker then defines watermarking *security* as "*the inability by unauthorized users to access* [i.e., to remove, to read, or to write the hidden message] *the communication channel*" established by a robust watermarking. Security deals with intentional attacks whose aims are not only the removal of the watermark signal, excluding those already encompassed in the robustness category since the watermarking technique is assumed to be robust.

Some seminal works have already warned the watermarking community that digital watermarking may not be a secure primitive (i.e., a tool providing information security) despite its robustness. However, they only deal with dedicated attacks relevant to particular applications. The deadlock attack concerns copyright protection and illustrates the impossibility to prevent somebody from watermarking content with his own technique and key (by embedding a watermark signal or by creating a fake original) [8]. This ruins the identification of the owner because two watermarking channels interfere in the same piece of content. The collusion attack (i.e., the mixing of several watermarked versions of the same content) is related to the fingerprinting application. Multiple problems in the field of copyright protection and authentication stems from the copy attack, where the attacker first copies a watermark and then pastes it in a different piece of content [9]. The oracle attack is a threat whenever the opponent has access to a watermarking detector, as in copy protection for consumer electronics devices [10]. The attacker first estimates the secret key, testing the detection process on different pieces of content [11]; this disclosure then helps him forging pirated content. Note that in this last case, the number of detection tries is of utmost importance.

Articles proposing a complete analysis of robust watermarking security are extremely rare. The authors are only aware of the pioneering work of [12], where two digital modulation schemes achieve perfect secrecy, and more recent works sketching a general framework for security analysis [13], [14]. The main idea is here to adapt Shannon's definition of cryptography security to watermarking. At the beginning of the game, the watermarker selects a watermarking technique and randomly picks up a private key. According to the Kerckhoff's principle, the opponent knows the selected algorithm but not the private key. Then, the watermarker starts producing some marked pieces of content. The opponent has access to some observations, and his aim is to estimate the private key. The main idea of Shannon's theory is that information about the private key might leak from the observations. Hence, the *a posteriori* uncertainty of the opponent decreases as he makes more and more observations. However, the above-mentioned

works have only translated the cryptanalysis methodology into watermarking terminology.

The goal of this paper is to offer a complete and workable theory of watermarking security. It completes the Barni *et al.* approach, assessing for the really first time security levels of substitution and, especially, spread spectrum-based watermarking methods. For this purpose, the first section summarizes the methodology and introduces the basic notation. Measurement of the information leakages are based on Shannon's mutual information for a substitutive watermarking method in Section III and on Fisher's information for a spread spectrum-based watermarking method in Section IV. This yields estimation of security levels for three types of attack. Yet, these information theory tools do not reveal any insight for practical hacking of spread spectrum-based watermarking. Section V tackles this algorithmic issue. Tools from the blind source separation (BSS) field appear to be extremely helpful for the attacker, especially Principal Component Analysis (PCA) and Independent Component Analysis (ICA).

II. Methodology

A. Notation

Let us first list some notational conventions used in this paper. Vectors are sets in bold font, matrices in calligraphic font, and sets in blackboard font. Data are written in lowercase letters and random variables in capitals. The length of the vectors considered in this paper is N_v, and $x(i)$ is the ith component of vector $\mathbf{x}$. The probability density function of random variable $\mathbf{X}$ (or its probability mass function if $\mathbf{X}$ is discrete) is denoted by $p_{\mathbf{X}}(.)$. Hidden messages have N_c bits, and secret keys are usually composed of N_c elements, e.g., several carriers: $\mathbf{u}_\ell$ is the ℓth carrier. Finally, N_o vectors are considered: $\mathbf{x}^{N_o}$ represents this collection of vectors, and $\mathbf{x}_j$ is the vector $\mathbf{x}$ associated to the jth observation.

B. Cryptanalytic Approach

The methodology presented in this section is clearly inspired by the cryptanalysis. It has already been presented in [14] and is based on three key articles: Kerckhoffs [15], Shannon [16], and Diffie-Hellman [17]. We first briefly present these concepts before formalizing them in the following subsections.

Kerckhoff's Principle: It was stated in 1883 that keeping an encryption algorithm secret for years is not realistic, and this principle is now used in any cryptographic study. In watermarking, the situation is similar, and it is assumed that the opponent knows the watermarking algorithm. Hence, for a given design and implementation of an algorithm, the security stems from the secrecy of the key. The designer's challenge is the following: "Am I sure that an opponent will not exploit some weaknesses of the algorithm to disclose the secret key?". Watermarking processes are often split into three functions. The first one extracts some features from content [issued by a classical transform, such as the discrete cosine transform (DCT), wavelet, fast Fourier transform (FFT), Fourier Mellin, ...], which are stored in a so-called extracted vector. The second one mixes the extracted vector with the secret watermark signal, giving a watermarked vector. Then, an insertion function reverses the extraction process to come back in the original world, putting out the watermarked document. Fig. 1 illustrates the embedding process. The detection follows an analogous process, as sketched in Fig. 2. According to the Kerckhoff's principle, the opponent knows all the involved functions. He thus observes the watermarked vectors from contents to which he has access because the extraction function has no secret parameter.

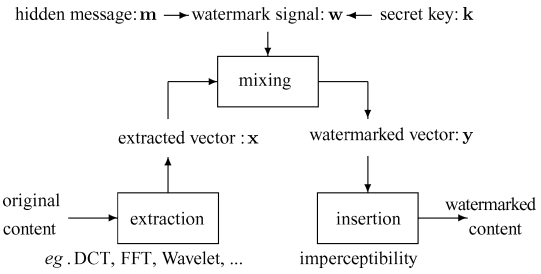


Fig. 1. Global point of view of the embedding process.

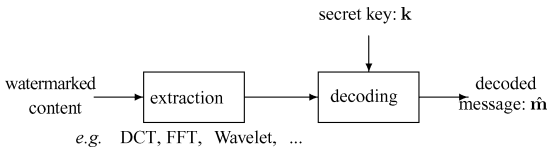


Fig. 2. Global point of view of the detection process.

Shannon's Approach: The methodology that Shannon exposed for studying the security of encryption schemes is here transposed to watermarking. The embedder has randomly picked up a secret key and used it to watermark several pieces of content. The opponent observes these pieces of watermarked content, which are all related to the same secret key but hiding different messages. The watermarking technique is *perfectly secure* if and only if no information about the secret key leaks from the observations. If it is not the case, the *security level* is defined as the number of observations that are needed to disclose the secret key. The bigger the information leakage is, the smaller the security level of the watermarking scheme will be.

Diffie-Hellman's Terminology: According to the context of the attack, the opponent may have access to several kinds of data. The opponent has at least access to watermarked content, but in some cases, he might also observe the hidden messages (for instance, the name of the author in copyright protection or the status of a movie in copy protection) or to the original data (for instance, imagine DVD movies are watermarked for copy protection; the original version of old movies were not protected). This implies that a security level is assessed for a given context. In this paper, we study the following:

- the Watermarked Only Attack (WOA), in which the opponent only has access to N_o watermarked vectors $\mathbf{y}^{N_o}$;
- the Known Message Attack (KMA), in which the opponent only has access to N_o watermarked vectors and the associate messages $(\mathbf{y}, \mathbf{m})^{N_o}$;
- the Known Original Attack (KOA), in which the opponent only has access to N_o watermarked vectors and the corresponding original ones $(\mathbf{y}, \mathbf{x})^{N_o}$.

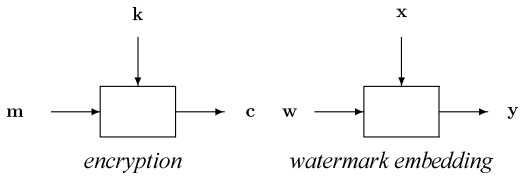


Fig. 3. Analogy with cryptography: plaintext $\mathbf{m} \rightarrow$ watermark $\mathbf{w}$, key $\mathbf{k} \rightarrow$ original $\mathbf{x}$, ciphertext $\mathbf{c} \rightarrow$ watermarked content $\mathbf{y}$.

The reader might be surprised that the KOA context deserves any attention. Seemingly, there is no need to attack watermarked content when one has the original version. The pirate does not hack these pieces of content, but his goal is to gain information about the secret key in order to, later on, hack different pieces of content watermarked with the same key.

C. Perfect Covering

Although cryptographic encryption and watermarking are two different security primitives, they might look like the same at first sight. Fig. 3 illustrates this analogy investigated in this subsection.

Shannon defined *perfect secrecy* of a crypto-system by the inability of opponents to refine the probability distribution of plaintexts $\mathbf{m}$ by observing related cipher texts that are all encrypted by key $\mathbf{k}$. We adapt this definition to watermarking, stating that the most important thing to be hidden is the watermark signal and not the original content. The equivalent of the plaintext is, here, the watermark signal.

Definition 1: A watermark embedding makes a *perfect covering* if

$$p_{\mathbf{W}}(\mathbf{w}) = p_{\mathbf{W}}(\mathbf{w}|\mathbf{y}), \qquad \text{for any } (\mathbf{y}, \mathbf{w}). \tag{1}$$

This means that in a perfect covering scheme, the observations of only watermarked pieces of contents will never reveal any information on the watermark signal: $I(\mathbf{Y};\mathbf{W}) = 0$. Whenever $\mathbf{K} \rightarrow \mathbf{W} \rightarrow \mathbf{Y}$ is a Markov chain, $I(\mathbf{Y};\mathbf{W}) \geq I(\mathbf{Y};\mathbf{K})$ holds. Consequently, perfect covering implies perfect secrecy.

Shannon easily found a necessary condition to get perfect secrecy by using his information theory tools: $H(\mathbf{M}) \leq H(\mathbf{K})$, where $H(.)$ denotes the entropy, that is, $H(\mathbf{M}) = -\sum_{\mathbf{m}} p(\mathbf{m}) \log p(\mathbf{m})$. Yet, the same proof yields the following necessary condition to get perfect covering: $H(\mathbf{W}) \leq H(\mathbf{X})$. This deeply reveals the difference between cryptography and watermarking. As suggested by the greek word $\kappa\rho\upsilon\pi\tau\omega$ (meaning "I hide"), the role of the secret key is, in encryption, to hide the meaning of the plaintext. Hence, its entropy should be greater or equal to the one of the plaintext, whereas steganography ($\sigma\tau\epsilon\gamma\alpha\nu\omega$ means "I cover") hides the watermark covered by the host signal.

D. Measure of Information Leakages and Physical Interpretation

If a watermarking scheme does not provide perfect secrecy, then one would like to measure the information leakage on the secret key. For this purpose, this subsection presents several tools from information theory, which will later be useful to analyze classical watermarking schemes.

1) Shannon's Measure: In the case where the secret key $\mathbf{K}$ is a discrete variable, and more usually a binary word, the entropy $H(\mathbf{K})$ measures the uncertainty of the opponent on the true value of $\mathbf{k}$. When he makes some observations[1] $\mathbf{O}^{N_o}$, his uncertainty is now evaluated through a conditional entropy, which Shannon named *equivocation*: $H(\mathbf{K}|\mathbf{O}^{N_o}) = H(\mathbf{K}) - I(\mathbf{K};\mathbf{O}^{N_o})$. The information leakage is measured by the mutual information between the observations and the secret key. The bigger the information leakage, the smaller the uncertainty of the opponent. Equivocation is a nonincreasing function with N_o. It goes from $H(\mathbf{K})$, ideally down to 0. When it becomes null, this means that the opponent has enough observations to uniquely determine the secret key. Shannon defined the *unicity distance* as the first value of N_o for which the equivocation becomes null, meaning that the set of all possible keys is now reduced to only one element. This is a way to measure the security level $N_o{}^{\star}$ of a primitive.

Unfortunately, these tools are not suitable for any watermarking scheme. It is well known that entropy (or conditional entropy) of a continuous random variable does not measure a quantity of information. Mutual information $I(\mathbf{K};\mathbf{O}^{N_o})$ is always pertinent as a measure of information leakages, but the physical interpretation of the equivocation as the remaining uncertainty does not hold when the secret key is regarded as a continuous random variable, as in Section IV. For instance, the equivocation can take positive or nonpositive values, ruining the concept of unicity distance.

2) Fisher's Measure: This is the reason why another information measurement is proposed. In statistics, Fisher was one of the first to introduce the measure of the amount of information supplied by the observations about an unknown parameter to be estimated. Suppose observation $\mathbf{O}$ is a random variable with a probability distribution function depending on a parameter vector $\boldsymbol{\theta}$. The *Fisher Information Matrix* (FIM) concerning $\boldsymbol{\theta}$ is defined as

$$\text{FIM}(\boldsymbol{\theta}) = E\boldsymbol{\psi}\boldsymbol{\psi}^T \quad \text{with} \quad \boldsymbol{\psi} = \nabla_{\boldsymbol{\theta}} \log p_{\mathbf{O}}(\mathbf{o};\boldsymbol{\theta}) \tag{2}$$

where E is the mathematical expectation operator, and $\nabla_{\boldsymbol{\theta}}$ is the gradient vector operator defined by $\nabla_{\boldsymbol{\theta}} = (\partial/\partial\theta[1], \ldots, \partial/\partial\theta[N_\theta])^T$. The Cramér–Rao theorem gives a lower bound of the covariance matrix of an unbiaised estimator of parameter vector $\boldsymbol{\theta}$ whenever the FIM is invertible:

$$\mathcal{R}_{\hat{\boldsymbol{\theta}}} \geq \text{FIM}(\boldsymbol{\theta})^{-1} \tag{3}$$

in the sense of non-negative definiteness of the difference matrix. In our framework, the parameter vector can be the watermark signal or the secret key. Equation (3) provides us a physical interpretation: The bigger the information leakage, the more accurate the estimation of the secret parameter.

The FIM is also an additive measure of the information, provided the observations are statistically independent. Suppose that the watermark signal has been added in N_o pieces of content whose extracted vectors are independent and identically distributed (i.i.d.) as $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathcal{R}_{\mathbf{X}})$. The observations are N_o watermarked signals. Then, $\log p_{\mathbf{O}}(\mathbf{o};\mathbf{w}) = -1/2 \sum_{j=1}^{N_o} (\mathbf{y}_j - \mathbf{w}) \mathcal{R}_{\mathbf{X}}^{-1} (\mathbf{y}_j - \mathbf{w})^T + const.$ Calculation readily gives $\text{FIM}(\mathbf{w}) =$

[1] e.g., observations can be "cipher texts" or "pairs of plain/cipher texts".

$N_o\mathcal{R}_{\mathbf{X}}^{-1}$. This models applications that detect the presence of (and not decode) watermarks as well as template signals that resynchronize content transformed by a geometric attack.

The mean square error $E\{||\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}||^2\}$ is the trace of $\mathcal{R}_{\hat{\boldsymbol{\theta}}}$, and thus, its lower bound decreases in ${N_o}^{-1}$. However, the rate $N_o^\star = N_o \text{tr}(\text{FIM}(\boldsymbol{\theta})^{-1})$ depends on the statistical model and, consequently, the kind of observations (see Section IV). This means that the estimation is significantly more accurate when the number of independent observations increases of an order of $N_o^\star$. The bigger $N_o^\star$ is, the more difficult the disclosure of the secret key will be. This notion is close to the unicity distance of the above subsection. This is the reason why we use the same notation $N_o^\star$ (although absolutely not defined in the same way).

III. Security Analysis of the Substitutive Method

A. Mathematical Model

In such a scheme, a binary vector $\mathbf{x} = (x(1)\dots x(N_v))^T$ is extracted from the content. For instance, in the famous Burgett, Koch, and Zao technique [18], N_v pairs of DCT coefficients of an image are compared in absolute value. The message to be hidden is a binary vector $\mathbf{m} = (m(1)\dots m(N_c))^T$. The secret key is a list of N_c integers $\mathbf{k} = [k(1),\dots,k(N_c)]$ with $1 \le k(\ell) \le N_v$ and $k(\ell) \neq k(\ell')$ if $\ell \neq \ell'$. The embedding process copies $\mathbf{x}$ in $\mathbf{y}$ and then substitutes the $k(\ell)$th bit of $\mathbf{y}$ by the ℓth bit of the message to be hidden: $y(k(\ell)) = m(\ell)$. The inverse extraction function maps back the watermarked vector $\mathbf{y}$ into the content. The decoding simply reads the bits whose indices are given by the secret key.

Example 1: $N_v = 8$ and $N_c = 4$:

$$\begin{aligned} \mathbf{m} &= (1101) & \mathbf{k} &= [2,8,5,3] \\ \mathbf{x} &= (01001011) & \mathbf{y} &= (0\mathbf{1}10\mathbf{0}01\mathbf{1}). \end{aligned}$$

The uncertainty of the opponent is given by the entropy of the secret key that the embedder has randomly selected among $N_v!/(N_v-N_c)!$ possible keys. Thus

$$H(\mathbf{K}) = \log_2 \frac{N_v!}{(N_v-N_c)!}. \tag{4}$$

B. Perfect Covering

Theorem 1: As defined above, a substitutive watermarking scheme provides perfect covering.

Proof: We can model the substitutive watermarking as follows: Let $\mathbf{x}$ be a binary N_v-length random vector, whose probability mass function is uniform and equal to 2^{-N_v}, and let $\mathbf{w}$ be a binary N_v-length vector whose bits equal to 1 indicates the bits to be flipped. Hence, we have $\mathbf{y} = \mathbf{x} \oplus \mathbf{w}$, giving

$$\begin{aligned} p_{\mathbf{Y}}(\mathbf{y}) &= \sum_{\mathbf{w}\in\mathbb{W}} p_{\mathbf{Y}}(\mathbf{y}|\mathbf{w})p_{\mathbf{w}}(\mathbf{w}) = \sum_{\mathbf{w}\in\mathbb{W}} p_{\mathbf{X}}(\mathbf{y}\oplus\mathbf{w})p_{\mathbf{w}}(\mathbf{w}) \\ &= 2^{-N_v}\sum_{\mathbf{w}\in\mathbb{W}} p_{\mathbf{w}}(\mathbf{w}) = 2^{-N_v} \\ p_{\mathbf{Y}}(\mathbf{y}|\mathbf{w}) &= p_{\mathbf{X}}(\mathbf{y}\oplus\mathbf{w}) = 2^{-N_v}. \end{aligned}$$

The Bayes rule $p_{\mathbf{Y}}(\mathbf{y}|\mathbf{w})p_{\mathbf{W}}(\mathbf{w}) = p_{\mathbf{W}}(\mathbf{w}|\mathbf{y})p_{\mathbf{Y}}(\mathbf{y})$ then gives $p_{\mathbf{W}}(\mathbf{w}) = p_{\mathbf{W}}(\mathbf{w}|\mathbf{y})$.

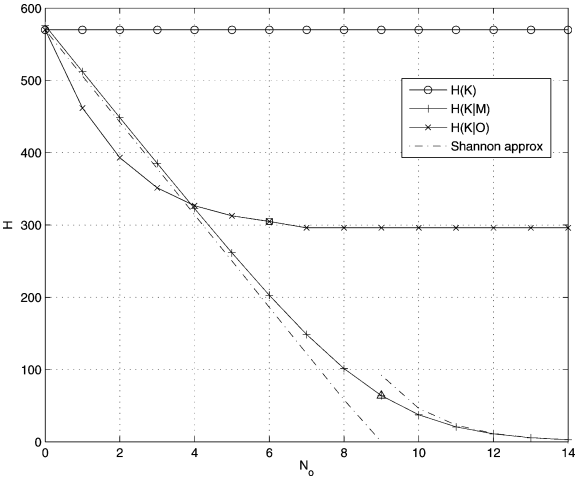


Fig. 4. Substitutive watermarking: Equivocations for WOA, KMA, and KOA against the number of observations. $N_c = 64$, $N_v = 512$. The triangle and the square, respectively, mark the security levels for the KMA and KOA.

C. Watermarked Only Attack

When the substitutive method provides perfect covering, it is then very easy to show that $I(\mathbf{Y};\mathbf{W}) = 0$, which implies that $I(\mathbf{Y};\mathbf{K}) = 0$. There is no information leakage, and the equivocation is equal to $H(\mathbf{K})$ whatever the number of observations. In a way, one can say that security level $N_o^\star = +\infty$.

D. Known Message Attack

If the opponent observes only one watermarked content $\mathbf{y}_1$ and its hidden message $\mathbf{m}_1$, the indices i such that $y_1(i) = m_1(\ell)$ are possible values of $k(\ell)$. Denote $\mathbb{S}_1(\ell)$ this set. As $P(y_1(i) = m_1(\ell)|i \neq k(\ell)) = 1/2$, there are, in expectation, $1 + (N_v-1)/2$ elements in this set.

Now, assume that the opponent observes several contents $\mathbf{y}^{N_o}$ and their hidden messages $\mathbf{m}^{N_o}$. Set $\mathbb{S}_{N_o}(\ell)$ is now defined by $\mathbb{S}_{N_o}(\ell) = \{i : y_j(i) = m_j(\ell)\ \forall j,\ 1 \le j \le N_o\}$. The probability that $y_j(i) = m_j(\ell)\,\forall j$ knowing that $i \neq k(\ell)$ is $1/2^{N_o}$. Thus, in expectation, $|\mathbb{S}_{N_o}| = 1+(N_v-1)/2^{N_o}$, and the equivocation about $k(\ell)$ is equal to $\log_2(1+2^{-N_o}(N_v-1))$. However, there might be some overlapping between the N_c sets $\mathbb{S}_{N_o}(\ell)$, and the total equivocation is smaller than the sum of the equivocations about $k(\ell)$. As the calculus is quite complex, we stay with this approximation:

$$H(\mathbf{K}|(\mathbf{Y},\mathbf{M})^{N_o}) \lesssim N_c \log_2(1+2^{-N_o}(N_v-1)). \tag{5}$$

Shannon approximated this equivocation by $N_c(\log_2(N_v-1)-N_o)$ when $N_o \ll \log_2(N_v-1)$ and by $2^{-N_o}N_c(N_v-1)/\log(2)$ when $N_o \gg \log_2(N_v-1)$ (see Fig. 4). He also approximated the unicity distance by $N_o^\star = \log_2 N_v$ [16, Sec. 14].

E. Known Original Attack

If the opponent observes only one watermarked content $\mathbf{y}_1$ and its original version $\mathbf{x}_1$, the indices i such that $x_1(i) \neq y_1(i)$ are possible values for the key samples. There are in expectation $N_c/2$ of such indices, as $p(x_1(k(\ell)) = m_1(\ell)) = 1/2$. When the opponent observes j pairs, the set $\mathbb{S}_j = \{\ell : \exists j', 1 \le$

$j' \leq j, x_{j'}(\ell) \neq y_{j'}(\ell)\}$ grows up. However, the event that an index revealed by a new pair was already known happens with a probability $|\mathbb{S}_{j-1}|/N_c$. This leads to the following series:

$$|\mathbb{S}_j| = |\mathbb{S}_{j-1}| + \frac{N_c\left(1 - \frac{|\mathbb{S}_{j-1}|}{N_c}\right)}{2} = N_c(1 - 2^{-j}). \quad (6)$$

Yet, it is not possible to assign a key sample to one of these indices. The equivocation is then the sum of two terms: One is due to the $N_c - |\mathbb{S}_{N_o}|$ undisclosed indices to be picked up randomly among the remaining candidates, and the second one is due to the $N_c!$ possible permutations of the chosen indices

$$H(\mathbf{K}|(\mathbf{Y},\mathbf{X})^{N_o}) = \log_2\left(\frac{(N_v - \lceil|\mathbb{S}_{N_o}|\rceil)!}{(N_v - N_c)!(N_c - \lceil|\mathbb{S}_{N_o}|\rceil)!}\right) + \log_2(N_c!). \quad (7)$$

The security level (in the unicity distance sense) is not defined as the equivocation is always greater than zero. This is due to the term $\log_2(N_c!)$ reflecting the ambiguity in the order of the estimated key samples. We preferably consider that within a number of observations greater than $N_o^\star = \log_2 N_c$, the opponent learns all the indices store in the secret key. This information is helpful for watermark jamming. He can also notice if two hidden messages are the same. Yet, the ambiguity prevents him reading the hidden messages (he cannot put the hidden bits in the right order) and writing hidden messages.

Fig. 4 gives a good synthesis of the results. In the WOA case, the opponent cannot get any information on the key and then cannot do anything. In the KMA case, he is able to completely disclose the key, and then, he will be able to read, erase, write, or modify hidden messages. In the KOA case, he is able to recover the components of the key but up to a permutation, and then, he will be able to erase the hidden message but not read or write a proper one.

IV. Security Analysis of Spread Spectrum-Based Techniques

Spread spectrum is a military communication scheme invented during World War II [19]. It was designed to be good at combatting interference due to jamming, hiding a signal by transmitting it at low power, and achieving secrecy. These properties make spread spectrum very popular in present-day digital watermarking. Theoretical studies [6] and practical implementations [20] focus on the optimization of operational capacity-robustness functions for a given embedding distortion.

A. Mathematical Model

Denote by $\mathbf{x}$ a vector of N_v samples extracted from original content. The embedding is the addition of the watermark signal that is the modulation of N_c private carriers $\mathbf{u}_\ell$:

$$\mathbf{w} = \frac{\gamma}{\sqrt{N_c}} \sum_{\ell=1}^{N_c} a(\ell)\mathbf{u}_\ell \quad (8)$$

where $\gamma > 0$ is a small gain fixing the embedding strength, and $\|\mathbf{u}_\ell\| = 1$, $1 \leq \ell \leq N_c$. The Watermark-to-Content power Ratio (WCR) equals $\gamma^2\sigma_a^2/\sigma_x^2$ (or $10\log_{10}(\gamma^2\sigma_a^2/\sigma_x^2)$ if expressed in decibels). An inverse extraction function puts back vector $\mathbf{y} = \mathbf{x} + \mathbf{w}$ into the media to produce the watermarked content.

Symbol vector $\mathbf{a}$ represents the message to be hidden/transmitted through content. In the case of a Direct Sequence Spread Spectrum (DSSS), the modulation is a simple binary phase shift keying (BPSK): $a(\ell) = (-1)^{m(\ell)}$, $1 \leq \ell \leq N_c$, and $\sigma_a^2 = 1$. Yet, the scope of this model is far broader than the sole case of DSSS. Spread spectrum is a very common process used to increase the signal-to-noise ratio (SNR) by projecting signals on a smaller subspace of dimension $N_c < N_v$. This also covers some side-informed watermarking techniques (sometimes called spread transform) [5], [21]–[23]. Symbols $a(\ell)$ are then continuous real values (see Section V-D).

For security reasons, the carriers are private and issued by a pseudo-random generator fed by a seed. Many people think the secret key is the seed. This is not false as the disclosure of the seed obviously gives the carriers and allows watermarking channel access. However, the knowledge of the carriers is sufficient, and the pirate has no interest in getting back to the seed. Hence, in this paper, the secret key, which is defined as the object the opponent is keen on revealing, is constituted by the carriers.

In the sequel, the security analysis considers several watermarked vectors $\mathbf{y}_j$, $1 \leq j \leq N_o$, with different embedded messages $\mathbf{a}_j = (a_j(1)\dots a_j(N_c))^T$ being linearly mixed by the $N_v \times N_c$ matrix $\mathcal{U} = (\mathbf{u}_1 \dots \mathbf{u}_{N_c})$. To cancel intersymbol interferences at the decoding side, carriers are two-by-two orthogonal vectors $\mathcal{U}^T\mathcal{U} = \mathcal{I}_{N_c}$, where $\mathcal{I}_N$ is the $N \times N$ identity matrix. Index i denotes the ith samples of a signal, whereas j indices denote the different signals. Thus, there are N_o watermarked vectors given by

$$\mathbf{y}_j = \mathbf{x}_j + \frac{\gamma}{\sqrt{N_c}}\mathcal{U}\mathbf{a}_j \quad (9)$$

or, equivalently, concatenating N_o vectors $\mathbf{x}_j$ (resp. $\mathbf{y}_j$ or $\mathbf{a}_j$) column-wise in the $N_v \times N_o$ matrix $\mathcal{X}$ (resp. $\mathcal{Y}$ or the $N_c \times N_o$ matrix $\mathcal{A}$):

$$\mathcal{Y} = \mathcal{X} + \frac{\gamma}{\sqrt{N_c}}\mathcal{U}\mathcal{A}. \quad (10)$$

B. Perfect Covering

Assume that $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathcal{R}_\mathbf{X})$ and that $\mathbf{w}$ is picked up randomly among sequences distributed as $\mathcal{N}(\mathbf{0}, \mathcal{R}_\mathbf{W})$. Then, $p_\mathbf{Y} = \mathcal{N}(\mathbf{0}, \mathcal{R}_\mathbf{X} + \mathcal{R}_\mathbf{W})$ and $p_{\mathbf{Y}|\mathbf{W}=\mathbf{w}} = \mathcal{N}(\mathbf{w}, \mathcal{R}_\mathbf{X})$. The Bayes rule shows that spread spectrum-based watermarking does not provide perfect covering. Even if the attacker has only access to watermarked pieces of content, some information about the watermark signal is leaking from these observations. The following subsections investigate whether the opponent can, thanks to this leakage on the watermark signal, gain some knowledge about the secret carriers.

C. Known Message Attack

In this subsection, the opponent has access to (watermarked signals/hidden messages) pairs. Moreover, only the DSSS technique (i.e., a BPSK modulation) is considered. Our attack may not work with side information embedding because the opponent still ignores symbols $\mathbf{a}$, as they also depend on the original

signal. Formally, the observations considered in this subsection are $(\mathbf{y},\mathbf{a})^{N_o}$.

Assume, for simplicity reasons, that each occurrence of random vector $\mathbf{X}$ is independently drawn from $\mathcal{N}(\mathbf{0},\sigma_x^2\mathcal{I}_{N_v})$. The following theoretical derivations (as well as the algorithm used in experiments in Section V) can be adapted to colored original signals and even nonstationary original signals [24]. Another motivation is that according to the Power Spectrum Constraint [25], watermark signals usually adopt the statistical structure of host signals in order to increase their robustness, i.e., $\mathcal{R}_{\mathbf{W}} = \gamma^2\mathcal{R}_{\mathbf{X}}$. Hence, the Karhunen–Loève transform simultaneously whitens both signals.

The likelihood is the probability of observing the data $\mathbf{y}^{N_o}$, while knowing the model

$$L(\mathbf{y}^{N_o}) = \frac{1}{(\sqrt{2\pi}\sigma_x)^{N_o N_v}} e^{\left(-(1/2\sigma_x^2)\sum_{j=1}^{N_o}\|\mathbf{y}_j-(\gamma/\sqrt{N_c})\mathcal{U}\mathbf{a}_j\|^2\right)} \tag{11}$$

and the log-likelihood is $\log L = K - (1/2\sigma_x^2)\sum_{j=1}^{N_o}\|\mathbf{y}_j - (\gamma/\sqrt{N_c})\mathcal{U}\mathbf{a}_j)\|^2$. The opponent wants to estimate the private carriers $\mathbf{u}^{N_c}$. Therefore, the derivative implied in the FIM is $\boldsymbol{\psi} = \partial\log L/\partial(\mathbf{u}_1^T\dots\mathbf{u}_{N_c}^T)^T$ with

$$\frac{\partial\log L}{\partial\mathbf{u}_\ell} = \frac{\gamma}{\sigma_x^2\sqrt{N_c}}\sum_{j=1}^{N_o} a_j(\ell)\mathbf{x}_j. \tag{12}$$

The expectation of the products gives the following $N_v\times N_v$ sub-blocks:

$$E\left(\frac{\partial\log L}{\partial\mathbf{u}_\ell}\right)\left(\frac{\partial\log L}{\partial\mathbf{u}_k}\right)^T = \frac{\gamma^2}{N_c\sigma_x^2}(\mathcal{F}_{uu})_{\ell,k} = \frac{\gamma^2}{N_c\sigma_x^2}\sum_{j=1}^{N_o} a_j(\ell)a_j(k)\mathcal{I}_{N_v}.$$

The FIM is then the following block matrix:

$$\text{FIM} = \frac{\gamma^2}{N_c\sigma_x^2}\begin{bmatrix}(\mathcal{F}_{uu})_{1,1} & \dots & (\mathcal{F}_{uu})_{1,N_c}\\ \vdots & & \vdots\\ (\mathcal{F}_{uu})_{N_c,1} & \dots & (\mathcal{F}_{uu})_{N_c,N_c}\end{bmatrix} = \frac{\gamma^2}{N_c\sigma_x^2}\mathcal{F}_{uu}\xrightarrow{N_o\to+\infty} N_o\frac{\gamma^2\sigma_a^2}{N_c\sigma_x^2}\mathcal{I}_{N_vN_c}. \tag{13}$$

With a BPSK modulation, $\sigma_a = 1$. The information leakage is linear with the number of observations, thanks to the assumption of independence, and the rate is given by the Watermark-to-Content power Ratio per carrier $\gamma^2/N_c\sigma_x^2$. The security level of spread spectrum-based watermarking techniques against KMA is $N_o{}^\star = N_c\sigma_x^2/\gamma^2$ of (watermarked signals/hidden messages) pairs.

D. Known Original Attack

The opponent observes $(\mathbf{y},\mathbf{x})^{N_o}$. The vector difference of each observation j gives the source signals $\mathbf{a}_j$ being linearly mixed by the $N_v\times N_c$ matrix $\mathcal{U}$:

$$\mathbf{d}_j = \mathbf{y}_j - \mathbf{x}_j = \frac{\gamma}{\sqrt{N_c}}\mathcal{U}\mathbf{a}_j. \tag{14}$$

Assume that $N_o \geq N_c$ and that there are at least N_c linearly independent messages. The difference matrix $\mathcal{D} = \mathcal{Y} - \mathcal{X} \propto \mathcal{U}\mathcal{A}$ is then full rank, and Span $(\mathcal{D})$ = Span $(\mathcal{U})$. The observation of difference vectors discloses the secret subspace Span $(\mathcal{U})$, provided symbol matrix $\mathcal{A}$ is full rank. However, this does not reveal the private carriers. Denote by $\mathcal{E}$ a matrix whose columns constitute an orthonormal basis of the subspace Span $(\mathcal{D})$. We have $\mathcal{E} = \mathcal{U}\mathcal{P}^T$, where $\mathcal{P}$ is a unitary $N_c\times N_c$ matrix. *A priori*, there is no reason for which $\mathcal{P} = \mathcal{I}_{N_c}$. Hence, decoding the symbols with matrix $\mathcal{E}$ gives the following mixture: $\mathbf{v} = \sqrt{N_c}\mathcal{E}^T\mathbf{d}/\gamma = \mathcal{P}\mathbf{a}$. This is a BSS problem with a square mixing matrix. Comon proved that it is possible to identify $\mathcal{P}$ (and, thus, $\mathcal{U}$), but up to a permutation and scale ambiguity, only if at most one source is Gaussian [26]. The scale ambiguity is indeed a sign ambiguity in our problem, as we set $\mathcal{U}^T\mathcal{U} = \mathcal{I}$. In conclusion, at best, the mixing matrix is identified by $\hat{\mathcal{U}} = \Pi\Sigma\mathcal{U}$, where Π is a permutation matrix and Σ a diagonal matrix whose elements are ±1. At best for the opponent, the secret carriers are identified up to a signed permutation (i.e., matrix $\Pi\Sigma$) ambiguity.

The likelihood to observe $\mathbf{v}$ for a given matrix $\mathcal{P}$ is $p(\mathbf{v};\mathcal{P}) = |\det\mathcal{P}|^{-1}p_{\mathbf{A}}(\mathcal{P}^{-1}\mathbf{v})$, and its score is

$$\frac{\partial}{\partial\mathcal{P}}\log p(\mathbf{v};\mathcal{P}) = -\mathcal{P}^{-T} + \mathcal{P}^{-T}\boldsymbol{\chi}(\mathcal{P}^{-1}\mathbf{v})\mathbf{v}^T\mathcal{P}^{-T} \tag{15}$$

with $\boldsymbol{\chi}(\mathbf{x}) = -(\partial/\partial\mathbf{x})\log p_{\mathbf{A}}(\mathbf{x})$ [27]. The asymptotic accuracy of the estimations is known to be only dependent on the symbol distribution and, especially, on its non-Gaussianity. As, in our case, symbols are i.i.d., denote by $\chi(.)$ the score function of $a_j(i)$ and by $\chi_n(.)$ the score function of a Gaussian random variable sharing the same variance (i.e., $\chi_n(x) = x/\sigma_a^2$). The trace of the Cramér–Rao Bound is then shown to be proportional to $(g^{-1}+1/2)/2N_o$ for large N_o [28], with g defined as

$$g = \frac{E\{(\chi(a)-\chi_n(a))^2\}}{E\{\chi_n(a)^2\}}. \tag{16}$$

However, g is not above bounded and tends to $+\infty$ when the symbols tend to have a discrete or bounded support. This is typically the case in watermarking, as the embedder would not allow the use of unbounded symbols for a perceptual distortion reason. In the case of discrete symbols, error-free mixing matrix recovery is possible within a finite number of observations. For instance, [29] shows a workable algorithm needing $N_o > {N_c}^2$ observations for BPSK symbols. In the case of bounded support symbols, the trace of the CRB decreases at a faster rate than $1/N_o$ [28], [30].

E. Watermarked Only Attack

In this section, the sources are unknown and can then be regarded as nuisance parameters [31], [32]. Vector $\boldsymbol{\psi}$ then equals $\partial\log L/\partial(\mathbf{u}_1^T\dots\mathbf{u}_{N_c}^T\mathbf{a}_1^T\dots\mathbf{a}_{N_o}^T)^T$, with the following $N_c\times 1$ vectors:

$$\frac{\partial\log L}{\partial\mathbf{a}_j} = \frac{\gamma}{\sigma_x^2\sqrt{N_c}}\mathcal{U}^T\mathbf{x}_j \qquad \forall j\in\{1,\dots,N_o\}. \tag{17}$$

The expectations of the products give the following subblocks:

$$E\left(\frac{\partial\log L}{\partial\mathbf{a}_j}\frac{\partial\log L}{\partial\mathbf{a}_k}^T\right) = \frac{\gamma^2}{N_c\sigma_x^2}(\mathcal{F}_{aa})_{j,k} = \frac{\gamma^2}{N_c\sigma_x^2}\mathcal{I}_{N_c}\delta_{j,k}$$

$$E\left(\frac{\partial\log L}{\partial\mathbf{u}_\ell}\frac{\partial\log L}{\partial\mathbf{a}_j}^T\right) = \frac{\gamma^2}{N_c\sigma_x^2}(\mathcal{F}_{ua})_{\ell,j} = \frac{\gamma^2}{N_c\sigma_x^2}(\mathcal{F}_{au})_{j,\ell}^T$$

where $\delta_{i,j}$ is the Kronecker function. We write with explicit notation

$$\text{FIM} = \frac{\gamma^2}{N_c\sigma_x^2}\begin{bmatrix}\mathcal{F}_{uu} & \mathcal{F}_{ua}\\ \mathcal{F}_{au} & \mathcal{F}_{aa}\end{bmatrix}. \tag{18}$$

Note that $\mathcal{F}_{aa} = \mathcal{I}_{N_oN_c}$. The CRB for estimated $\text{Vect}(\mathcal{U}) = (\mathbf{u}_1^T,\dots,\mathbf{u}_{N_c}^T)^T$ is given by

$$\text{CRB}(\text{Vect}(\mathcal{U})) = \frac{N_c\sigma_x^2}{\gamma^2}\widetilde{\mathcal{F}}_{uu}^{-1} \tag{19}$$

with $\widetilde{\mathcal{F}}_{uu} = (\mathcal{F}_{uu} - \mathcal{F}_{ua}\mathcal{F}_{aa}^{-1}\mathcal{F}_{au}) = (\mathcal{F}_{uu} - \mathcal{F}_{ua}\mathcal{F}_{au})$. It is known that in the general case, $\widetilde{\mathcal{F}}_{uu}^{-1} \geq \mathcal{F}_{uu}^{-1}$ (i.e., $\widetilde{\mathcal{F}}_{uu}^{-1} - \mathcal{F}_{uu}^{-1}$ is non-negative definite). In other words, nuisance parameters render the estimation of $\mathcal{U}$ less accurate [27]. However, the situation is even worse here as the FIM becomes singular. Indeed

$$(\mathcal{F}_{ua}\mathcal{F}_{au})_{\ell,k} = \sum_{j=1}^{N_o}(\mathcal{F}_{ua})_{\ell,j}(\mathcal{F}_{au})_{j,k} = \sum_{j=1}^{N_o}a_j(\ell)a_j(k)\mathcal{U}\mathcal{U}^T \tag{20}$$

and therefore, $\widetilde{\mathcal{F}}_{uu} = \mathcal{A}\mathcal{A}^T \otimes (\mathcal{I}_{N_v} - \mathcal{U}\mathcal{U}^T)$. As $(\mathcal{I}_{N_v} - \mathcal{U}\mathcal{U}^T)\mathbf{u}_k = \mathbf{0}$, $\widetilde{\mathcal{F}}_{uu}$ is singular.

This problem stems from two facts. First, we did not integrate some constraints during our derivation. Especially, we know that $\mathbf{u}_\ell^T\mathbf{u}_k = \delta_{\ell,k}$. An alternative expression for the bound in the case where the unconstrained problem is unidentifiable and the FIM noninvertible is given in [31].

However, the integration of the above-mentioned constraints in the derivation of the FIM is not sufficient for $N_c > 1$. The second fact is that an ambiguity remains about the order and "phase" of the carriers. The system is only identifiable up to a signed permutation. The case $N_c = 1$ is interesting, as constraint integration removes the FIM singularity because the ambiguity of the permutation does not exist.

1) One Carrier: The parameter vector to be estimated is composed of the unique carrier and the hidden symbols as nuisance parameters: $(\mathcal{U}^T\mathcal{A})$. Please note that $\mathcal{U}^T$ and $\mathcal{A}$ are row vectors in this case. The constraint on $\mathbf{u}_1$ is $(\|\mathbf{u}_1\|^2 - 1)/2 = 0$. The sequel is only the strict application of [31]. The $1 \times (N_v + N_o)$ gradient matrix of the constraint is equal to $\mathcal{G} = (\mathbf{u}_1^T\mathbf{0}_{N_o}^T)$, where $\mathbf{0}_N$ is an N zero vector. There exists a matrix $\mathcal{H} \in \mathbb{R}^{(N_v+N_o)\times(N_v+N_o-1)}$, whose columns form a basis for the nullspace of $\mathcal{G}$, that is, such that $\mathcal{G}\mathcal{H} = \mathbf{0}$. In our case, one particular choice of $\mathcal{H}$ is readily verified to be

$$\mathcal{H} = \begin{bmatrix}\mathcal{U}^\perp & \mathbf{0}\\ \mathbf{0} & \mathcal{I}_{N_o}\end{bmatrix} \tag{21}$$

where $\mathcal{U}^\perp$ is a basis of the complementary subspace of $\text{Span}(\mathbf{u}_1)$ in $\mathbb{R}^{N_v}$. Then, according to [31, Th. 1], the CRB under the above-mentioned constraint is $\text{CRB}(\mathcal{U}^T\mathcal{A}) = \mathcal{H}(\mathcal{H}^T\ \text{FIM}\ \mathcal{H})^{-1}\mathcal{H}^T$. With our choice of $\mathcal{H}$, this yields

$$\text{CRB}(\mathcal{U}^T\mathcal{A}) = \frac{\sigma_x^2}{\gamma^2}\begin{bmatrix}(\mathcal{A}\mathcal{A}^T)^{-1}\mathcal{U}^\perp\mathcal{U}^{\perp T} & \mathbf{0}\\ \mathbf{0} & \mathcal{I}_{N_o}\end{bmatrix} \tag{22}$$

and we finally get

$$\text{CRB}(\mathcal{U}^T) = \frac{\sigma_x^2}{\gamma^2}(\mathcal{A}\mathcal{A}^T)^{-1}\mathcal{U}^\perp\mathcal{U}^{\perp T} \xrightarrow{N_o\to+\infty} \frac{\sigma_x^2}{N_o\sigma_a^2\gamma^2}\mathcal{U}^\perp\mathcal{U}^{\perp T}. \tag{23}$$

2) N_c Carriers ($N_c > 1$): The ambiguity renders the Fisher Information Matrix singular, even when considering the constraints. However, Section V shows that in practice, the opponent builds noisy estimation of the carriers up to a signed permutation. A possibility in [32] is to pretend that the opponent knows N_m messages (for instance, $\{\mathbf{a}_\ell\}_{\ell=1}^{N_m}$), in order to *artificially* remove the ambiguity. This adds N_mN_c constraints of the type $\hat{a}_j(\ell) = a_j(\ell)$. At the end, calculation leads to

$$\text{CRB}(\text{Vect}(\mathcal{U})) = \frac{N_c\sigma_x^2}{\gamma^2}\mathcal{H}_{uu}\mathcal{B}^{-1}{\mathcal{H}_{uu}}^T \tag{24}$$

where $\mathcal{B}$ is the $N_c(N_v - N_m) \times N_c(N_v - N_m)$ matrix whose $(N_v - N_m) \times (N_v - N_m)$ blocks are

$$(\mathcal{B})_{\ell,k} = (\mathcal{A}\mathcal{A}^T)_{\ell,k}\mathcal{U}_\ell^{\perp T}\mathcal{U}_k^\perp - (\mathcal{A}_{N_m:N_o}{\mathcal{A}_{N_m:N_o}}^T)_{\ell,k}\mathcal{U}_l^{\perp T}\mathcal{U}\mathcal{U}^T\mathcal{U}_k^\perp$$

and $\mathcal{H}_{uu}$ the $N_cN_v \times N_c(N_v - 1)$ diagonal matrix whose $N_v \times (N_v - 1)$ blocks on the diagonal are $(\mathcal{H}_{uu})_{\ell,\ell} = \mathcal{U}_\ell^\perp$. In these expressions, the columns of $\mathcal{U}_\ell^\perp$ form an orthonormal basis of the complementary subspace of $\text{Span}(\mathbf{u}_\ell)$, and $\mathcal{A}_{N_m:N_o} = (\mathbf{a}_{N_m+1}\dots\mathbf{a}_{N_o})$. However, the minimal number N_m to remove the ambiguity depends on the symbols' pdf [32].

Facing the difficulty of finding the right parameter N_m and some cumbersome calculus, we prefer to approximate the information leakage about a carrier by (23), where γ^2 is replaced by the power per carrier γ^2/N_c. The security level is then $N_o^\star = N_c\sigma_x^2/\sigma_a^2\gamma^2$, which is, by the way, coherent with (24). This result is quite surprising because the security level is the same against KMA and WOA. Yet, the estimation of the secret carriers remains, up to a signed permutation in the WOA.

F. Possible Hacks

The conclusion of this security analysis stands in the different possibilities to forge pirated content.

- The pirate discloses secret subspace $\text{Span}(U)$. He can now focus the attack's noise in this subspace to jam the communication far more efficiently. He can also nullify the watermarked signals projection in this subspace to remove the watermark.
- The pirate discloses the secret carriers up to a signed permutation. The above-mentioned hacks are still possible. Besides, he can detect whether two watermarked pieces of content share the same hidden message. He can also flip some randomly chosen bits. Moreover, the accidental knowledge of hidden messages in few watermarked pieces of content might remove this ambiguity. This extra security analysis indeed pertains to Section III-D.
- The pirate discloses the secret carriers. He has a full access to the watermarking channel to read, write, or erase a hidden message.

Of course, the quality of the pirated pieces of content depends on the accuracy of his estimation. The authors focus on this aspect in [33].

V. Algorithms for Spread Spectrum-Based Techniques

Section III not only gives security levels of the substitutive method but also contains almost practical implementations of

workable algorithms. On the contrary, Section IV only presents theoretical assessment of security levels. Hence, this section deals with practical algorithms that are useful to hack spread spectrum-based watermarking schemes. For each attack, an algorithm is presented and tested on synthetic data as supposed by the model of (9) with BPSK symbols and Gaussian host vectors. At the end of the section, these algorithms are applied on spread transform side information methods and one still image technique.

This section has an intensive use of PCA and ICA algorithms, which is completely new in watermarking security analysis, as the only other papers mentioning PCA/ICA in the watermarking community have different purposes. ICA was used in [34] and [35] to design a watermarking embedder. A technique for estimating the watermark by observing only one image was presented in [36]. Their purpose is the simple erasure of the whole watermark signal and not the disclosure of the secret parameters, whereas the approach here allows a complete access to the watermarking communication channel to remove, read, or write hidden data.[2]

The following average normalized correlation measures the efficiency of our attack:

$$\eta = \frac{1}{N_c} \sum_{\ell=1}^{N_c} \frac{\hat{\mathbf{u}}_\ell^T \mathbf{u}_\ell}{\|\hat{\mathbf{u}}_\ell\|}. \tag{25}$$

Although the normalization renders estimators $\hat{\mathbf{u}}_j/\|\hat{\mathbf{u}}_j\|$ biased [38], the normalized correlation is preferred because it is an extremely popular measure in the watermarking community. $\eta \lesssim 1$ means that the opponent discloses vectors that are almost collinear with the secret carriers. When existing, we manually removed the ambiguity of the signed permutation. Measures of η are done by averaging $N_t = 128$ experimental results.

The relation with the theoretical security levels is not difficult to find. Equation (25) is in expectation the cosine of the angle between $\mathbf{u}_\ell$ and $\hat{\mathbf{u}}_\ell = \mathbf{u}_\ell + \mathbf{n}$, where $\mathbf{n}$ is the estimation noise (orthogonal to $\mathbf{u}_\ell$ and whose norm is $\sqrt{\text{tr}(\text{CRB}(\text{Vect}(\mathcal{U})))/N_c}$, where $\text{tr}(A)$ is the trace of matrix A). The following relation holds:

$$\eta \approx \frac{\|\mathbf{u}_\ell\|}{\sqrt{\|\mathbf{u}_\ell\|^2 + \frac{\text{tr}(\text{CRB}(\text{Vect}(\mathcal{U})))}{N_c}}}. \tag{26}$$

A. Known Message Attack

Observing $(\mathbf{y}, \mathbf{a})^{N_o}$, the opponent can use the Maximum Likelihood Estimator (MLE) related to (11). This estimator is also defined by $\partial \log L/\partial \mathbf{u}_\ell = \mathbf{0} \quad \forall \ell \in \{1, \ldots, N_c\}$, which gives

$$\hat{\mathcal{U}} = \frac{\sqrt{N_c}}{\gamma} (\mathcal{Y}\mathcal{A}^T)(\mathcal{A}\mathcal{A}^T)^{-1}. \tag{27}$$

The MLE is known to be unbiased and consistent, i.e., it asymptotically achieves the CRB derived in Section IV-D. Fig. 5 shows experimental values of η against N_o and WCR $= \gamma^2/\sigma_x^2$ for the DSSS case. The locus of points such that $\eta = const$ are projected on the plane $\eta = 0$. They appear to be parallel with the curve $N_o = N_c\sigma_x^2/\gamma^2$. Tests done with different N_v confirm that the efficiency of the attack does not depend on the vector length. This asserts the theoretical security level of Section IV-C.

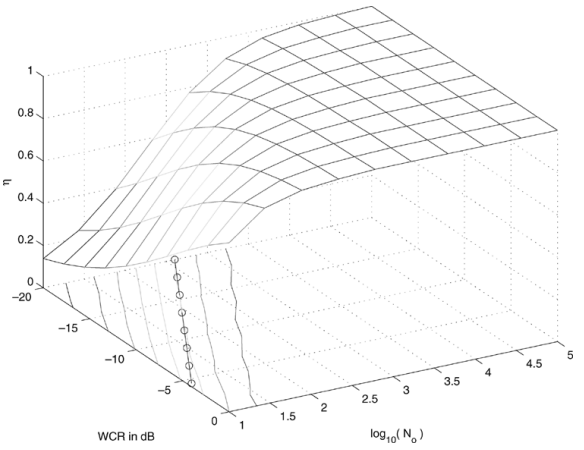


Fig. 5. KMA for DSSS ($N_c = 4, N_v = 512$). η against $\log_{10}(N_o)$ and WCR in decibels. The curve $N_o = N_c\sigma_x^2/\gamma^2$ is plotted with small circles.

B. Known Original Attack

In this case, the opponent observes several instances of $\mathbf{d}_j = (\mathbf{y}_j - \mathbf{x}_j) \propto \mathcal{U}\mathbf{a}_j$. As seen in Section IV-D, this is related to the well-known problem of signal processing called BSS with no noise. A lot of papers have already been written on BSS, and we just recall here its most common algorithms. Note that spread spectrum corresponds to the BSS overdetermined case (i.e., $N_v \geq N_c$).

The most classical algorithm in BSS is the PCA. Denote $\mathcal{D} = \mathcal{Y} - \mathcal{X}$. This technique makes an eigendecomposition of the matrix $\mathcal{D}\mathcal{D}^T = \gamma^2\mathcal{U}\mathcal{A}\mathcal{A}^T\mathcal{U}^T/N_c$. This corresponds to a Gram–Schmidt orthogonalization of vectors $\mathbf{d}^{N_o}$. Please note that $\rho \triangleq \text{Rank}(\mathcal{A})$ is also the rank of $\mathcal{D}\mathcal{D}^T$. Hence, the decomposition outputs ρ orthonormal vectors lying in Span $(\mathcal{U})$. In the best case, the opponent has $\rho = \min(N_o, N_c)$. Nevertheless, in reality, he may have $\rho \leq \min(N_o, N_c)$ if the N_o symbol vectors are linearly dependent.

When successful (i.e., when $\rho = N_c$), the PCA technique yields a orthonormal basis of the secret subspace Span $(\mathcal{U})$. The possibilities to hack watermarked pieces of content when Span $(\mathcal{U})$ is disclosed are summarized in Section IV-F. Yet, the vectors of this basis are not necessarily collinear with the private carriers. This is due to the unitary matrix $\mathcal{P}$ mentioned in Section IV-D. The opponent cannot decode, as projection of watermarked signals onto this basis gives a mixture of the hidden symbols. This is illustrated by Fig. 6. The same reason prevents him transmitting information in the hidden channel.

Nevertheless, under the assumption that the symbol vectors are *statistically* independent, the opponent can resort to a more powerful tool: the Independent Component Analysis (ICA). It is an extension of PCA, constraining the output estimated symbol vectors to be independent [26]. Good tutorials on ICA and on its

[2]We discovered after submission a similar approach that is uniquely devoted to watermark removal and only based on PCA in [37].

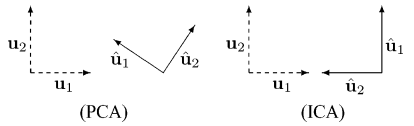


Fig. 6. PCA versus ICA. PCA finds the secret carriers up to a rotation, whereas ICA succeeds in aligning the estimated carriers $\hat{\mathbf{u}}^{N_c}$ with $\mathbf{u}^{N_c}$ (Here, $N_c = 2$). An ambiguity remains in their order (permutation) and orientation (sign).

links with BSS are [28] and [39]. A very general ICA algorithm named FastICA [40] has been preferred to algorithms dedicated to specific symbol distribution [29], [30].

In short, ICA algorithms usually work in the basis recovered by a PCA. This basis describes exactly the secret subspace (provided that $\rho = N_c$). The problem is now reduced to the estimation of the $N_c \times N_c$ matrix $\mathcal{P}$. Hence, parameter N_v has absolutely no influence on the attack. Then, in an iterative process, the ICA "rotates" the basis until it nullifies an objective function (often called a constrast function) of the estimated sources $\hat{\mathbf{a}}^{N_o}$. This function can be an approximation of the mutual information of the estimated sources. Contrast functions depend on the distribution of the symbol sources. However, this measure reflects statistical independence only for large N_o. For a finite number of observations, ICA algorithms usually search for a minimum of the contrast function with the help of a gradient descent technique.

When successful, ICA reduces the set of ambiguity matrices $\mathcal{P}$ to the one of signed permutations. This is illustrated by Fig. 6. Section IV-F lists the possibilities to hack watermarked pieces of content when the carriers are disclosed up to a signed permutation.

C. Watermarked Only Attack

The WOA case is quite similar to KOA, as it is related to BSS in a noisy environment. The covariance matrix $\mathcal{R}_{\mathbf{y}}$ has the following expression:

$$\mathcal{R}_{\mathbf{y}} = \mathcal{R}_{\mathbf{x}} + \frac{\gamma^2}{N_c}\mathcal{U}\mathcal{R}_{\mathbf{a}}\mathcal{U}^T = \sigma_x^2\mathcal{I} + \frac{\gamma^2\sigma_a^2}{N_c}\mathcal{U}\mathcal{U}^T. \qquad (28)$$

Its diagonalization leads to N_c eigenvalues equaling $\sigma_x^2 + \gamma^2\sigma_a^2/N_c$ and $N_v - N_c$ eigenvalues equaling σ_x^2. Hence, the eigenvectors related to the N_c biggest values constitute a basis of Span $(\mathcal{U})$, which is also known as the signal space in blind equalization for digital communications.

PCA estimates covariance matrix $\mathcal{R}_{\mathbf{y}}$ by $\mathcal{Y}\mathcal{Y}^T/N_o$ and outputs N_c eigenvectors whose eigenvalues are the biggest ones. Due to this rough estimation, these vectors do not live exactly in Span $(\mathcal{U})$. Compared to Fig. 6, these noisy estimation vectors would not lie in the plan of the page, which is regarded as subspace Span $(\mathcal{U})$ in this simple example. However, ICA will still try to rotate them in order to render the decoded symbols independent. Fig. 7 shows the locus of points such that $\eta = const$ for different values of N_c and N_o with the DSSS method (i.e., a BPSK modulation). The ICA algorithm meets the theoretical limit only for large N_o and high energy of watermark signal per carrier: $\gamma^2 N_v/N_c$. Note that for $N_c = 4$, the gap between experimental performances and the theoretical limit gets larger.

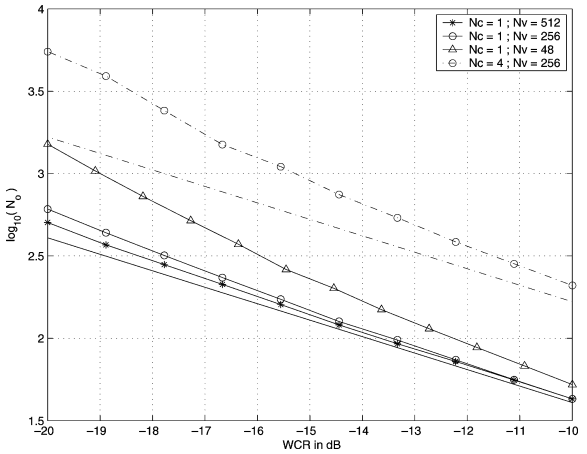


Fig. 7. WOA for DSSS. Operating points achieving $\eta = 0.8$ for different parameters N_c and N_v. The solid line is the theoretical limit for $N_c = 1$, and curves with stars, circles, and triangles are the experimental results. They capture the efficiency of the PCA, as only one carrier is used. The dashed line is the theoretical limit for $N_c = 4$ (i.e., the solid line translated of $\log_{10}(N_c)$), and the dashed curve with circles is the experimental results with the FastICA algorithm [40].

D. Extension to Spread Transform Side Information Watermarking

This subsection presents experiments with side information watermarking using the process on spread spectrum. In these methods, the symbols $a_j(\ell)$ depend on the host signal in the following way:

$$a_j(\ell) = f(m_j(\ell), \mathbf{u}_\ell^T\mathbf{x}_j). \qquad (29)$$

Three techniques were investigated: Improved Spread Spectrum (ISS) [23], the Scalar Costa Scheme (SCS) [21], and Maximized Robustness Embedding (MRE) [22]. Two implementations of SCS have been done. The carriers have disjoint supports in the first one, which is a possible interpretation of [21]: $\mathbf{u}_1 = (\mathbf{u}^T\mathbf{0}_\tau^T \ldots \mathbf{0}_\tau^T)^T$, $\mathbf{u}_2 = (\mathbf{0}_\tau^T\mathbf{u}^T \ldots \mathbf{0}_\tau^T)^T$, and so on, with $\tau N_c = N_v$. The second implementation is called SCS with Subspace Projection (SSP) [41]. The carriers have full support and are orthonormal. The embedding distortion, the vector length, and the number of hidden bits are the same for a fair comparison.

The KMA case has not been investigated. The knowledge of the messages does not usually imply the disclosure of the symbols. In SCS, function $f(.)$ of (29) is private and depends on a secret key (i.e., a dithering vector). However, information about the symbols may leak from the message. Symbols are Gaussian variables centered on $\gamma(-1)^{m_j(\ell)}$ for the ISS technique:

$$a_j(\ell) = \gamma(-1)^{m_j(\ell)} - \lambda\mathbf{u}_\ell^T\mathbf{x}_j. \qquad (30)$$

We foresee that the MLE algorithm could easily be tuned to exploit this information leakage.

The KOA is simpler, as the basic assumption is still valid: $\mathbf{u}_\ell^T\mathbf{x}_j$ and $\mathbf{u}_k^T\mathbf{x}_j$ ($k \neq \ell$) are Gaussian distributed and noncorrelated; thus, the symbols are statistically independent. Yet, the efficiency of BSS depends on the symbols distribution; therefore, we expect different performances. Once again, in our simulation, the opponent always uses the same generic ICA algorithm.

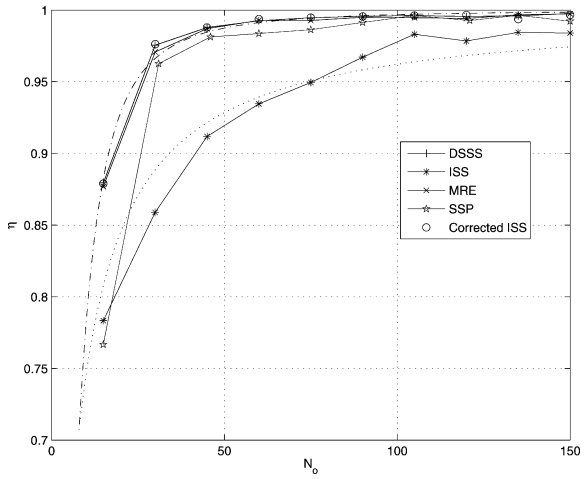


Fig. 8. KOA for four different watermarking techniques ($N_c = 4$, $N_v = 512$). Dotted line: $\eta = (1 + k/N_o)^{-1}$. Dash-dotted line: $\eta = (1 + (k/N_o)^2)^{-1}$.

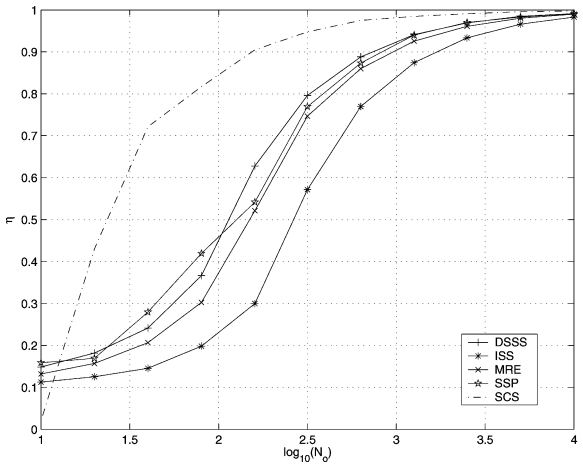


Fig. 9. WOA for five different watermarking methods ($N_v = 512$, $N_c = 4$, WCR $= -15$ dB). $\tau = 128$ for SCS. For SCS, SSP, and ISS, the embedding parameters are optimal for an expected noise attack whose distortion equals the embedding distortion: WNR $= 0$ dB.

No fine tuning according to the expected symbols distribution is done. Fig. 8 shows the results, except for SCS.[3] Surprisingly, the rate of the noise estimation variance is in $1/{N_o}^2$ for DSSS, SSP, and MRE. This seems to be due to the bounded support feature of the symbols in these methods, despite the use of a generic algorithm. For ISS, the rate is in $1/N_o$. Please note that according to (30), the KOA for ISS is similar to a WOA for the SS method, with a watermark to host power ratio of $\gamma^2/\lambda^2\sigma_x^2$. A smarter attack on ISS stems from this remark. First, difference vectors are used to disclose the secret subspace with a PCA. Then, they are corrected in adding the projection of the original vectors scaled by a factor λ. We are now in a situation similar to a KOA with DSSS. Finally, ICA finishes the job by working on the corrected vectors. The last curve, called "Corrected ISS" in Fig. 8, shows the dramatic improvement. The security level of ISS is in practice as low as the DSSS one.

The WOA is also straightforward as we applied the same ICA algorithm for DSSS, ISS, MRE, and SSP. For SCS, the observed watermarked vectors are split by chunks of τ samples. Thus, the opponent has ${N_o}' = N_o\tau$ vectors, whose length is ${N_v}' = N_v/\tau$ and watermarked with an ${N_c}' = 1$ secret carrier. The algorithm is thus a simple PCA in this case. Fig. 9 shows the results. SCS (or, more precisely, the way we have implemented it) is obviously the less secure. However, the simple change brought in the implementation of SSP is sufficient to correct this security flaw.[4] The other techniques share the same security level. ISS seems to be slightly more secure; however, remember that we did not tune the contrast function of the ICA algorithm. In the same way, the embedding parameters (γ, λ) play a big role in the symbol's distribution, and the attack might thus perform differently. This is the reason why we prefer to look at the global shape of the curves, rather than to draw erroneous conclusions from these meager differences.

[3]For SCS, $N_o = 1$ is enough to disclose small length carrier $\mathbf{u}$ up to a sign.

[4]We only analyze here the security of the spreading transform. Yet, the dithering vector in the SCS-like technique constitutes a second barrier, which will be the subject of a future work.

(a) (b)

Fig. 10. Comparison between the two pirated Lena images. This is their best quality for a successful attack. Pirate A: PSNR $= 21.8$ dB, Pirate B: PSNR $= 35.8$ dB. (a) Pirate A. (b) Pirate B.

E. Application to a Robust Watermarking Technique for Still Images

The goal of this last subsection is to demonstrate the power of "smart" attacks based on secret carrier estimation. So far, this paper has investigated the first phase of the attack: the secret disclosure. Now, in a second phase, the opponent uses this *a posteriori* information to hack pieces of content, which were watermarked with the same secret key. To this end, the subsection deals with real still images. The robust watermarking technique from [20] has been chosen.

A challenge is proposed to two opponents: They attack a watermarked image with an increasing attack distortion, until an oracle warns them that the decoded message is different from the embedded message ($N_c = 8$ bits, PSNR $= 38$ dB). Pirate A uses *blind* attacks (i.e., pertaining to the robustness issue–except any geometric attack). For instance, in this paper, he scales the size of the image by a quarter, JPEG compresses it with a decreasing quality factor, and finally scales back the image. Pirate B uses *smart* attacks. He has estimated the secret carriers by a WOA,

with $N_o \sim 1000$ images such that $\eta = 0.5$,[5] and he tries to remove the hidden information for one carrier. Details of algorithm adaptations to real images may be found in [33]. Fig. 10 shows the result of the challenge for the Lena image. For a panel of 50 pictures (512×512 pixels), pirate B, on average, produces an attack distortion that is 15 dB smaller than Pirate A to successfully hack watermarked pictures.

VI. Conclusion

As in cryptanalysis, measurement of information leakage is the fundamental principle underlying the theoretical framework for robust watermarking security assessment presented in this paper. A watermarking technique, even if it is robust, is not secure if the opponent can refine his knowledge on the presumably secret key while pieces of content are watermarked with the same key. The security level is then defined by the number of observations the opponent needs in order to accurately estimate the secret key.

The conclusion of this paper is not that spread spectrum-based watermarking techniques or substitutive schemes are broken. The goal is to warn the watermarking community that security is a crucial issue. Designers should not only control the imperceptibility and the robustness of their schemes but also assess their security levels. Depending on the application designers are targeting (and especially on the observations available to the pirate), watermarking several pieces of content with the same key might bring threats. This potentially raises difficulties on the key management. For instance, it is not clear how a blind watermarking decoder will be informed of the secret key if this later one is to be changed according to the security levels assessed in this paper.

References

[1] I. Cox, M. Miller, and J. Bloom, *Principles and Practice*. San Francisco, CA: Morgan Kaufmann, 2001.

[2] J. O'Ruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Process.*, vol. 66, no. 3, pp. 303–17, 1998.

[3] S. Pereira and T. Pun, "Fast robust template matching for affine resistant image watermarks," in *Proc. IHW*, A. Pfitzmann, Ed. Dresden, Germany: Springer-Verlag, 1999, pp. 199–210.

[4] I. Cox, M. Miller, and A. McKellips, "Watermarking as communication with side information," *Proc. IEEE*, vol. 87, no. 7, pp. 1127–1141, Jul. 1999.

[5] B. Chen and G. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[6] P. Moulin, "The role of information theory in watermarking and its application to image watermarking," *Signal Process.*, vol. 81, pp. 1121–1139, 2001.

[7] T. Kalker, "Considerations on watermarking security," in *Proc. MMSP*, Cannes, France, Oct. 2001, pp. 201–206.

[8] S. Craver, N. Memon, B.-L. Yeo, and M. M. Yeung, "On the invertibility of invisible watermarking technique," in *Proc. IEEE ICIP* Washington, DC, 1997, pp. 540–543.

[5]Of course, the opponent cannot know this last value, as he does not have the secret carriers. However, nothing prevents him from running simulations with his own private carriers in order to get an estimation of η.

[9] M. Kutter, S. Voloshynovskiy, and A. Herrigel, "Watermark copy attack," in *Proc. SPIE Security and Watermarking of Multimedia Contents II*, vol. 3971, P. W. Wong and E. Delp, Eds., San Jose, CA, Jan. 2000.

[10] I. Cox and J.-P. Linnartz, "Some general methods for tampering with watermarks," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 5, pp. 587–93, May 1998.

[11] J. P. Linnartz and M. van Dijk, "Analysis of the sensitivity attack against electronic watermarks in images," in *Proc. IHW*, D. Aucsmith, Ed. Portland, OR: Springer-Verlag, 1998, vol. 1525, Lecture Notes in Computer Science.

[12] T. Mittelholzer, "An information-theoritic approach to steganography and watermarking," in *Proc. IHW*, A. Pfitzmann, Ed. Dresden, Germany: Springer-Verlag, 1999, pp. 1–17.

[13] T. Furon and P. Duhamel, "An asymmetric watermarking method," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 981–995, Apr. 2003.

[14] M. Barni, F. Bartolini, and T. Furon, "A general framework for robust watermarking security," Signal Process., vol. 83, no. 10, pp. 2069–2084, Oct. 2003, to be published.

[15] A. Kerckhoffs, "La cryptographie militaire," *J. Des Sci. Militaires*, vol. 9, pp. 5–38, Jan. 1883.

[16] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, pp. 656–715, Oct. 1949.

[17] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 6, pp. 644–654, Nov. 1976.

[18] S. Burgett, E. Koch, and J. Zhao, "Copyright labeling of digitized image data," *IEEE Commun. Mag.*, vol. 36, no. 3, pp. 94–100, Mar. 1998.

[19] D. Kahn, "Cryptology and the origins of spread spectrum," *IEEE Spectrum*, vol. 21, pp. 70–80, Sep. 1984.

[20] S. Pateux and G. Le Guelvouit, "Practical watermarking scheme based on wide spread spectrum and game theory," *Signal Processing: Image Commun.*, vol. 18, pp. 283–296, Apr. 2003.

[21] J. Eggers, R. Baüml, R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1003–1019, Apr. 2003.

[22] M. Miller, I. Cox, and J. Bloom, "Informed embedding: Exploiting image and detector information during watermark insertion," in *Proc. ICIP*, Vancouver, BC, Canada, Sep. 2000.

[23] H. S. Malvar and D. A. F. Florêncio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 868–905, Apr. 2003.

[24] D. T. Pham and J. F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sep. 2001.

[25] J. Su, J. Eggers, and B. Girod, "Analysis of digital watermarks subjected to optimum linear filtering and additive noise," *Signal Process.*, vol. 81, pp. 1141–1175, 2001.

[26] P. Comon, "Independent component analysis, A new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[27] S.-I. Amari and J. F. Cardoso, "Blind source separation; Semiparametric statistical approach," *IEEE Trans. Signal Process.*, vol. 45, 1997.

[28] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[29] A.-J. van der Veen, "Blind separation of BPSK sources with residual carriers," *Signal Process.*, vol. 73, no. 10, pp. 67–79, Jan. 1999.

[30] F. Gamboa and E. Gassiat, "Source separation when the input sources are discrete or have constant modulus," *IEEE Trans. Signal Process.*, vol. 45, no. 12, pp. 3062–3072, Dec. 1997.

[31] P. Stoica and B. C. Ng, "On the Cramér-Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, Jul. 1998.

[32] Y. Yao and G. B. Giannakis, "On regularity and identifiability of blind source separation under constant-modulus constraints," *IEEE Trans. Signal Process.*, vol. 53, no. 4, Apr. 2005.

[33] F. Cayre, C. Fontaine, and T. Furon, "Watermarking attack: Security of wss techniques," in *Proc. IWDW*, Seoul, Korea, 2004.

[34] F. J. González-Serrano and J. J. Murillo-Fuentes, "Independent component analysis applied to image watermarking," in *Proc. IEEE ICASSP*, 2001.

[35] S. Bounkong, B. Toch, D. Saad, and D. Lowe, "ICA for watermarking digital images," *J. Machine Learning Res.*, vol. 1, pp. 1–25, 2002.

[36] J. Du, C.-H. Lee, H.-K. Lee, and Y. Suh, "Watermark attack based on blind estimation without priors," in *Proc. IWDW*, Lecture Notes in Computer Science, 2002.

[37] G. Doërr and J.-L. Dugelay, "Danger of low-dimensional watermarking subspaces," in *Proc. IEEE ICASSP* Montreal, QC, Canada, 2004, vol. 3.

[38] P. Stoica and B. Ng, "Performance bounds for blind channel estimation," in *Signal Processing Advances in Wireless and Mobile Communications*. Englewood Cliffs, NJ: Prentice-Hall, 2001, vol. 1, pp. 41–62.

[39] A. Hyvärinen and E. Oja, "Independent component analysis: A tutorial," *Neural Networks*, vol. 13, no. 4–5, pp. 411–430, 2000.

[40] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 3, pp. 626–634, May 1999.

[41] R. Fischer, R. Tzschoppe, and R. Bäuml, "Lattice costa schemes using subspace projection for digital watermarking," *Eur. Trans. Telecommun.*, vol. 15, no. 4, pp. 351–362, Aug. 2004.

[42] A. Akansu, E. Delp, T. Kalker, B. Liu, N. Memon, P. Moulin, and A. Tewfik, "Special issue on signal processing for data hiding in digital media and secure content delivery," *IEEE Trans. Signal Process.*, vol. 51, no. 4, Apr. 2003.

François Cayre received the Computer Science and M.S. degrees from the Université de Technologie de Compiègne, Compiègne, France, in 2000. In 2003, he received the Ph.D. degree from both the Université Catholique de Louvain, Louvain-le-Neuve, Belgium, and the École Nationale Supérieure des Télécommunications, Paris, France.

He is currently a post-doctoral fellow at INRIA, Rennes, France, where he is investigating the security of watermarking schemes. His main interests include image and 3-D watermarking, watermarking security and benchmarking, and 3-D mesh coding.

Mr. Cayre was co-recipient of the IWDW'04 Best Paper Award.

Caroline Fontaine received the Ph.D. degree in computer science from the University of Paris 6, Paris, France, in 1998, for a work dealing with cryptography, error-correcting codes, and digital watermarking.

She has been with the Computer Science Lab at the University of Lille 1, Lille, France, as an Associate Professor from 1999 to 2002, where she is now a CNRS permanent assistant researcher. Her research interests include cryptography and cryptanalysis (mainly of symmetric encryption schemes), digital watermarking, and security of mobile *ad hoc* networks.

Teddy Furon received the M.S. degree in digital communications in 1998 and the Ph.D. degree in signal and image processing in 2002 from the Ecole Nationale Supérieure des Télécommunications, Paris, France.

From 1998 to 2001, he was a research engineer with the Security Lab of THOMSON Multimedia, Rennes, France, working on digital watermarking in the framework of copy protection. He continued working on digital watermarking as a postdoctoral researcher at the TELE Lab., Université Catholique de Louvain, Louvain-la-Neuve, Belgium. He is now a researcher with the INRIA institute, working within the TEMICS project of the IRISA public research center, Rennes, France.

Dr. Furon was a co-recipient of the IWDW'04 Best Paper Award.

An Asymmetric Fingerprinting Scheme Based on Tardos Codes

Ana Charpentier[1,⋆], Caroline Fontaine[2], Teddy Furon[1], and Ingemar Cox[3]

[1] INRIA-Rennes research center, Campus de Beaulieu, Rennes, France
[2] CNRS/Lab-STICC/CID, Télécom Bretagne/ITI, Brest, France
[3] University College London, Dpt. of Computer Science, London, United Kingdom

Abstract. Asymmetric fingerprinting protocols are designed to prevent an untrustworthy Provider incriminating an innocent Buyer. These protocols enable the Buyer to generate their own fingerprint by themself, and ensure that the Provider never has access to the Buyer's copy of the Work. Until recently, such protocols were not practical because the collusion-resistant codes they rely on were too long. However, the advent of Tardos codes means that the probabilistic collusion-resistant codes are now sufficiently short that asymmetric fingerprint codes should, in theory, be practical.

Unfortunately, previous asymmetric fingerprinting protocols cannot be directly applied to Tardos codes, because generation of the Tardos codes depends on a secret vector that is only known to the Provider. This knowledge allows an untrustworthy Provider to attack traditional asymmetric fingerprinting protocols. We describe this attack, and then propose a new asymmetric fingerprinting protocol, specifically designed for Tardos codes.

1 Introduction

This paper considers a problem arising in the fingerprinting of digital content. In this context, a fingerprint is a binary code that is inserted into a Work for the purpose of protecting it from unauthorized use, or, more precisely, for the purpose of identifying individuals responsible for its unauthorized use. In such a scenario, it is assumed that two or more users may collude in order to try to hide their identities. Under the *marking assumption* [2], colluders cannot alter those bits of the code that are identical for all colluders. However, where bits differ across colluders, these bits may be assigned arbitrary values. A key problem is resistance to collusion, i.e. if c users create a pirated copy of the Work, its tampered fingerprint (i) should not implicate innocent users, and (ii) should identify at least one of the colluders.

This problem has received considerable attention since Boneh and Shaw [2] discussed it. They introduced the concept of a c-secure code such that the probability of framing an innocent user is lower than ϵ. Unfortunately, the length of

⋆ Supported by National Project MEDIEVALS ANR-07-AM-005.

T. Filler et al. (Eds.): IH 2011, LNCS 6958, pp. 43–58, 2011.

their codes, $O(c^4 \log(\frac{n}{\epsilon}) \log(\frac{1}{\epsilon}))$ where n is the number of users, was too long to be practical. Following Boneh and Shaw's paper, there has been considerable effort to design shorter codes. In 2003, Tardos [19] proposed an efficient code construction that, for the first time, reduced the code length to the theoretical lower bound, $O(c^2 \log(\frac{n}{\epsilon}))$, thereby making such codes practical. Tardos codes are currently the state-of-the-art for collusion-resistant fingerprinting.

Contemporaneously, some papers considered the scenario where the Provider is untrustworthy. Given knowledge of a Buyer's fingerprint, the Provider creates a pirated copy of a Work, implicating the innocent Buyer. To prevent this, Pfitzman and Schunter [16] first introduced the concept of asymmetric fingerprinting in which the Provider does not need to know the Buyer's fingerprint. The Buyer first commits to a secret (the fingerprint) that only he/she knows. The Buyer and Provider then follow a protocol which results in the Buyer receiving a copy of the Work with his/her secret fingerprint (and some additional information coming from the Provider) embedded within it. The Provider does not learn the Buyer's secret, and cannot therefore create a forgery. Unfortunately, the early implementations of this concept were not practical due to the very long length of the collusion resistant codes. The advent of Tardos codes has reduced the length of the collusion resistant codes to a practical size. However, generation of these codes depends on a probability distribution based on a secret vector that is only known to the Provider. This knowledge is sufficient for the Provider to circumvent traditional asymmetric fingerprinting protocols.

In the next Section, we briefly summarize the design of Tardos codes. We then describe how an untrustworthy Provider, with knowledge of the secret vector needed to generate the Tardos codes, can false accuse an innocent Buyer. Section 3 then describes a new asymmetric fingerprinting protocol specific to the use of Tardos codes, that prevents both the Buyer and the Provider from cheating. Practical aspects of the fingerprints embedding and accusation are discussed in Section 4, while security and efficiency of the whole scheme are discussed in Section 6.

2 Untrustworthy Provider with the Tardos Code

For readers unfamiliar with Tardos codes, we now provide a brief introduction. Further details can be found in [18].

2.1 Introduction to Tardos Codes

Let n denote the number of buyers, and m the length of the collusion-resistant codes. The fingerprints can then be arranged as a binary $n \times m$ matrix $\mathbf{X}$, where Buyer j's binary fingerprint is the jth row of the matrix, i.e. $\mathbf{X}_j = (X_{j1}, X_{j2}, \ldots, X_{jm})$.

To generate this matrix, m real numbers $p_i \in [t, 1-t]$ are generated, each of them being randomly and independently drawn according to the probability density function $f : [t, 1-t] \to \mathbb{R}^+$ with $f(z) = \kappa(t)(z(1-z))^{-1/2}$ and

$\kappa(t)^{-1} = \int_t^{1-t}(z(1-z))^{-1/2}dz$. The parameter $t \ll 1$ is referred to as the cutoff whose value is around $1/300c$. The resulting vector, $\mathbf{p} = (p_1, \ldots, p_m)$ is a secret only known by the Provider. Each element of the matrix $\mathbf{X}$ is then independently randomly drawn, such that the probability that the element X_{ji} is set to symbol '1' is $\mathbb{P}(X_{ji} = 1) = p_i$. The collusion-resistant fingerprint, $\mathbf{X}_j$, is then embedded into Buyer j's copy of the Work. This embedding can be accomplished by a variety of watermarking techniques.

When an unauthorized copy is found, a binary sequence, $\mathbf{Y}$, is extracted from the copy thanks to the watermark decoder. Due to collusion and possible distortions such as transcoding, this binary sequence is unlikely to exactly match one of the fingerprints in the matrix $\mathbf{X}$. To determine if Buyer j is involved in the creation of the unauthorized copy, a score, referred to as an accusation score, S_j is computed. If this score is greater than a given threshold Z, then Buyer j is considered to have colluded. The value of the threshold Z theoretically guarantees that the probability of accusing an innocent person is below a significance level, ϵ.

The scores are computed according to an accusation function g, reflecting the impact of the correlation between the fingerprint $\mathbf{X}_j$, associated with Buyer j, and the decoded sequence $\mathbf{Y}$:

$$S_j = G(\mathbf{Y}, \mathbf{X}_j, \mathbf{p}) = \sum_{i=1}^{m} g(Y_i, X_{ji}, p_i). \tag{1}$$

In the usual symmetric codes [18], the function g is constrained (for example, for an innocent person, the expectation of the score is zero and its variance is m), giving $g(1,1,p) = g(0,0,1-p) = -g(0,1,p) = -g(1,0,1-p) = \sqrt{\frac{1-p}{p}}$.

2.2 Untrustworthy Content Provider

We now consider the case where the Provider is no longer trusted, and wishes to frame Buyer j. There are a number of scenarios, depending on the knowledge available to the Provider. We briefly outline these and discuss our specific scenario in detail.

The Provider Knows the Buyer's Fingerprint and How to Embed the Corresponding Watermark. This scenario provides no protection to the Buyer. The Provider can simply watermark a Work with the fingerprint of Buyer j, place the Work in an incriminating location and then accuse Buyer j.

The Provider Knows the Buyer's Fingerprint. In this scenario the Provider does not have the ability to watermark a Work. Instead, upon a Provider's request, a trusted Technology Provider embeds the fingerprint into a Work and sends the fingerprinted Work to the Buyer. We emphasize that the Technology Provider is trusted, and as such, the Provider cannot embed the same fingerprint into a Work and have it delivered to two different users, one of which is colluding with the Provider to frame the other user. If the Technology Provider were not trusted, we would be back to the previous scenario.

All the Provider needs is fingerprinted copies from $c \geq 3$ fake users or colluders. There is nothing special about the particular fingerprints. For a given Buyer j, whom the Provider wishes to frame, the Provider knows where the elements of the Buyer's fingerprint $X_{ji} = 1$. This happens with probability p_i. At least one of the accomplices has the same symbol as the Buyer with a probability of $1 - (1 - p_i)^c$. Therefore, given that the Provider knows the Buyer's fingerprint, $\mathbf{X}_j$, the accomplices can forge a sequence very similar to the fingerprint of Buyer j. More specifically, if $Y_i = X_{ji}$ whenever the marking assumption allows it, then the forgery is such that, in expectation, the score of Buyer j becomes:

$$\begin{aligned} S_j &= m \int_t^{1-t} f(p) \left[p(1-(1-p)^c) g(1,1,p) + (1-p)(1-p^c) g(0,0,p) \right. \\ &\quad \left. + \; p(1-p)^c g(0,1,p) + (1-p)p^c g(1,0,p) \right] dp \\ &= 2m\kappa(t) \left((1-2t) - 2\frac{(1-t)^{c+1} - t^{c+1}}{c+1} \right) \approx 2m\kappa(t)(1 - \frac{2}{c+1}) \end{aligned} \tag{2}$$

In comparison, the colluders have scores equalling $2m\kappa(t)c^{-1}$ in expectation. This means that with only $c = 3$ accomplices, the score of Buyer j is bigger than the ones of the colluders, which are bigger than Z if the code is long enough to face a collusion of size 3 (depending on the parameters (n, ϵ)). The Provider sends $(\mathbf{X}_j, \mathbf{Y}, \mathbf{p}, Z)$ to the Judge as an evidence to accuse Buyer j. This attack is just an example, there certainly exists a better way to frame an innocent.

The Provider Knows the Bias Vector p. The previous two scenarios demonstrate that the Provider must not know the fingerprints of the Buyers, if the Buyers are to be protected. This is well known in the literature of asymmetric fingerprinting. However, another threat occurs when dealing with Tardos codes. In this scenario, the Provider has no knowledge of the Buyer's fingerprint, nor the underlying watermark method. We therefore assume that the Provider cannot forge an unauthorized copy, either on his/her own or with accomplices. On receipt of a pirated copy, the sequence is extracted by the trusted Technology Provider. Given the extracted sequence $\mathbf{Y}$, the scores of all Buyers are computed using Equation (1). It is here that the Provider can lie, since the probabilities in $\mathbf{p}$ are only known by the Provider.

Specifically, an untrustworthy Provider can create a fake vector of probabilities $\hat{\mathbf{p}}$ that implicates Buyer j. However, the distribution $f(p)$ is publicly known, so the question becomes how to generate a $\hat{\mathbf{p}}$ that (i) implicates Buyer j, and (ii) has an arbitrarily high probability of been drawn from the distribution $f(p)$?

The following method shows that it is simple to do so. However, we do not claim that this attack is unique or optimal. Let us focus on a column where $p_i = p$ and $Y_i = X_{j,i}$. The true summand in Equation (1) is $g(1,1,p)$ or $g(0,0,p)$ (with equal probability). Suppose that the content provider replaces the secret value p by a fake secret $\hat{p}$ which is drawn independently according to f. On average, this summand takes the new value:

$$\Delta(t) = \int_t^{1-t} f(\hat{p}) \frac{g(1,1,\hat{p}) + g(0,0,\hat{p})}{2} d\hat{p} = \kappa(t) \ln \frac{1-t}{t}.$$

For a cutoff $t = 1/900$ (recommended by G. Tardos to fight against 3 colluders), $\kappa(t) \approx \pi^{-1}$ and the numerical value is surprisingly high: $\Delta(1/900) \approx 2.16$. Suppose now that the content provider applies the same strategy on an index i where $Y_i \neq X_{j,i}$. Then the expectation is the opposite. However, in a Tardos code, even for an innocent Buyer j, the proportion α of indices where symbols Y_i and $X_{j,i}$ agree is above $1/2$ for common collusion strategies. For instance, with an interleaving collusion attack [18], $\alpha = 3/4$ whatever the collusion size c.

Based on this fact, we propose the following attack. The Provider computes the score for all Buyers, which on average equals 0 for innocent Buyers and $2m\kappa(t)c^{-1}$ for the colluders [18]. The Provider initializes $\hat{\mathbf{p}} = \mathbf{p}$. Then, he/she randomly selects a column i and randomly draws a fake secret $\hat{p_i} \sim f$. He/She re-computes the score of Buyer j with this fake secret and iterates selecting a different column until S_j is above the threshold Z. On average, $m(c\kappa(t)^{-1}\Delta(t)(\alpha - 1/2))^{-1}$ secret values p_i need to be changed in this way, e.g. only 20% of the code length if the copy has been made using an interleaving attack.

Figure 1 illustrates this attack for the case where the code length is $m = 1000$ and the number of colluders is $c = 3$. The solid coloured lines depict the accusation scores of 10 randomly selected innocent buyers. We observe that after 20 to 30% of the elements of $\mathbf{p}$ have been altered, the accusation scores of the innocent Buyers exceed the *original* scores of the colluders. In fact, the colluders' accusation scores also increase. However, we are not concerned by the highest score, but rather by the fact that the Provider is able to exhibit a couple $(\hat{\mathbf{p}}, \mathbf{X}_j)$ such that $S_j > Z$. Thus, it is sufficient to raise the score of the innocent Buyer, even if this raises all other Buyers' scores as well.

Randomly selecting some p_i's (independently from $\mathbf{X}_j$ and $\mathbf{Y}$) and re-drawing them according to the same law ensures that $\hat{p_i} \sim f$, $\forall i$. Therefore, the Judge observing $\hat{\mathbf{p}}$ cannot distinguish the forgery. For this reason, the Judge might request to see the matrix $\mathbf{X}$ to statistically test whether the elements of $\mathbf{X}$ are drawn from the distribution $\hat{\mathbf{p}}$. In this case, the Provider can give a fake matrix $\hat{\mathbf{X}}$ where the columns whose p_i have been modified are re-drawn such that $\mathbb{P}(X_{ki} = 1) = \hat{p_i}$, $\forall k \neq j$. The only way to prevent this deception would be for the Judge to randomly asked an innocent Buyer $k \neq j$ for his copy in order to verify the authenticity of $\hat{\mathbf{X}}$. This latter step seems somewhat odd. We arrive at the strange situation where the Judge has to contact innocent buyers when Buyer j is accused.

3 An Asymmetric Tardos Code Construction

The previous section underlines the difficulty of constructing an asymmetric fingerprinting protocol using Tardos codes. The constraints are:

- The Provider should not know the fingerprints.
- The Provider should not change the secret $\mathbf{p}$ used for the code construction during the accusation score computation.
- The Buyer should know neither the secret $\mathbf{p}$ nor the fingerprint of any other user.

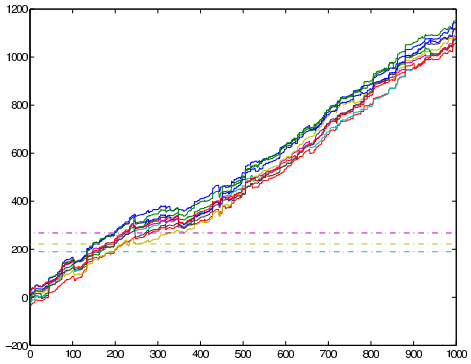


Fig. 1. Accusation score as a function of the number of changed elements of the vector $\mathbf{p}$ for the case where $m = 1000$ and $c = 3$. The solid coloured lines show how the accusation scores of 10 randomly selected innocent buyers increases. The dotted horizontal lines show the original scores for the colluders before the modification.

- His fingerprint must be drawn according to the statistical distribution induced by $\mathbf{p}$.
- The Buyer should not be able to modify his fingerprint.

These constraints prevent the application of previous asymmetric fingerprinting schemes to a Tardos code. This section proposes a solution to this problem, which consists of two phases: the generation of the fingerprint and the disclosure of a halfword. Both phases rely on a primitive which we present first.

3.1 Pick a Card, Any Card!

What we need is a scheme that enables a receiver $\mathbf{R}$ to pick k elements at random in a list of N elements provided by a sender $\mathbf{S}$, in such a way that:

1. $\mathbf{R}$ gets elements that belong to the list;
2. $\mathbf{R}$ does not get any information on the elements he did not pick;
3. $\mathbf{S}$ does not know which elements have been picked.

Functionally speaking, this is precisely what is called *Oblivious Transfer* by cryptographers. A k-out-of-N Oblivious Transfer protocol is denoted by OT_k^N. In the literature we can find OT_1^2, OT_1^N and OT_k^N protocols. When $k \geq 1$, if the k elements are picked one-by-one adaptively, we speak of *adaptive OT protocols*, denoted by $OT_{k\times 1}^N$; if they are picked simultaneously, we speak of *non-adaptive OT protocols*, simply denoted OT_k^N .

Technically speaking, the oblivious transfer problem has been independently tackled by two communities. First, Cryptographers have been working on it since 1981. We will refer to this quite long and mature framework as "traditional" OT. Second, in 2001 other researchers proposed a different approach based on *Commutative Encryption* and *Two-lock Cryptosystems*. Both are considered and

discussed in Sec. 4, according to their respective advantages. We provide more details on the use of OT protocols based on Commutative Encryption or Two-lock crypto-systems, as they are less known but particularly interesting in our case.

3.2 Phase 1: Generation of the Fingerprint

Fingerprint generation consists of two steps. During Step 1, the Provider generates lists from the secret $\mathbf{p}$, and commits them in order to avoid any *a posteriori* cheating. During Step 2, the Buyer picks elements in the lists to generate his own fingerprint. This step is addressed by oblivious transfer protocols.

Step 1. We use the commutative encryption protocol m times to generate the fingerprint of the j-th Buyer $\mathbf{X}_j = (X_{j,1}, \ldots, X_{j,m})$. $\mathbf{S}$ is the Provider, and $\mathbf{R}$ is Buyer j. The Provider generates a secret vector $\mathbf{p}$ for a Tardos code. Each p_i is quantized such that $p_i = L_i/N$ with $L_i \in [N-1]$.

For a given index i, the objects are the concatenation of a binary symbol and a text string. There are only two versions of an object in list $\mathcal{C}_i$. For L_i objects, $O_{k,i} = (1\|\mathtt{ref}_{1,i})$, and $O_{k,i} = (0\|\mathtt{ref}_{0,i})$ for the $N - L_i$ remaining ones. The use of the text strings $\{\mathtt{ref}_{X,i}\}$ depends on the content distribution mode as detailed in Sec. 4.3. The object $O_{k,i}$ is committed with key $K_{k,i}$ and stored in the list $\mathcal{C}_i = \{C_{k,i}\}_{k=1}^N$. There are thus as many different lists $\mathcal{C}_i$ as the length m of the fingerprint. These lists are the same for all buyers, and are published in a public Write Once Read Many (WORM) directory [**?**] whose access is granted to all users. As the name, nobody can modify or erase what is initially written in a WORM directory, but anyone can read from it.

$$\begin{array}{lcll} p_1 & \overset{Quantize}{\longrightarrow} & (0\|\mathtt{ref}_{0,1}, 1\|\mathtt{ref}_{1,1}, \ldots, 1\|\mathtt{ref}_{1,1}) & \overset{Commit}{\longrightarrow} \quad \mathcal{C}_1 = (C_{1,1}, \ldots, C_{N,1}) \\ p_2 & \longrightarrow & (0\|\mathtt{ref}_{0,2}, 0\|\mathtt{ref}_{0,2}, \ldots, 0\|\mathtt{ref}_{0,2}) & \longrightarrow \quad \mathcal{C}_2 = (C_{1,2}, \ldots, C_{N,2}) \\ & & \vdots & \\ p_m & \longrightarrow & (1\|\mathtt{ref}_{1,m}, 0\|\mathtt{ref}_{0,m}, \ldots, 1\|\mathtt{ref}_{1,m}) & \longrightarrow \quad \mathcal{C}_m = (C_{1,m}, \ldots, C_{N,m}) \end{array}$$

Fig. 2. The lists $\mathcal{C}_i = \{C_{k,i}\}_{k=1}^N$ are stored in a WORM

Step 2. If we use a traditional Oblivious Transfer protocol, the Buyer and Provider run it to get the corresponding key $K_{\mathsf{ind}(j,i),i}$: the Provider proposes the list of the keys $\{\pi_j(k)\|K_{\pi_j(k),i}\}$ and the Buyer picks one with an OT_1^N. This key allows him to open one of the commitments $C_{\pi_j(k),i}$. Provider and Buyer will have to keep in a log file some elements of the exchange in order to run the Phase 2. It is specific to the *OT* protocol and we have not studied this problem in detail.

Let us now describe how to solve the problem with a Commutative Encryption scheme. Contrary to the $\mathcal{C}$-lists, the $\mathcal{D}$-lists are made specific to a given Buyer

j. The Provider picks a secret key S_j and a permutation $\pi_j(.)$ over $[N]$. The Buyer is given a list $\mathcal{D}_{j,i} = \{D_{j,i,k} = \mathbf{CE}(S_j, (\pi_j(k)\|K_{\pi_j(k),i}))\}_{k=1}^{N}$. Therefore, the lists $\{\mathcal{C}_i\}_{i=1}^{m}$ are common for all users, whereas the lists $\{\mathcal{D}_{j,i}\}_{i=1}^{m}$ are specific to Buyer j. We have introduced here a slight change with respect to protocol 4.1, i.e. the permutation π_j whose role is explained below. Buyer j chooses one object in the list, say the $k(j,i)$-th object. He/she sends the corresponding ciphertext $U_{k(j,i),i} = \mathbf{CE}(R_{j,i}, D_{j,i,k(j,i)})$ decrypted by the provider with S_j and sent back to the Buyer who, at the end, gets the index $\mathsf{ind}(j,i) = \pi_j(k(j,i))$ and the key $K_{\mathsf{ind}(j,i),i}$, which grants him/her the access to the object $O_{\mathsf{ind}(j,i),i}$, stored in encrypted form in the WORM. It contains the symbol $b_{\mathsf{ind}(j,i),i}$. This becomes the value of the i-th bit of his/her fingerprint, $X_{j,i} = b_{\mathsf{ind}(i,j),i}$, which equals '1' with probability p_i. The provider keeps in a log file the values of S_j and $U_{k(j,i),i}$, and the user keeps $R_{j,i}$ in his/her records.

	P		B
1.	**Commit** $C_{\pi_j(k),i} = \mathbf{Com}(K_{\pi_j(k),i}, O_{\pi_j(k),i})$ Only one time for all users and all bits		
2.	$D_{j,i,k} = \mathbf{CE}(S_j, (\pi_j(k)\|K_{\pi_j(k),i}))$		
3.		$\longleftarrow$	$U_{k(j,i),i} = \mathbf{CE}(R_{j,i}, D_{j,i,k(j,i)})$
4.	$V = \mathbf{CE}^{-1}(S_j, U_{k(j,i),i})$	$\longrightarrow$	
5.			$\pi_j(k)\|K_{\pi_j(k),i} = \mathbf{CE}^{-1}(R_{j,i}, V)$
6.			$O_{\pi_j(k),i} = \mathbf{E}^{-1}(K_{\pi_j(k),i}, C_{\pi_j(k),i})$

Fig. 3. Generation of a fingerprint bit using the Commutative Encryption Scheme

3.3 Phase 2: Disclosure of the Halfword

The accusation process detailed in Sec. 4.4 allows the Provider to list a set of suspected users to be forwarded to the judge for verification. After phase 1 is completed, the Provider orders Buyer j to reveal $m_h < m$ bits of his fingerprint. These disclosed symbols compose the so-called halfword [16]. The following facts must be enforced: Buyer j does not know which bits of his/her fingerprint are disclosed even if the Provider asks for the same bit indices to all the users. The Provider discloses m_h bits of the fingerprints without revealing any knowledge about the others. Of course, Buyer j refuses to follow the protocol for more than m_h objects.

Commutative Encryption. Again, we propose to use the double-blind random selection protocol of Sec. 3.1. Now, Buyer j plays the role of **S**, and the Provider the role of **R**, $N = m$, and object $O_i = (R_{i,j}\|\mathtt{alea}_{i,j})$. These items are the m secret keys selected by Buyer j during phase 1 (Sec. 3.2) concatenated with random strings $\mathtt{alea}_{i,j}$ to be created by Buyer j. This alea finds its use during the personalization of the content (see Sec. 4.3). Following the protocol, the Provider selects m_h such object. The decryption of message $U_{k(i,j),j}$ received during phase 1 thanks to the disclosure of the key $R_{i,j}$ yields $D_{i,j,k(i,j)}$ which

in turn is decrypted with key S_j, provides the index of the selected object, otherwise the protocol stops. This prevents a colluder from denying the symbol of his fingerprint and from copying the symbol of an accomplice. At the end, the Provider learns which item was picked by Buyer j at index i. Therefore, he/she ends up with m_h couples $(X_{j,i}, \texttt{alea}_{k(i,j),i})$ associated to a given Buyer j.

Generic Oblivious Transfer protocols. At phase 2, any $OT^N_{k\times 1}$ can be used to allow the Provider to get m_h objects from the list of the $O_i = (R_{i,j} \| \texttt{alea}_{i,j})$ owned by the Buyer. The problem is if another OT scheme was used at the precedent step, there is no such things as the $R_{i,j}$ values. In order to prevent the Buyer from denying the symbol of his fingerprint, the $R_{i,j}$ values have to be replaced by a number which was part of the exchange during the generation of the fingerprint. This element is specific to the OT protocol.

4 Implementation Details

The previous section has detailed the core of our scheme which is the construction of the codewords based on oblivious transfer. This section deals with the details of this primitive and the remaining elements, namely the watermarking of video content, the distribution and the accusation process.

4.1 Details of the Oblivious Transfer Protocol

This protocol can be implemented by two approaches, 'classical' Oblivious transfer and Commutative encryption, which have been studied with different security models. Both are interesting for us, and we will now summarize them and discuss their usefulness

Traditional Oblivious Transfer protocols. Oblivious Transfer Protocols have been introduced by cryptographers in [17] and led to a huge number of papers in the cryptographic community, *e.g.* [13,5,10]. These protocols are studied in the same framework as multi-party computation. Their security is studied under different models below, listed from the weakest to the strongest: honest-but-curious model (where no one cheats during the protocol execution), half simulation (introduced by [14], cheating sender or cheating receiver studied separately; local security study), full simulation (introduced in [3], studying cheating sender and receiver globally; global security study). In addition, the UC (Universally Composable) model has been introduced in [4] to study the behavior and security of protocols that are based on concurrent and composable cryptographic primitives.

Oblivious Transfer based on Commutative Encryption. An encryption primitive **CE** is said to be a *Commutative Encryption* if for any two keys k_R and k_S and any plaintext m, we have (usual definition in the literature)

$$\mathbf{CE}(k_R, \mathbf{CE}(k_S, m)) = \mathbf{CE}(k_S, \mathbf{CE}(k_R, m)). \tag{3}$$

Based on such a primitive, a *Commutative Encryption Scheme (CES)* can be defined as follows [1].

1. Let m_1, m_2, ...,m_N be the N inputs of the Sender **S**. **S** chooses N secret keys K_1, K_2, ..., K_N for a symmetric cryptosystem **E** (*e.g.* AES, DES) and a key k_S for the commutative encryption primitive **CE**. **S** provides

$$\begin{aligned} C_1 &= \mathbf{E}(K_1, m_1)\ , & D_1 &= \mathbf{CE}(k_S, K_1) \\ C_2 &= \mathbf{E}(K_2, m_2)\ , & D_2 &= \mathbf{CE}(k_S, K_2) \\ &\dots & &\dots \\ C_N &= \mathbf{E}(K_N, m_N)\ , & D_N &= \mathbf{CE}(k_S, K_N) \end{aligned}$$

   Note that the couples $\langle C_j, D_j \rangle$ can be publicly accessed.

2. Now, let us assume that the receiver **R** wants to pick the i-th element of the list. **R** loads $\langle C_i, D_i \rangle$ and chooses a secret key k_R for **CE**. He encrypts D_i with it and sends the result $U = \mathbf{CE}(k_R, D_i)$ to **S**.
3. **S** decrypts U with S and sends $W = \mathbf{CE}^{-1}(k_S, U)$ to **R**. **R** computes K_i, and can get to $m_i = \mathbf{E}^{-1}(K_i, C_i)$.

A *Two-lock Cryptosystem* is a variant that uses two different primitives **CE1** and **CE2** instead of **CE**:

$$\mathbf{CE1}(k_R, \mathbf{CE2}(k_S, m)) = \mathbf{CE2}(k_S, \mathbf{CE1}(k_R, m)). \tag{4}$$

Both approaches are interesting for us, as we will discuss now. First of all, the security of Oblivious Transfer Protocols has been much stronger studied than the one of the Commutative Encryption Schemes. Hence, we will use them each time it is possible, leaning on well known protocols.

But, at some steps of the protocol we prefer to use Commutative Encryption Schemes, as its structure fits really well to our purpose. It is for example the case during fingerprint generation, as we also want the Provider to commit on the lists elements, which correspond to the secret vector Tardos accusation will rely on. This ensures that the same secret vector will be used during the accusation process. Such commitments are easily included in a Commutative Encryption Scheme, it is more difficult in a traditional Oblivious Transfer protocol. In addition, we use some elements exchanged during the course of the protocol in phase 1 (Sec. 3.2) to ensure the correct conduct of the Phase 2 (Sec. 3.3).

Designing the right Commutative Encryption Scheme is not so easy, as the literature does not provide us a scheme that fulfill our requirements. First of all, notice that using a symmetric or asymmetric encryption primitive as **CE**, or in the variant scheme **CE1** and **CE2**, does not matter here, functionally speaking, as encryption and decryption will be performed by the same person. Hence, only security and eventually efficiency may guide our choice. Of course, we would like to use the most secure encryption primitives. The highest security level, *unconditional security* is only reached by the One-Time Pad, and cannot be achieved here because it would require to use a different key for each encryption whereas here the same key k_S is used to encrypt all the keys K_i. Hence, *semantic security* is the best security class we might achieve [9,20,7]. Moreover, semantic security

is necessary in our case, because we have to encrypt binary symbols and do not want the Receiver to be able to distinguish encrypted 0's from encrypted 1's during both the fingerprint generation or the halfword disclosure steps. This implies the use of a probabilistic encryption scheme. Unfortunately, semantic security has not yet been tackled in the Commutative Encryption literature [1,11,21]. Nevertheless, semantic security should be achieved in a near future, making this kind of OT particularly interesting for us.

Concerning the variant called Two-lock Cryptosystem, a few implementations have been proposed: a first one based on the Knapsack problem [21], which has been broken [22], a second one based on the discrete logarithm problem [21], and a third one based on RSA [11]. None of them achieve semantic security at the moment.

4.2 Watermarking

A nowadays trend is the application of fingerprinting to premium video contents. Premium means movies in very high quality available for home cinema shortly after their release in theaters. Personalization of the copies are usually done as follows: Before distribution, the content is divided into sequential blocks (e.g. Group of Pictures of few seconds of a video). Offline, a robust watermarking technique creates two versions of some blocks embedding the symbol '0' and respectively '1'. This is done by the Technology Provider. Quality is very important for premium movies and watermarking under that constraint involves a lot of processing. This motivates this offline preprocessing.

In some scenarios (screeners for jurys, marketing, blu-ray discs, premium downloads), the physical medium storage or bandwitth is so large that both versions of the blocks are encrypted and transmitted to the software client or the device of the Buyers. This latter is trusted and the strings $\{\mathtt{ref}_{X,i}\}$ it got from phase 1 are parameters needed to get access to the i-th block watermarked with symbol X.

4.3 Content Personalization at the Server Side

As for Video On Demand where the client is not trusted, personalization of the content is usually made at the server side, which raises an issue since the Provider doesn't know user fingerprints. There exist Buyer-Seller protocols for embedding a sequence $\mathbf{X}_j$ into a content c_o without disclosing $\mathbf{X}_j$ to the Seller and c_o to the Buyer. They are based on homomorphic encryption scheme and work with some specific implementations of spread spectrum [12] or Quantization Index Modulation watermarking [6]. In other words, not any watermarking technique can be used, and this is not the route we have chosen so far. Due to space limitations, a brief sketch of the adaptation of [6] is presented hereafter.

Let $\mathbf{c}_i^{(0)} = (c_{i,1}^{(0)}, \ldots, c_{i,Q}^{(0)})$ be the Q quantized components (like pixels, DCT coefficients, portion of streams etc) of the i-th content block watermarked with symbol '0' (resp. $\mathbf{c}_i^{(1)}$ with symbol '1'). Denote $\mathbf{d}_i = \mathbf{c}_i^{(1)} - \mathbf{c}_i^{(0)}$. Assume as in

[6, Sect. 5], an additive homomorphic and probabilistic encryption $E[.]$ such as the Pallier cryptosystem. Buyer j has a pair of public/private keys (pk_j, sk_j) and sends $(E_{pk_j}[X_{j,1}], \ldots, E_{pk_j}[X_{j,m}])$. The provider sends him/her the ciphers

$$E_{pk_j}[c_{i,\ell}^{(0)}].E_{pk_j}[X_{j,i}]^{d_{i,\ell}}, \forall (i,\ell) \in [m] \times [Q]. \tag{5}$$

Thanks to the homomorphism, Buyer j decrypts this with sk_j into $c_{i,\ell}^{(0)}$ if $X_{j,i} = 0$, $c_{i,\ell}^{(1)}$ if $X_{j,i} = 1$. Since $X_{j,i}$ is constant for the Q components of the i-th block, a lot of bandwidth and computer power will be saved with a composite signal representation as detailed in [6, Sect. 3.2.2].

A crucial step in this kind of Buyer-Seller protocols is to prove to the Provider that what is sent by the Buyer is indeed the encryption of bits, and moreover bits of the Buyer's fingerprint. This usually involves complex zero-knowledge subprotocols [12,6]. Here, we avoid this complexity by taking advantage of the fact that the Provider already knows some bits of the fingerprint $\mathbf{X}_j$, i.e. those belonging to the halfword (see Sec. 3.3), and the Buyers do not know the indices of these bits. Therefore, in $m_v < m_h$ random indices of the halfword, the Provider asks Buyer j to open his/her commitment. For one such index i_v, Buyer j reveals the random value r_{i_v} of the probabilistic Pallier encryption (with the notation of [6]). The Provider computes $g^{X_{j,i_v}} h^{r_{i_v}} \mod N$ and verifies it equals the i_v-th cipher, which Buyer j pretended to be $E_{pk_j}[X_{j,i}]$.

One drawback of this simple verification scheme is that the Buyer discovers m_v indices of the halfword. This may give rise to more elaborated collusion attacks. For example, Buyer j, as a colluder, could try to enforce $Y_{i_v} \neq X_{j,i_v}$ when attempting to forge a pirated copy. Further discussion of this is beyond the scope of this paper.

This approach may also introduce a threat to the Buyer. An untrustworthy Provider can ask to open the commitments of non-halfword bits in order to disclose bits he/she is not supposed to know. For this reason, the Provider needs to send $\texttt{alea}_{k(i_v,j),i_v}$ as defined in Sec. 3.3 to show Buyer j that his/her verification duly occurs on a halfword bit.

4.4 The Accusation Procedure

The accusation is straightforward and similar to other fingerprinting protocols. A Scouting Agency is in charge of catching a forgery. The Technology Provider decodes the watermark and extracts sequence $\mathbf{Y}$ from the pirated content. The Provider computes the halfscores by applying Eq. (1) only on the halfwords. This produces a list of suspects, e.g. those users whose score is above a threshold, or those users with the highest scores.

Of course, this list cannot be trusted, since the Provider may be untrustworthy. The list is therefore sent to a third party, referred to as the Judge, who first verifies the computation of the halfscores. If different values are found, the Provider is black-listed. Otherwise, the Judge computes the scores of the full fingerprint.

To do so, the Judge needs the secret $\mathbf{p}$: he/she asks the Provider for the keys $\{K_{k,i}\}$, $\forall(k,i) \in [N] \times [m]$ and thereby obtains from the WORM all the objects $\{O_{k,i}\}$, and the true values of $(p_1, \ldots, p_m)$. The Judge must also request suspected Buyer j for the keys $R_{j,i}$ in order to decrypt the messages $U_{k(j,i),i}$ in $D_{i,j,k(i,j)}$ which reveal which object Buyer j picked during the i-th round of Sec. 3.2 and whence $X_{j,i}$. Finally, the Judge accuses the user whose score over the full length fingerprint is above a given threshold (related to a probability of false alarm).

5 Discussion

5.1 Security

Suppose first that the Provider is honest and denote by c the collusion size. A reliable tracing capability on the halfwords is needed to avoid false alarms. Therefore, as proven by G. Tardos, $m_h = O(c^2 \log n \epsilon^{-1})$, where ϵ is the probability of suspecting some innocent Buyers. Moreover, successful collusions are avoided if there are secret values such that $p_i < c^{-1}$ or $p_i > 1 - c^{-1}$(see [8]). Therefore, N should be sufficiently big, around a hundred, to resist against collusion of size of some tens. During the generation of the fingerprint in Sec. 3.2, permutation $\pi_j(.)$ makes sure that Buyer j randomly picks up a bit '1' with probability $p_i = L_i/N$ as needed in the Tardos code. In particular, a colluder cannot benefit from the discoveries made by his accomplices.

We now analyze why colluders would cheat during the watermarking of their version of the Work described in Sec. 4.3. By comparing their fingerprints, they see indices where they all have the same symbols, be it '0' or '1'. As explained in the introduction, they won't be able to alter those bits in the tampered fingerprint except if they cheat during the watermarking: If their fingerprint bits at index i all equal '1', one of them must pretend he/she has a '0' in this position. If they succeed to do so for all these positions, they will able to forge a pirated copy with a null fingerprint for instance.

How many times do the colluders need to cheat? With probability p_i^c (resp. $(1-p_i)^c$), they all have bit '1' (resp. '0') at index i. Thus, there are on average $m_c(c) = m \int_t^{1-t} (p^c + (1-p)^c) f(p) dp$ such indices. The Provider asks for a bit verification with probability m_v/m_h. The probability of a successful attack for a collusion of size c is therefore $(1 - m_v/m_h)^{m_c(c)}$. Our numerical simulations (see figure 4 (a)) show that m_v shouldn't be more than 50 bits for typical code length and collusion size below a hundred. Thus, m_v is well below m_h.

Suppose now that the Provider is dishonest. The fact that the m lists $\mathcal{C}_i$, $\forall i \in [m]$ are public and not modifiable prevents the Provider from altering them for a specific Buyer in order to frame him/her afterwards. Moreover, it will raise the Judge's suspicion if the empirical distribution of the p_i is not close to the pdf f. Yet, biases can be introduced on the probabilities for the symbols of the colluders' fingerprint only if there is a coalition between them and the untrustworthy Provider. For instance, the Provider can choose a permutation such that by selecting the first item (resp. the last one) in the list $\mathcal{D}_{j,i}$ an accomplice colluder

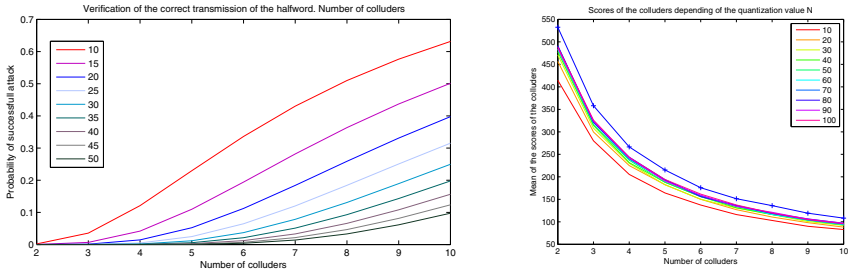


Fig. 4. (a) m_v goes from 10 to 50 by 5, $m = 3000$ and $m_h = 1500$. (b) N goes from 10 to 100 and $m = 1500$. The gray curve with crosses is for the unquantified Tardos code.

is sure to pick up a symbol '1' (resp. '0'). This ruins the tracing property of the code, but this does not allow the Provider to frame an innocent. First, it is guaranteed that $\mathbf{p}$ used in Eq. (1) is the one which generated the code. Second, the Provider and his accomplices colluders must ignore a significant part of the fingerprints of innocent Buyers. To this end, $m - m_h$ must also be in order of $O(c^2 \log n\epsilon^{-1})$. If this holds, the Judge is able to take a reliable decision while discarding the halfword part of the fingerprint. Consequently, $m \approx 2m_h$, our protocol has doubled the typical code length, which is still in $O(c^2 \log n\epsilon^{-1})$.

5.2 Efficiency

Parameters. The parameters of the Tardos code are chosen according to the formulas linking length, number of colluders, and number of users. We have found out that the value m_v doesn't need to be more than 50, see Sec. 4. We consider the value N, the quantization parameter, with the interleaving collusion attack. In the figure 4 (b), we can see that up to a small value of N (around 20), there is no gain of efficiency. The red line shows that the results with the unquantized Tardos parameters remain better.

Complexity. The cost of phase 1 is $m \times N$ commitments for the lists that will be stored in the Worm file, and $mn \times (N+4)$ exponentiations for the OT phase. Regarding the use of a non specific OT, still $m \times N$ commitments, plus the cost of mn 1-out-of-N Oblivious Transfers. This cost depends of course of the chosen protocol, it is in $O(N)$ for a lot of protocols. For Phase 2, the cost is that of an m_h-out-of-m Oblivious transfer. If this OT is performed with the use of a *Commutative Encryption*, the cost is $2m+4m_h$ for the communication, and $4m_h$ rounds, for another OT scheme, the communication is in $O(m)$ and the number of rounds depends of the protocol, it is usually in $O(m_h)$.

6 Conclusion

Tardos codes are currently the state-of-the-art in collusion-resistant fingerprinting. However, the previous asymmetric fingerprint protocols cannot be applied to this particular construction. There are mainly two difficulties. First, the Buyer has to generate his/her secret fingerprint but according to vector $\mathbf{p}$, which is kept secret by the Provider. Second, the secret $\mathbf{p}$ used in the accusation process must be the same as the one which generated the fingerprints.

We have proposed the first asymmetric fingerprinting protocol dedicated to Tardos codes. The construction of the fingerprints and their embedding within pieces of Work do not need a trusted third party. Note, however, that during the accusation stage, a trusted third party is necessary like in any asymmetric fingerprinting scheme we are aware of. Further work is needed to determine if such a third party can be eliminated. In particular, we anticipate that some form of secure multi-party computation can be applied.

We considered two forms of oblivious transfer protocols, the first based on traditional cryptographic techniques and the second based on less well known Commutative Encryption or Two-Lock crypto-systems. These latter techniques are less mature than traditional Oblivious Transfer protocols in terms of security, but offers interesting properties that are convenient to our application. Further work is needed to improve their semantic security, so that their advantages do not come at the cost of decreased security.

Acknowledgement. We would like to thank Boris Škorić, and the three anonymous reviewers for their useful comments, which helped to improve the presentation of our results.

References

1. Bao, F., Deng, R.H., Feng, P.: An efficient and practical scheme for privacy protection in the E-commerce of digital goods. In: Won, D. (ed.) ICISC 2000. LNCS, vol. 2015, pp. 162–170. Springer, Heidelberg (2001)
2. Boneh, D., Shaw, J.: Collusion-secure fingerprinting for digital data. IEEE Trans. Inform. Theory (1998)
3. Camenisch, J., Neven, G., Shelat, A.: Simulatable adaptive oblivious transfer. In: Naor, M. (ed.) EUROCRYPT 2007. LNCS, vol. 4515, pp. 573–590. Springer, Heidelberg (2007)
4. Canetti, R.: Universally composable security: A new paradigm for cryptographic protocols. In: 42nd IEEE Symposium on Foundations of Computer Science, pp. 136–145. IEEE, Los Alamitos (2002)
5. Chu, C., Tzeng, W.: Efficient k-out-of-n oblivious transfer schemes with adaptive and non-adaptive queries. In: Vaudenay, S. (ed.) PKC 2005. LNCS, vol. 3386, pp. 172–183. Springer, Heidelberg (2005)
6. Deng, M., Bianchi, T., Piva, A., Preneel, B.: An efficient Buyer-Seller watermarking protocol based on composite signal representation. In: ACM MM&Sec 2009, pp. 9–18 (2009)

7. Fontaine, C., Galand, F.: A survey of homomorphic encryption for nonspecialists. EURASIP Journal on Information Security 15 (2007)
8. Furon, T., Pérez-Freire, L.: Worst case attack against binary probabilistic traitor tracing codes. In: IEEE WIFS 2009, pp. 46–50 (2009)
9. Goldreich, O.: Foundations of cryptography: Basic applications. Cambridge Univ. Pr., Cambridge (2004)
10. Green, M., Hohenberger, S.: Blind identity-based encryption and simulatable oblivious transfer. In: Kurosawa, K. (ed.) ASIACRYPT 2007. LNCS, vol. 4833, pp. 265–282. Springer, Heidelberg (2007)
11. Huang, H., Chang, C.: A new design for efficient t-out-n oblivious transfer scheme (2005)
12. Kuribayashi, M.: On the Implementation of Spread Spectrum Fingerprinting in Asymmetric Cryptographic Protocol. EURASIP Journal on Inf. Security (2010)
13. Naor, M., Pinkas, B.: Oblivious transfer with adaptive queries. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, p. 791. Springer, Heidelberg (1999)
14. Naor, M., Pinkas, B.: Computationally secure oblivious transfer. Journal of Cryptology 18(1), 1–35 (2005)
15. Oprea, A., Bowers, K.D.: Authentic time-stamps for archival storage. In: Backes, M., Ning, P. (eds.) ESORICS 2009. LNCS, vol. 5789, pp. 136–151. Springer, Heidelberg (2009)
16. Pfitzmann, B., Schunter, M.: Asymmetric fingerprinting. In: Maurer, U.M. (ed.) EUROCRYPT 1996. LNCS, vol. 1070, pp. 84–95. Springer, Heidelberg (1996)
17. Rabin, M.: How to exchange secrets by oblivious transfer. Tech. rep., Technical Report TR-81, Harvard Aiken Computation Laboratory (1981)
18. Skoric, B., Katzenbeisser, S., Celik, M.: Symmetric Tardos fingerprinting codes for arbitrary alphabet sizes. Designs, Codes and Cryptography 46(2), 137–166 (2008)
19. Tardos, G.: Optimal probabilistic fingerprint codes. In: STOC 2003, pp. 116–125. ACM, New York (2003), http://www.renyi.hu/~tardos/publications.html
20. van Tilborg, H.: Encyclopedia of cryptography and security. Springer, Heidelberg (2005)
21. Wu, Q., Zhang, J., Wang, Y.: Practical t-out-n oblivious transfer and its applications. In: Qing, S., Gollmann, D., Zhou, J. (eds.) ICICS 2003. LNCS, vol. 2836, pp. 226–237. Springer, Heidelberg (2003)
22. Zhang, B., Wu, H., Feng, D., Bao, F.: Cryptanalysis of a knapsack based two-lock cryptosystem. In: Jakobsson, M., Yung, M., Zhou, J. (eds.) ACNS 2004. LNCS, vol. 3089, pp. 303–309. Springer, Heidelberg (2004)

1494 IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 47, NO. 4, MAY 2001

On Cryptographic Properties of the Cosets of $R(1, m)$

Anne Canteaut, Claude Carlet, Pascale Charpin, and Caroline Fontaine

***Abstract*—We introduce a new approach for the study of weight distributions of cosets of the Reed–Muller code of order 1. Our approach is based on the method introduced by Kasami in [1], using Pless identities. By interpreting some equations, we obtain a necessary condition for a coset to have a "high" minimum weight. Most notably, we are able to distinguish such cosets which have three weights only. We then apply our results to the problem of the nonlinearity of Boolean functions. We particularly study the links between this criterion and the propagation characteristics of a function.**

***Index Terms*—Boolean function, derivation, nonlinearity, propagation criterion, Reed–Muller codes.**

Main Notation

- $\mathcal{B}_m$ is the set of Boolean functions of m variables;
- Ω_f, $f \in \mathcal{B}_m$ is the codeword of length 2^m equal to the ordered list of all values of f; $\mathbf{0}$ and $\mathbf{1}$ denote, respectively, the zero codeword and the all-one codeword;
- $x \cdot y$ denotes the usual dot product between two vectors x and y;
- $V^{\perp}$ denotes the dual of a subspace $V \subset \boldsymbol{F}_2^m$, i.e.,

$$V^{\perp} = \{x \in \boldsymbol{F}_2^m \,|\, \forall y \in V, x \cdot y = 0\};$$

- $\{\varphi_\alpha | \alpha \in \boldsymbol{F}_2^m\}$ is the subset of $\mathcal{B}_m$ consisting of all linear functions

$$\varphi_\alpha \colon x \mapsto \alpha \cdot x;$$

- H_α denotes the kernel of φ_α;
- $\mathcal{F}(f)$, $\mathcal{L}(f)$, $\mathcal{MD}(f)$, $\mathcal{V}(f)$, and $\mathcal{N}(f)$ are, respectively, defined by (1), Definition II.1, (5), and (6);
- $\boldsymbol{F}_q$ is the finite field of order q;
- $\mathcal{A}$ is the group algebra $\boldsymbol{F}_2[\{\boldsymbol{F}_2^m, +\}]$;
- $\mathcal{W}$ is the set of two-dimensional affine subspaces of $\boldsymbol{F}_2^m$;
- $\mathcal{W}_0$ is the set of two-dimensional linear subspaces;
- $\langle e_1, \ldots, e_k \rangle$ is the linear space spanned by $e_1, \ldots, e_k$.

I. Introduction

THE general framework of this paper is double: coding theory (and in particular the class of Reed–Muller codes) on one hand and symmetric cryptography (block ciphers and stream ciphers) on the other hand. In both of these general domains, the Boolean functions defined on the set $\boldsymbol{F}_2^m$ of all binary words of length m play an important role. Some open problems on Boolean functions are of most interest in both fields. One of them is the determination of those functions which lie at large Hamming distance from the Reed–Muller code of order 1, $R(1, m)$. This code can be viewed as the set of all affine forms on the m-dimensional vector space $\boldsymbol{F}_2^m$ (an affine form is the sum of a linear form and of one of the constants 0 or 1). The Hamming distance between two Boolean functions is equal to the number of words of $\boldsymbol{F}_2^m$ at which they take different values. The maximum Hamming distance between a general Boolean function and $R(1, m)$ is the covering radius of this code. Its value is known only when m is even or when $m = 1, 3, 5, 7$.

The covering radius of a code is an important parameter, which can be used for analyzing and improving the decoding algorithms devoted to this code. The knowledge of the covering radius of $R(1, m)$ has therefore theoretical and practical importance for coders. It is also a serious challenge for cryptographers: the design of conventional cryptographic systems relies on two fundamental principles introduced by Shannon [2]: *confusion* and *diffusion*. The distance from a Boolean function to the set of all affine functions is called the *nonlinearity* of the function and it allows to quantify some kind of confusion. More precisely, the Boolean functions used in block ciphers must have a large nonlinearity to resist linear attacks [3]; in stream ciphers, the use of highly nonlinear Boolean functions prevents fast correlation attacks [4]. The knowledge of the maximum nonlinearity of Boolean functions is therefore necessary to appreciate (together with other criteria) the practical interest of a given Boolean function for cryptographic applications. Unfortunately, the covering radius of $R(1, m)$ for odd $m \geq 9$ is unknown. We know only that it lies between $2^{m-1} - 2^{\frac{m-1}{2}}$ and $2^{m-1} - 2^{\frac{m}{2}-1}$ (the lower bound can be slightly improved for $m \geq 15$). One aim of this paper is studying, for m odd, those functions whose nonlinearities lie between these two numbers.

For m even, the situation seems better since we know the exact value of the covering radius of $R(1, m)$: $2^{m-1} - 2^{\frac{m}{2}-1}$ (except that the *bent* functions, whose nonlinearity is maximum, are not all determined and that their determination is considered as a difficult open problem). However, from a cryptographic point of view, the case m even is in fact not better than the case m odd, since bent functions are not *balanced* (i.e., their values are not uniformly distributed); bent functions are then usually improper for use in cryptosystems. For this reason, it is also necessary to study those functions which have large but not optimal nonlinearity, say between $2^{m-1} - 2^{\frac{m}{2}}$ and $2^{m-1} - 2^{\frac{m}{2}-1}$. This is what we do also in this paper. Among these functions there are some balanced functions. The maximum nonlinearity of balanced functions is unknown for any $m \geq 8$.

Manuscript received March 8, 2000; revised November 28, 2000.

A. Canteaut and P. Charpin are with the INRIA, Projet CODES, Domaine de Voluceau, Rocquencourt, 78153 Le Chesnay Cedex, France (e-mail: Anne.Canteaut@inria.fr; Pascale.Charpin@inria.fr).

C. Carlet is with GREYC, University of Caen, 14032 Caen Cedex, France (e-mail: Claude.Carlet@inria.fr).

C. Fontaine is with LIFL, University of Sciences and Technology of Lille, 59655 Villeneuve d'Ascq Cedex, France (e-mail: Caroline.Fontaine@lifl.fr).

Communicated by I. F. Blake, Associate Editor for Coding Theory.

Publisher Item Identifier S 0018-9448(01)02714-6.

We study also other cryptographic criteria related to the notion of diffusion. The *strict avalanche criterion* (SAC) was introduced by Webster and Tavares [5] and this concept was generalized into the *propagation criterion* (PC) by Preneel [6] (see also [7]). The SAC, and its generalizations, are based on the properties of the derivatives of Boolean functions. These properties describe the behavior of a function whenever some input coordinates are complemented. We want to point out the relations between the propagation criterion and the nonlinearity. These two criteria are of most interest and form the subject of many current works. The general idea we develop, with these aims, is that the whole Fourier spectra of the functions have to be taken in account. This point of view leads us to consider both the Fourier spectrum of any given Boolean function and the coset of the Reed–Muller code of order 1 generated by the associated codeword. Therefore, several representations are proposed, in particular in the context of group codes, the aim being to have in hand all useful tools.

The paper is organized as follows. Section II is devoted to the presentation of the main tools. We first give basic properties on Boolean functions on $\boldsymbol{F}_2^m$ where the functions are implicitly represented by their *algebraic normal forms*. The study of algebraic properties of Boolean functions of m variables leads us to the study of binary codewords of length 2^m and of their relation with Reed–Muller codes. On the other hand, we need to use any basis in $\boldsymbol{F}_2^m$ and to treat some permutations on $\boldsymbol{F}_2^m$. So the codewords are viewed as formal sums in the binary group algebra $\mathcal{A}$ of the elementary 2-group $\{\boldsymbol{F}_{2^m}, +\}$. Section II is also devoted to the derivation and its significance considering the operations in $\mathcal{A}$.

These tools are applied in Section III, where we study the *maximal odd-weighting subspace* of a given Boolean function f. This concept was recently introduced in [8] and was shown to be linked with the nonlinearity of f. By replacing this concept in the ambient space of Reed–Muller codes, we prove the existence of maximal odd-weighting subspaces, for any f (Theorem III.1).

Section IV is devoted to the study of weight distributions of cosets of $R(1, m)$. By Theorem IV.1 we establish general results on the weight polynomial of any binary linear code of length 2^m and dimension $m+2$. We introduce *almost-optimal cosets* of $R(1, m)$ which correspond to functions with a high nonlinearity (see Definitions II.1 and IV.1). Considering the code $D \cup R(1, m)$, where D is any coset of $R(1, m)$, Corollary IV.1 is then deduced: we show that it is possible to distinguish among almost-optimal cosets those which have three weights only, the *three-valued almost-optimal cosets*. The next subsection is an extension of Corollary IV.1. We exhibit as an indicator of the nonlinearity, the number b_4 of codewords of weight 4 in the dual code. We are more explicit about the computation of b_4 for cosets which are contained in the third-order Reed–Muller code $R(3, m)$.

Note that, when m is odd, the main open problem is the determination of almost-optimal cosets of $R(1, m)$ with unknown weight distributions. But the context is similar for m even, if we consider the problem of the nonlinearity of balanced Boolean functions.

Section V deals with the propagation criterion and its relations with the nonlinearity. A function is said to be almost-optimal (resp., three-valued almost-optimal) if the associated coset of $R(1, m)$ satisfies this property.

In Section V-A, we study the *sum-of-squares indicator* $\mathcal{V}(f)$ of a Boolean function f, which measures the *global avalanche criterion* (GAC)—introduced in [9]. We first give an upper bound on $\mathcal{V}(f)$ in the case where f is almost-optimal (Proposition V.2). This result will have a lot of applications in the sequel of the paper. For instance, we show in this section that an almost-optimal function of degree 3 must have "many" balanced derivatives (Corollary V.1).

We next study the restrictions of a Boolean function f to each coset of any linear subspace of $\boldsymbol{F}_2^m$ (Section V-B). The main result is given by Theorem V.1, where we establish a relation between the Fourier spectrum of f and the Fourier spectra of its restrictions to these subspaces.

In Section V-C, we examine the cases where the derivatives $D_e f$ of a given function f are balanced for any $e \neq 0$ belonging to a subspace of codimension 1 or 2. These cases allow us to obtain some characterizations of bent functions and of three-valued almost-optimal functions. Theorem V.3 is most surprising since it provides a full explanation of links between bent functions and three-valued almost-optimal functions.

In the last section, we consider Boolean functions whose nonbalanced derivatives $D_a f$ exist when a belongs to a subset of rank $k < m$ only. In this case, we can be more precise, by applying the results of Section V-B. We notably characterize the almost-optimal functions which have a linear structure (Corollaries V.4 and V.5). By Theorem V.5, we show that the links between such functions and some of their decompositions are of most interest.

II. Definitions and Basic Properties

The *distance* between two codewords will always be the Hamming distance. The *weight* of any binary vector $\boldsymbol{a} = (a_1, \ldots, a_n)$ will be the Hamming weight

$$\mathrm{wt}(\boldsymbol{a}) = \sum_{i=1}^{n} a_i.$$

The support of $\boldsymbol{a}$, denoted by $\mathrm{supp}(\boldsymbol{a})$, is the set of all labels i such that $a_i \neq 0$.

A. Boolean Functions

We denote by $\mathcal{B}_m$ the set of Boolean functions of m variables. Let $f \in \mathcal{B}_m$; thus, f is a function from $\boldsymbol{F}_2^m$ to $\boldsymbol{F}_2$. The classical representation of f is its *algebraic normal form*

$$f(x_1, \ldots, x_m) = \sum_{u \in \boldsymbol{F}_2^m} \lambda_u \left(\prod_{i=1}^{m} x_i^{u_i} \right), \qquad \lambda_u \in \boldsymbol{F}_2.$$

The *degree* of f, denoted by $\deg(f)$, is the maximal value of $\mathrm{wt}(u)$ such that $\lambda_u \neq 0$. On the other hand, let us denote by Ω_f the codeword equal to the list of all values $f(x)$, $x \in \boldsymbol{F}_2^m$. Then we denote by $\mathcal{F}$ the mapping $\mathcal{F}(f) = \sum_{x \in \boldsymbol{F}_2^m} (-1)^{f(x)}$ related to the *Fourier transform* (see below). It is also related to the weight of the codeword Ω_f

$$\mathcal{F}(f) = \sum_{x \in \boldsymbol{F}_2^m} (-1)^{f(x)} = 2^m - 2\mathrm{wt}(\Omega_f). \tag{1}$$

We denote by φ_α, $\alpha \in \boldsymbol{F}_2^m$, the linear function $x \longmapsto \alpha \cdot x$. Note that the algebraic normal form of φ_α is

$$\varphi_\alpha(x_1, \ldots, x_m) = \sum_{i=1}^{m} \alpha_i x_i, \qquad \alpha_i \in \boldsymbol{F}_2.$$

We now give a list of basic definitions and properties; we keep the above stated notation.

Definition II.1: When $\mathcal{F}(f) = 0$, the function f is said to be *balanced*. The mapping $\alpha \in \boldsymbol{F}_2^m \mapsto \mathcal{F}(f + \varphi_\alpha)$ is called the *Fourier transform* of f. The multiset

$$\{\pm\mathcal{F}(f + \varphi_\alpha) | \alpha \in \boldsymbol{F}_2^m\}$$

is called the *Fourier spectrum* of f. The *nonlinearity* $\mathcal{N}(f)$ of f is the minimum Hamming distance between Ω_f and all codewords associated to the affine functions φ_α and $\varphi_\alpha + 1$. It is equal to $2^{m-1} - \frac{\mathcal{L}(f)}{2}$, where

$$\mathcal{L}(f) = \max_{\alpha \in \boldsymbol{F}_2^m} |\mathcal{F}(f + \varphi_\alpha)|.$$

Note that we are not only interested in the values appearing in the Fourier spectrum, but also in the number of times they occur. The multiset $\{\pm\mathcal{F}(f + \varphi_\alpha)\}$ is often called the *extended* Walsh spectrum (see, for instance, [10]).

The nonlinearity of f being the minimum Hamming weight of the coset $\Omega_f + R(1, m)$ we have $\mathcal{N}(f) \leq \rho(R(1, m))$ where $\rho(R(1, m))$ is the *covering radius* of $R(1, m)$:

$$\rho(R(1, m)) = \max_{y \in \boldsymbol{F}_2^{2^m}} \left(\min_{x \in y + R(1, m)} \mathrm{wt}(x) \right).$$

When m is even, it is known that $\rho(R(1, m)) = 2^{m-1} - 2^{m/2-1}$ and that the Fourier spectrum of functions of maximal nonlinearity is unique [11]. In particular, it does not contain 0 (hence those functions are not balanced).

Definition II.2: A Boolean function $f \in \mathcal{B}_m$, m even, is said to be bent when

$$\mathcal{N}(f) = \rho(R(1, m)) = 2^{m-1} - 2^{m/2-1}.$$

The Fourier spectrum of such a function is $\{\pm 2^{m/2}\}$.

The case where m is odd is completely different. A recent review is given in [12]. We have [13]

$$2^{m-1} - 2^{\frac{m-1}{2}} \leq \rho(R(1, m)) < 2^{m-1} - 2^{\frac{m}{2}-1}.$$

For $m = 3, 5, 7$, $\rho(R(1, m))$ is equal to $2^{m-1} - 2^{\frac{m-1}{2}}$. But the exact lower bound is not known for $m \geq 9$.

A function has a *good nonlinearity* if its nonlinearity is large, i.e., if $\mathcal{L}(f)$ is small. We say that $\mathcal{L}(f)$ is small when

$$2^{m/2} < \mathcal{L}(f) \leq 2^{(m+1)/2}.$$

This corresponds to the case where

$$2^{m-1} - 2^{\frac{m-1}{2}} \leq \mathcal{N}(f) < 2^{m-1} - 2^{\frac{m}{2}-1}.$$

The SAC was generalized into the *propagation criterion* (PC) by Preneel [6]. More recently, Zhang and Zheng introduced the *global avalanche criterion* (GAC), in order to measure the global avalanche characteristics of cryptographic functions [9]. These criteria are based on the properties of the functions $x \longmapsto f(x) + f(x + a)$, $a \in \boldsymbol{F}_2^m$.

Definition II.3: Let f be a Boolean function on $\boldsymbol{F}_2^m$ and $a \in \boldsymbol{F}_2^m$. We denote by $D_a f$ the derivative of f with respect to a

$$D_a f(x) = f(x) + f(a + x).$$

i) The *linear space* of f is the linear subspace of those a such that $D_a f$ is a constant function. Such a, $a \neq 0$, is said to be a linear structure of f [14].

ii) Let $E \subset \boldsymbol{F}_2^m$. The function f satisfies *PC with respect to* E if for all $e \in E$ the function $D_e f$ is balanced.

iii) The function f is said to have a good GAC if $|\mathcal{F}(D_a f)|$ is zero or is very close to zero for most nonzero a.

We now recall some fundamental formulas. Parseval's relation

$$\sum_{\alpha \in \boldsymbol{F}_2^m} \mathcal{F}^2(f + \varphi_\alpha) = 2^{2m} \tag{2}$$

and a formula which states the link between f and its derivatives

$$\begin{aligned} \mathcal{F}^2(f + \varphi_\alpha) &= \sum_{a \in \boldsymbol{F}_2^m} \mathcal{F}(D_a f + \varphi_\alpha(a)) \\ &= \sum_{a \in \boldsymbol{F}_2^m} (-1)^{\alpha \cdot a} \mathcal{F}(D_a f). \end{aligned} \tag{3}$$

This was proved by Carlet in [15] and [16], giving particularly

$$\mathcal{F}^2(f) = \sum_{a \in \boldsymbol{F}_2^m} \mathcal{F}(D_a f). \tag{4}$$

In [9], the authors propose two indicators related to the GAC: we shall denote by $\mathcal{MD}(f)$ the *absolute indicator*

$$\mathcal{MD}(f) = \max_{a \in \boldsymbol{F}_2^m, a \neq 0} |\mathcal{F}(D_a f)| \tag{5}$$

and by $\mathcal{V}(f)$ the second moment of the autocorrelation coefficients called the *sum-of-squares* indicator

$$\mathcal{V}(f) = \sum_{a \in \boldsymbol{F}_2^m} \mathcal{F}^2(D_a f) = \sum_{a, b \in \boldsymbol{F}_2^m} \mathcal{F}(D_a D_b f). \tag{6}$$

Note that obviously $\mathcal{V}(f) \geq 2^{2m}$, since $\mathcal{F}^2(D_0 f) = 2^{2m}$. The next formula provides a relation between $\mathcal{V}(f)$ and the Fourier spectrum of f, i.e., the values $|\mathcal{F}(f + \varphi_\alpha)|$, $\alpha \in \boldsymbol{F}_2^m$.

Proposition II.1: For any Boolean function $f \in \mathcal{B}_m$, we have

$$\begin{aligned} \forall e \in \boldsymbol{F}_2^m, \quad \sum_{\alpha \in \boldsymbol{F}_2^m} \mathcal{F}^2(f + \varphi_\alpha) \mathcal{F}^2(f + \varphi_{\alpha + e}) \\ = 2^m \sum_{a \in \boldsymbol{F}_2^m} \mathcal{F}^2(D_a f) (-1)^{e \cdot a} \end{aligned}$$

providing, for $e = 0$, a relation between the Fourier spectrum of f and the sum-of-squares indicator defined by (6)

$$\sum_{\alpha \in \boldsymbol{F}_2^m} \mathcal{F}^4(f + \varphi_\alpha) = 2^m \mathcal{V}(f). \tag{7}$$

Proof: Let

$$G = \sum_{\alpha \in \boldsymbol{F}_2^m} \mathcal{F}^2(f + \varphi_\alpha) \mathcal{F}^2(f + \varphi_{\alpha + e}).$$

According to (3), we have for all $e \in \boldsymbol{F}_2^m$

$$G = \sum_{\alpha\in\boldsymbol{F}_2^m}\left(\sum_{a\in\boldsymbol{F}_2^m}\mathcal{F}(D_af)(-1)^{a\cdot\alpha}\right) \times\left(\sum_{b\in\boldsymbol{F}_2^m}\mathcal{F}(D_bf)(-1)^{b\cdot(\alpha+e)}\right)$$
$$= \sum_{\alpha\in\boldsymbol{F}_2^m}\sum_{a,\,b\in\boldsymbol{F}_2^m}\mathcal{F}(D_af)\mathcal{F}(D_bf)(-1)^{b\cdot e+\alpha\cdot(a+b)}$$
$$= \sum_{a,\,b\in\boldsymbol{F}_2^m}\mathcal{F}(D_af)\mathcal{F}(D_bf)(-1)^{b\cdot e}\sum_{\alpha\in\boldsymbol{F}_2^m}(-1)^{\alpha\cdot(a+b)},$$

where

$$\sum_{\alpha\in\boldsymbol{F}_2^m}(-1)^{\alpha\cdot(a+b)} = 0$$

unless $a = b$. Then we deduce, for any $e \in \boldsymbol{F}_2^m$

$$G = 2^m\sum_{a\in\boldsymbol{F}_2^m}\mathcal{F}^2(D_af)(-1)^{e\cdot a}.$$

We complete the proof by using the definition of the sum-of-squares indicator given by (6). □

Our purpose is to point out that there are interesting connections between the GAC and the nonlinearity. Note, as a trivial example, that the bent functions—i.e., the functions which have the best nonlinearity $2^{m-1}-2^{\frac{m}{2}-1}$ for m even—have a perfect GAC, since their derivatives are all balanced. *For such a function f, we have $\mathcal{MD}(f) = 0$ and $\mathcal{V}(f) = 2^{2m}$. Moreover these equalities hold for bent functions only.*

On the other hand, a function f which has a linear space V satisfies

$$|\mathcal{F}(D_af)| = 2^m, \qquad \forall a \in V \tag{8}$$

(see Definition II.3). Hence, $\mathcal{MD}(f)$ takes the maximal value and one can say that f has not a good GAC. We obviously deduce a lower bound for $\mathcal{V}(f)$.

Lemma II.1: A function f, which has a linear space V of dimension k, $k \geq 1$, satisfies (8) and is such that $\mathcal{V}(f) \geq 2^{2m+k}$.

However, the nonlinearity of a function f which has a linear structure is not always so bad. We will show later that there exist such functions satisfying $\mathcal{L}(f) = 2^{(m+2)/2}$ for even m and $\mathcal{L}(f) = 2^{(m+1)/2}$ for odd m (see Corollaries V.4 and V.5).

For clarity, we notice that $\mathcal{MD}(f)$ and $\mathcal{V}(f)$ are invariant if we change f into $f + \varphi_\alpha$: since $D_a\varphi_\alpha$ is a constant function, we have

$$|\mathcal{F}(D_af)| = |\mathcal{F}(D_a(f+\varphi_\alpha))|, \qquad \text{for any } \alpha$$

implying the next property.

Lemma II.2: For any $\alpha \in \boldsymbol{F}_2^m$, we have

$$\mathcal{MD}(f+\varphi_\alpha) = \mathcal{MD}(f) \quad \text{and} \quad \mathcal{V}(f+\varphi_\alpha) = \mathcal{V}(f).$$

We want to end this section with few elements on *resilient functions*. In this paper, we do not emphasize the criterion of correlation immunity. However, this concept is strongly related to the properties of balanced functions and thus with our next results (see [10, Sec. 7]).

Definition II.4: Let $\boldsymbol{e} = (e_1, \ldots, e_m)$ be any basis of $\boldsymbol{F}_2^m$. A function $f \in \mathcal{B}_m$ is said to be *correlation-immune of order k, $1 \leq k \leq m$, with respect to $\boldsymbol{e}$* if for any vector $\alpha = (\alpha_1, \ldots, \alpha_m)$ in $\boldsymbol{F}_2^m$ such that $0 < \mathrm{wt}(\alpha) \leq k$, the function

$$f + \varphi_\alpha, \qquad \alpha = \sum_{i=1}^m \alpha_i e_i$$

is balanced. The function f is said to be *resilient of order k* if it is additionally balanced.

We now recall the simplest link between nonlinearity and resiliency.

Proposition II.2: Let $f \in \mathcal{B}_m$. Let us denote by ν the number of 0's in the Fourier spectrum of f. Then we have

$$\nu \leq 2^m - \frac{2^{2m}}{\mathcal{L}(f)^2}$$

with equality if and only if the values occurring in the Fourier spectrum of f lie in $\{0, \pm\mathcal{L}(f)\}$.

Most notably, this implies:

- for m even, if $\mathcal{L}(f) \leq 2^{(m+2)/2}$ then $\nu \leq 2^{m-1} + 2^{m-2}$, with equality if and only if the values occurring in the Fourier spectrum of f lie in $\{0, \pm 2^{(m+2)/2}\}$;
- for m odd, if $\mathcal{L}(f) \leq 2^{(m+1)/2}$ then $\nu \leq 2^{m-1}$, with equality if and only if the values occurring in the Fourier spectrum of f lie in $\{0, \pm 2^{(m+1)/2}\}$.

Proof: We simply use Parseval's relation (see (2)). Let A be the set of all α such that $f + \varphi_\alpha$ is not balanced. Then we have

$$\sum_{\alpha\in A}\mathcal{F}^2(f+\varphi_\alpha) = 2^{2m}.$$

Since $|A| = 2^m - \nu$, we deduce that

$$\mathcal{L}(f)^2(2^m - \nu) \geq 2^{2m}$$

i.e.,

$$\nu \leq 2^m - \frac{2^{2m}}{\mathcal{L}(f)^2}.$$

Equality in the above formula holds if and only if all nonzero values of the Fourier spectrum are equal to $\pm\mathcal{L}(f)$. □

Remark II.1: By the previous property we give a significant upper bound on the number of balanced functions $f+\varphi_\alpha$, when f has a good nonlinearity. This contradicts a high order of resiliency.

B. Product and Derivation

The study of properties of Boolean functions of m variables leads us to the study of binary codewords of length 2^m. More generally, any set of Boolean functions provides a set of codewords and can be studied by means of tools of coding theory.

The main concern is with Reed–Muller codes as we first state in the next definition.

Definition II.5: The Reed–Muller code of length 2^m and order r, $1 \le r \le m$, denoted by $R(r, m)$, is the binary code of length 2^m composed of the codewords Ω_f where f is a Boolean function of m variables whose degree is less than or equal to r.

We described above some properties of $\mathcal{B}_m$ by taking the standard basis in $\boldsymbol{F}_2^m$. It is clear that any basis can be chosen. From now on, we will consider that $f \in \mathcal{B}_m$ is a function from $\boldsymbol{F}_2^m$ to $\boldsymbol{F}_2$ where $\boldsymbol{F}_2^m$ is viewed as an additive group. We will fix a basis in $\boldsymbol{F}_2^m$ when it will be necessary. However, we have to mention that generally, for cryptographic applications, the basis is fixed and the properties have to be considered relatively to the chosen basis.

The concept of "derivative" can be seen as a multiplication in a group algebra, the ambient space of binary codes of length 2^m. We begin by recalling some definitions and properties. An extensive study was made by Assmus and Key in [17] and Charpin in [18] and [19]; we only give basic elements for the use of the algebraic tools which are provided here.

Definition II.6: Let us denote by $\mathcal{A}$ the group algebra $\boldsymbol{F}_2[\boldsymbol{F}_2^m]$. The algebra $\mathcal{A}$ is the set of all binary words of length 2^m; such a word x is a formal polynomial

$$\mathrm{x} = \sum_{g \in \boldsymbol{F}_2^m} x_g X^g, \qquad x_g \in \boldsymbol{F}_2.$$

The operations are

$$\begin{aligned} a\mathrm{x} + b\mathrm{y} &= a \sum_{g \in \boldsymbol{F}_2^m} x_g X^g + b \sum_{g \in \boldsymbol{F}_2^m} y_g X^g \\ &= \sum_{g \in \boldsymbol{F}_2^m} (a x_g + b y_g) X^g, \\ \mathrm{xy} &= \sum_{g \in \boldsymbol{F}_2^m} x_g X^g \times \sum_{g \in \boldsymbol{F}_2^m} y_g X^g \\ &= \sum_{g \in \boldsymbol{F}_2^m} \left(\sum_{\substack{h,\, k \in \boldsymbol{F}_2^m \\ h+k=g}} x_h y_k \right) X^g, \end{aligned}$$

where $a \in \boldsymbol{F}_2$, $b \in \boldsymbol{F}_2$, $\mathrm{x} \in \mathcal{A}$, $\mathrm{y} \in \mathcal{A}$. Note that the multiplicative unit is X^0. The all-one vector and the null vector will be denoted by $\mathbf{1}$ and $\mathbf{0}$, respectively. By convention, X^0 is denoted 1. An ideal I of $\mathcal{A}$ is a subgroup (and, thus, a subspace) invariant under the multiplication by X^b, for some b. The algebra $\mathcal{A}$ has only one maximal ideal, called its *radical*, which is the set of all words of even weights

$$\mathcal{P} = \left\{ \sum_{g \in \boldsymbol{F}_2^m} x_g X^g \,\middle|\, \sum_{g \in \boldsymbol{F}_2^m} x_g = 0 \bmod 2 \right\}.$$

Thus, we can define the ideals $\mathcal{P}^j$, $1 \le j \le m+1$, generated by the products $\prod_{i=1}^{j} \mathrm{x}_i$, $\mathrm{x}_i \in \mathcal{P}$, providing the decreasing sequence

$$\mathcal{A} = \mathcal{P}^0 \supset \mathcal{P} \supset \cdots \supset \mathcal{P}^{m-1} \supset \mathcal{P}^m = \{\mathbf{0}, \mathbf{1}\}$$

where $\mathcal{P}^i \mathcal{P}^j = \mathcal{P}^{i+j}$ and $\mathcal{P}^{m+1} = \{\mathbf{0}\}$. Recall the fundamental result, due to Berman [20] (see also [17, Theorem 4.2]).

Theorem II.1: The powers of the radical of the algebra $\mathcal{A}$ are the Reed–Muller codes. More precisely, for any r, $R(r, m) = \mathcal{P}^{m-r}$.

In the sequel, we will generally use the notation $\mathcal{P}^j$ when we have to handle some multiplications in $\mathcal{A}$. Recall that $\mathcal{P}^j$ *is the subspace generated by the codewords whose supports are the* j*-dimensional subspaces* of $\boldsymbol{F}_2^m$ [17, Corollary 3.11]

$$\sum_{v \in V} X^v = \prod_{i=1}^{j} (X^{v_i} + 1), \; V = \langle v_1, \ldots, v_j \rangle. \tag{9}$$

The so-called *Jenning's Basis* provides a basis of $\mathcal{A}$ containing a basis of each $\mathcal{P}^j$ as we recall in the next proposition—a proof, for any characteristic, can be found in [17, p. 1299].

Proposition II.3: Let $(e_1, \ldots, e_m)$ be a basis of $\boldsymbol{F}_2^m$. Then the set

$$\left\{ \prod_{i=1}^{m} (X^{e_i} + 1)^{k_i} \,\middle|\, (k_1, \ldots, k_m) \in \{0, 1\}^m \right\}$$

is a basis of $\mathcal{A}$. Moreover, for each j, $1 \le j \le m$, the set

$$\left\{ \prod_{i=1}^{m} (X^{e_i} + 1)^{k_i} \,\middle|\, \sum_{i=1}^{m} k_i \ge j \right\}$$

is a basis of $\mathcal{P}^j$, the Reed–Muller code of order $m - j$.

Let $f \in \mathcal{B}_m$. The associated codeword of f is written as follows in $\mathcal{A}$:

$$\Omega_f = \sum_{g \in \boldsymbol{F}_2^m} f(g) X^g.$$

So we clearly have

$$\mathcal{N}(f) = \min_{\alpha \in \boldsymbol{F}_2^m} \left\{ \mathrm{wt}(\mathrm{x}) \,\middle|\, \mathrm{x} = \sum_{g \in \boldsymbol{F}_2^m} (f(g) + \varphi_\alpha(g)) X^g \right\}. \tag{10}$$

On the other hand, for any $a \in \boldsymbol{F}_2^m$, we have

$$X^a \Omega_f = \sum_{g \in \boldsymbol{F}_2^m} f(g) X^{g+a} = \sum_{g \in \boldsymbol{F}_2^m} f(g+a) X^g, \tag{11}$$

showing that $(X^a + 1)\Omega_f$ is the associated codeword of $D_a f$.

More generally, the concept of kth-*derivative*, given in the next definition, is actually a multiplication in the algebra $\mathcal{A}$.

Definition II.7: Let V be a k-dimensional subspace of $\boldsymbol{F}_2^m$. The kth-*derivative* of $f \in B_m$ *with respect to* V is the function

$$D_{a_1, \ldots, a_k} f = D_{a_1} D_{a_2} \cdots D_{a_k} f$$

where $(a_1, \ldots, a_k)$ is any basis of V.

Proposition II.4: Let V be a k-dimensional subspace of $\boldsymbol{F}_2^m$; $(a_1, \ldots, a_k)$ denotes any basis of V. Let $f \in \mathcal{B}_m$ be any function of degree r. Set $h = D_{a_1, \ldots, a_k} f$. Then

$$\Omega_h = \sum_{g \in \boldsymbol{F}_2^m} \left(\sum_{a \in V} f(g+a) \right) X^g = \left(\sum_{v \in V} X^v \right) \Omega_f.$$

The degree of h is less than or equal to $r-k$. When $r<k$, h is the zero function. In particular, the derivative of f with respect to a has degree at most $r-1$ and corresponds to the product by X^a+1 in $\mathcal{A}$

$$\Omega_{D_af}=(X^a+1)\Omega_f.$$

Proof: We deduce from (11)

$$(X^a+1)\Omega_f=\sum_{g\in\boldsymbol{F}_2^m}(f(g)+f(g+a))X^g=\Omega_{D_af}.$$

Set $\mathrm{y}=\sum_{v\in V}X^v$, the codeword of support V. The general formula is easily obtained by expanding the product $(\mathrm{y}\Omega_f)$. For instance,

$$D_{a_1}D_{a_2}f(g)=f(g)+f(g+a_1)+f(g+a_2)+f(g+a_1+a_2).$$

The codeword y is in $\mathcal{P}^k$, by definition (see (9)). Assume that f has degree r—this means that the codeword Ω_f is in the Reed–Muller code of order r. So, from Theorem II.1, $\Omega_f\in\mathcal{P}^{m-r}$ implying that the product $\mathrm{y}\Omega_f$ is in $\mathcal{P}^k\mathcal{P}^{m-r}=\mathcal{P}^{m+k-r}$, which is the Reed–Muller code of order $r-k$. So the degree of h is less than or equal to $r-k$. □

In the next section, we will develop a concept directly stemming from the concept of derivation. To end this section we give some obvious properties and mention an important class of functions. Note that $f=1$ (resp., $f=0$) means that the function f is constant, with associated codeword $\mathbf{1}$ (resp., $\mathbf{0}$).

Proposition II.5: Let $f\in\mathcal{B}_m$. Then we have the following.

1) If there exists $a\in\boldsymbol{F}_2^m$ such that $D_af=1$ then f is balanced.
2) When $\deg(f)\le 2$, f is balanced if and only if there exists $a\in\boldsymbol{F}_2^m$ such that $D_af=1$.
3) When $\deg(f)\le 3$, D_af is balanced if and only if there exists $b\in\boldsymbol{F}_2^m$ such that $D_aD_bf=1$.

Proof: For proving the first property, it is sufficient to notice that $\Omega_f+X^a\Omega_f=\mathbf{1}$ implies that $2\mathrm{wt}\,(\Omega_f)=2^m$. We recall the proof of the second property in Appendix I. The third property is then deduced, since D_af has degree at most 2 when $\deg(f)\le 3$. □

Example II.1: The above property allows us to characterize a large class of balanced functions by means of their associated codewords. Let H be any subspace of codimension 1 in $\boldsymbol{F}_2^m$. The weight of the following codewords x is 2^{m-1}:

$$\mathrm{x}=(X^e+1)\mathrm{y}+\mathrm{z},\quad \mathrm{z}=\sum_{h\in H}X^h,\quad\text{and}\quad e\notin H.$$

Indeed, x is balanced for any y, since

$$(X^e+1)\mathrm{x}=(X^e+1)\mathrm{z}=\mathbf{1}.$$

The corresponding functions have a linear structure.

The *partially bent functions* were introduced by Carlet in [16]. These functions are quadratic-like functions, in the sense that the dimension of their linear space is sufficient for determining their Fourier spectra. With our terminology we obtain directly, from [16, p. 137], the form of the codewords corresponding to partially bent functions.

Proposition II.6: A Boolean function f of m variables is said to be *partially bent* if there exists a basis $(e_1,\ldots,e_m)$ of $\boldsymbol{F}_2^m$ such that $f=g+\varphi_\alpha$, where g is a bent function on the $(m-k)$-dimensional space $\langle e_{k+1},\ldots,e_m\rangle$ for some $k<m$ such that $m-k$ is even, and φ_α is a linear function.

The codewords corresponding to partially bent functions have the following form:

$$\Omega_f=\prod_{i=1}^{k}(X^{e_i}+1)\left(\sum_{a\in\langle e_{k+1},\ldots,e_m\rangle}g(a)X^a\right)+\Omega_{\varphi_\alpha}.$$

Note that $\langle e_1,\ldots,e_k\rangle$ is the linear space of f and that Ω_f lies in $\mathcal{P}^{\frac{m-k}{2}}$. Moreover, *$f$ is a balanced partially bent function* if and only if there is $c\in\boldsymbol{F}_2^m$ such that $D_cf=1$ (see [16, Proposition 2] and Appendix I).

Open Problem II.1: Since any quadratic function is partially bent, the derivatives of any function of degree 3 are partially bent. Characterize a class of functions of degree r, $r>3$, whose derivatives are all partially bent.

Notice that there exist bent functions whose derivatives are not all partially bent. Consider, for instance, Maiorana–McFarland functions: we identify the elements of $\boldsymbol{F}_2^m$, $m=2t$, with the pairs (x,y) where $x=(x_1,\ldots,x_t)$ and $y=(y_1,\ldots,y_t)$ and we define

$$f(x,y)=x\cdot\pi(y)+g(y)$$

where π is some bijection from $\boldsymbol{F}_2^t$ to $\boldsymbol{F}_2^t$ (with the usual dot product "$\cdot$") and g is some function in $\mathcal{B}_t$. The derivative $D_{e_i}f$ of f with respect to the ith word of weight 1, e_i, for $1\le i\le t$, is equal to the ith coordinate function π_i of π. Since the m-variable function $D_{e_i}f$ only depends on t variables, its linear space has dimension at least $m-t=t$. Recall that the degree of a partially bent function is at most the half of the codimension of its linear space [16]. The derivative $D_{e_i}f$ cannot be partially bent if the degree of π_i is greater than $t/2$. This situation occurs, for example, if $\pi(y)=y^s$ where $\boldsymbol{F}_2^t$ is identified with the finite field with 2^t elements, and where s is such that $\gcd(s,2^t-1)=1$ and the binary expansion of s contains more than $t/2$ 1's. An example of such π is $\pi(y)=y^{2^{t-1}-1}$ for $t\ge 3$.

III. Maximal Odd-Weighting Subspaces of Boolean Functions

Zheng, Zhang, and Imai introduced in [8] the *maximal odd-weighting subspace* of a given Boolean function f. They indicated the link between this concept and the nonlinearity of f. Replacing their concept in the ambient space $\mathcal{A}$ of Reed–Muller codes, we deduce additional properties.

Lemma III.1: Let V be a k-dimensional subspace of $\boldsymbol{F}_2^m$. Set $\mathrm{y}=\sum_{v\in V}X^v$ and $\lambda=2^{m-k}$. We denote by $V_1,\ldots,V_\lambda$ the

λ cosets of V where $V_1 = V$. Let $\mathrm{x} \in \mathcal{A}$ and, for each i, denote by x_i the restriction of x to V_i. Then the product xy satisfies

$$\mathrm{xy} = \sum_{\substack{1 \le i \le \lambda \\ \mathrm{wt}(\mathrm{x}_i) \text{ is odd}}} \sum_{g \in V_i} X^g.$$

Furthermore,

i) $\mathrm{xy} = \mathbf{0}$ (resp., $= \mathbf{1}$) if and only if the weight of x_i is even (resp., odd) for all i, $1 \le i \le \lambda$;

ii) $\mathrm{wt}(\mathrm{xy}) = \lambda_o \times 2^k$, where λ_o is the number of x_i which have odd weights.

Proof: We have

$$\mathrm{x} = \sum_{g \in \boldsymbol{F}_2^m} x_g X^g$$

where $\boldsymbol{F}_2^m = \bigcup_{1 \le i \le \lambda} V_i$. Setting $V_i = a_i + V$, for each i, with $a_1 = 0$, we obtain

$$\mathrm{x_i} = \sum_{g \in V_i} x_g X^g = \sum_{u \in V} x_{a_i+u} X^{a_i+u} = X^{a_i} \sum_{u \in V} x_{a_i+u} X^u.$$

Now

$$\begin{aligned}
\mathrm{xy} &= \sum_{i=1}^{\lambda} \sum_{g \in V_i} x_g X^g \sum_{v \in V} X^v \\
&= \sum_{i=1}^{\lambda} X^{a_i} \sum_{u \in V} x_{a_i+u} X^u \sum_{v \in V} X^v \\
&= \sum_{i=1}^{\lambda} X^{a_i} \left(\sum_{u \in V} x_{a_i+u} \right) \left(\sum_{v \in V} X^v \right) \\
&= \sum_{i=1}^{\lambda} \left(\sum_{g \in V_i} X^g \right) \times (\mathrm{wt}(\mathrm{x}_i) \bmod 2)
\end{aligned}$$

giving the main formula. Note that $X^u \mathrm{y} = \mathrm{y}$, for any $u \in V$.

Since $\sum_{g \in V_i} X^g$ is the all-one vector of length 2^k and support V_i, i) and ii) are immediately deduced. □

Proposition III.1: Let $\mathrm{x} \in \mathcal{A}$ and $1 \le j \le m$. Then

i) x lies in $R(m-j, m)$ if and only if for any subspace W of $\boldsymbol{F}_2^m$ of dimension $m-j+1$, we have:

$$\mathrm{x} \left(\sum_{v \in W} X^v \right) = \mathbf{0}$$

— i.e., the restriction of x to each coset of W has an even weight.

ii) x lies in $R(m-j, m) \setminus R(m-j-1, m)$ if and only if $\mathrm{x} \in R(m-j, m)$ and there is a subspace V of $\boldsymbol{F}_2^m$ of dimension $m-j$ such that

$$\mathrm{x} \left(\sum_{v \in V} X^v \right) = \mathbf{1}$$

— i.e., the restriction of x to each coset of V has an odd weight.

Proof: Recall that $R(m-j, m) = \mathcal{P}^j$, implying

$$(\mathcal{P}^j)^\perp = R(m-j, m)^\perp = R(j-1, m) = \mathcal{P}^{m-j+1}.$$

Remember that any element can be represented with respect to a *Jenning's Basis* (see Theorem II.1 and Proposition II.3).

The code $\mathcal{P}^j$ is generated by the codewords whose supports are the subspaces U of dimension j [17, Corollary 3.11]. The dual of $\mathcal{P}^j$ is the code $\mathcal{P}^{m-j+1}$, which is generated by the codewords whose supports are the subspaces W of dimension $m-j+1$. Since

$$\mathcal{P}^j \mathcal{P}^{m-j+1} = \mathcal{P}^{m+1} = \{\mathbf{0}\}$$

we obviously have $\mathrm{x} \in \mathcal{P}^j$ if and only if the product of x with any generator of $\mathcal{P}^{m-j+1}$ is $\mathbf{0}$, completing the proof of i).

Assume that $\mathrm{x} \in \mathcal{P}^j$. The dual of $\mathcal{P}^{j+1}$ being $\mathcal{P}^{m-j}$, we have $\mathrm{x} \notin \mathcal{P}^{j+1}$ if and only if at least one generator with support V, say $\mathrm{y} = \sum_{v \in V} X^v$, where $\dim V = m-j$, satisfies $\mathrm{xy} \neq \mathbf{0}$. Since

$$\mathcal{P}^j \mathcal{P}^{m-j} = \mathcal{P}^m = \{\mathbf{0}, \mathbf{1}\}$$

we can conclude that $\mathrm{xy} = \mathbf{1}$, completing the proof of ii). □

Now we give the definition of Zheng *et al.* [8].

Definition III.1: Let f be a Boolean function on $\boldsymbol{F}_2^m$. Let U be some k-dimensional subspace of $\boldsymbol{F}_2^m$. Denote by f_U the restriction of f to U, i.e., the function on U defined by $f_U(x) = f(x)$.

Then U is said to be a *maximal odd-weighting subspace* of f if the weight of the codeword corresponding to f_U is odd and the weight of the codeword corresponding to $f_{U'}$ is even for all subspace U' which strictly contains U.

Using Proposition III.1 we are able to complete this definition.

Theorem III.1: Let f be a Boolean function of degree r. Recall that Ω_f denotes the corresponding codeword of f. Let U be a k-dimensional subspace of $\boldsymbol{F}_2^m$ and set $\mathrm{y} = \sum_{u \in U} X^u$. Then we have

a) U is a maximal odd-weighting subspace of f if and only if the product $\mathrm{y}\Omega_f$ is equal to the all-one codeword; or, equivalently, if the kth-derivative of f with respect to U, say $D_{e_1} \cdots D_{e_k} f$ for some basis $(e_1, \ldots, e_k)$ of U, is equal to the constant function 1.

b) If U is a maximal odd-weighting subspace of f, then $k \le r$. Moreover, there exists at least one r-dimensional maximal odd-weighting subspace of f.

Proof: By definition, U is a maximal odd-weighting subspace of f if and only if the weight of the restriction of Ω_f to U and to any coset L of U is odd. This is because the set $L \cup U$ is a subspace containing U and any subspace containing U is a union of an even number of cosets of U. In accordance with Lemma III.1, we obtain: U is a maximal odd-weighting subspace of f if and only if $\mathrm{y}\Omega_f = \mathbf{1}$. Since $\mathrm{y} = \prod_{i=1}^{k} (X^{e_i} + 1)$, then $\mathrm{y}\Omega_f$ is the codeword corresponding to $D_{e_1} \cdots D_{e_k} f$ (see Proposition II.4), completing the proof of a).

Since f has degree r, Ω_f is in $R(r, m) \setminus R(r-1, m)$. From Proposition III.1 ii) and from a), there exists V of dimension r which is a maximal odd-weighting subspace of f. Moreover,

from Proposition III.1 i), it is not the case for any W of dimension $k > r$. □

Remark III.1: In their paper, Zheng *et al.* noticed that if U is a maximal odd-weighting subspace of f of dimension k, $k > 2$, then $\mathcal{N}(f) \geq 2^{m-k}$. Note that it is simply because (with the above notation)

$$2^m = \mathrm{wt}\,(\mathrm{y}\Omega_f) \leq \mathrm{wt}\,(\mathrm{y})\mathrm{wt}\,(\Omega_f) = 2^k \mathrm{wt}\,(\Omega_f).$$

Moreover, when $k > 2$, this inequality holds for $f+\varphi_\alpha$, for any α, since any second derivative of φ_α is 0.

Note that, according to Proposition II.5, a Boolean function of degree 3 has a maximal odd-weighting subspace of dimension 2 as soon as it has a balanced derivative.

IV. The Weights of Cosets of the Reed–Muller Code of Order 1

In this section, we study the nonlinearity through the properties of weight polynomials of cosets of $R(1,\ m)$. To be more precise, we establish a necessary condition for such a coset to have a high minimum weight.

A. An Extension of the Results of Kasami

The major result of this section is presented in Theorem IV.1, providing a new point of view on the characterization of the weight distributions of the cosets $\mathrm{x} + R(1,\ m)$ for any m. This result is based on Pless identities, introduced by Pless in [21], and which are obtained from MacWilliams identities (see also [22, Ch. 5]).

Let C denote an $[n,\ k,\ \delta]$ binary linear code, and $C^\perp$ its dual, which has δ' as minimum distance. Let us denote by a_w (resp., b_w), $w \in [0,\ n]$, the number of codewords of C (resp., $C^\perp$) whose Hamming weight is w. If $\delta' \geq 4$, then we have the following Pless identities (see [22, p. 130]):

$$\begin{aligned}
\sum_{w=0}^{n} a_w &= 2^k \\
\sum_{w=0}^{n} w\,a_w &= 2^{k-1}\,n \\
\sum_{w=0}^{n} w^2\,a_w &= 2^{k-2}\,n(n+1) \\
\sum_{w=0}^{n} w^3\,a_w &= 2^{k-3}\,(n^2(n+3)) \\
\sum_{w=0}^{n} w^4\,a_w &= 2^{k-4}\,(n(n+1)\,(n^2+5n-2)+4!\,b_4).
\end{aligned} \tag{12}$$

In the next theorem, we treat linear binary codes C of length 2^m and dimension $m+2$. Note that we will focus later on the linear codes $(\mathrm{x} + R(1,\ m)) \cup R(1,\ m)$, for any $\mathrm{x} \notin R(1,\ m)$.

Theorem IV.1: Let m be a positive integer, $m \geq 3$. Consider any binary linear code C of length $n = 2^m$, dimension $k = m+2$, and minimum distance δ. Let us denote by a_w (resp., b_w) the number of codewords of weight w in C (resp., $C^\perp$) and by $\mathcal{I}(\lambda)$ the number

$$\mathcal{I}(\lambda) = \sum_{w=1}^{2^{m-1}-1} (w-2^{m-1})^2 \left((w-2^{m-1})^2 - \lambda^2\right) a_w. \tag{13}$$

Assume that C contains the all-one vector $\mathbf{1}$ and that $C^\perp$ is such that $b_1 = b_2 = b_3 = 0$. Then, for any positive integer $\lambda \leq 2^{m-1}$, we have

$$\mathcal{I}(\lambda) = 2^m \left(3b_4 - 2^{m-2}\left((2^{m-1}-1)^2 + (\lambda^2 - 2^{m-1})\right)\right). \tag{14}$$

If $\delta \geq 2^{m-1} - \lambda$ then $\mathcal{I}(\lambda) \leq 0$ which can be expressed as

$$b_4 \leq \tfrac{1}{3}\,2^{m-2}\left((2^{m-1}-1)^2 - 2^{m-1} + \lambda^2\right). \tag{15}$$

Equality holds in (15) if and only if $\delta = 2^{m-1} - \lambda$ and if the weight distribution of C is: $a_0 = a_{2^m} = 1$ and we get the expression shown at the bottom of the page for the other nonzero a_w's. Since $b_4 \neq 0$, the minimum distance of $C^\perp$ is exactly 4.

Proof: The proof is based on the study of the numbers

$$I_\ell = \sum_{w=1}^{n-1} (w-2^{m-1})^\ell a_w.$$

We are particularly interested in I_2 and I_4; according to (12), we have

$$\begin{aligned}
I_2 &= \sum_{w=0}^{n} (w-2^{m-1})^2 a_w - 2^{2m-1} \\
&= 2^{k-2} n(n+1) - 2^m 2^{k-1} n + 2^{2m-2} 2^k - 2^{2m-1}
\end{aligned}$$

which gives, replacing k by $m+2$ and n by 2^m

$$I_2 = 2^{2m}(2^m+1) - 2^m 2^{2m+1} + 2^{2m-2}(2^{m+2}-2) = 2^{2m-1}. \tag{16}$$

In the same way, we obtain

$$\begin{aligned}
I_4 &= \sum_{w=0}^{n} (w-2^{m-1})^4 a_w - 2^{4m-3} \\
&= 2^{k-4}(n(n+1)(n^2+5n-2)+4!b_4) \\
&\quad - 2^{m+k-2}(n^2(n+3)) + 3.2^{2m+k-3} n(n+1) \\
&\quad - 2^{3m+k-2} n + 2^{4m-4} 2^k - 2^{4m-3}
\end{aligned}$$

w	δ	2^{m-1}	$2^m - \delta$
a_w	$\dfrac{2^{2m-2}}{(\delta - 2^{m-1})^2}$	$2^{m+2} - \dfrac{2^{2m-1}}{(\delta - 2^{m-1})^2} - 2$	$\dfrac{2^{2m-2}}{(\delta - 2^{m-1})^2}$

which finally gives, replacing k by $m+2$ and n by 2^m

$$I_4 = 2^{m-2}\left(3.2^{2m} - 2^{m+1} - 2^{3m-1} + 4!\,b_4\right). \tag{17}$$

Since the codeword $\mathbf{1}$ belongs to C, we have $a_w = a_{n-w}$, for all $0 \le w \le n$, and thus,

$$I_\ell = \sum_{w=1}^{2^{m-1}-1} \left((w-2^{m-1})^\ell + (-1)^\ell (w-2^{m-1})^\ell\right) a_w.$$

Then

$$I_\ell = \begin{cases} 0, & \text{for odd } \ell \\ 2\displaystyle\sum_{w=1}^{2^{m-1}-1} (w-2^{m-1})^\ell a_w, & \text{for even } \ell. \end{cases} \tag{18}$$

Thus, we have $I_4 \ge 0$, and from (17) we deduce

$$\begin{aligned} 0 &\le 3.2^{2m-3} - 2^{m-2} - 2^{3m-4} + 3.b_4 \\ &\le 2^{m-2}(3.2^{m-1} - 1 - 2^{2m-2}) + 3.b_4. \end{aligned}$$

Thus, b_4 must satisfy

$$b_4 \ge \frac{1}{3}\,2^{m-2}\left((2^{m-1}-1)^2 - 2^{m-1}\right)$$

which implies $b_4 > 0$—i.e., the minimum distance of $C^\perp$ is 4.

Now we compute $I_4 - \lambda^2 I_2$. On one hand, we have by (18)

$$I_4 - \lambda^2 I_2 = 2\mathcal{I}(\lambda). \tag{19}$$

On the other hand, we express $I_4 - \lambda^2 I_2$ by means of (16) and (17). Therefore, we deduce from (19)

$$\begin{aligned} \mathcal{I}(\lambda) &= 3.2^{3m-3} - 2^{2m-2} - 2^{4m-4} + 3.b_4 2^m - 2^{2m-2}\lambda^2 \\ &= 2^m\left(2^{m-2}(3.2^{m-1} - 1 - 2^{2m-2} - \lambda^2) + 3.b_4\right) \\ &= 2^m\left(3.b_4 - 2^{m-2}((2^{m-1}-1)^2 - 2^{m-1} + \lambda^2)\right). \end{aligned}$$

Equation (13) implies that the quantity $\mathcal{I}(\lambda)$ consists of a sum of terms T_w, $1 \le w \le 2^{m-1}-1$, with $T_w \le 0$ for every w such that $(w-2^{m-1})^2 \le \lambda^2$. If $\lambda \ge 2^{m-1} - \delta$ then $T_w \le 0$ for every nonzero weight w, since $2^{m-1} - w \le 2^{m-1} - \delta$. Thus, if $\lambda \ge 2^{m-1} - \delta$ then $\mathcal{I}(\lambda) \le 0$, which exactly corresponds to the inequality (15).

Moreover, equality holds in (15) if and only if the values of w such that $a_w \ne 0$ lie in $\{2^{m-1}, 2^{m-1} \pm \lambda\}$ (i.e., $\mathcal{I}(\lambda) = 0$). Then $\lambda = 2^{m-1} - \delta$. We obtain the values a_δ by computing I_2 by means of (16) and (18). □

We now come back to the code $C_{\mathrm{x}} = (\mathrm{x} + R(1, m)) \cup R(1, m)$, $\mathrm{x} \notin R(1, m)$. Note that such a code satisfies the hypothesis of the previous theorem. Indeed, the code contains the all-one vector and, denoting by a_w (resp., b_w), $w \in [0, n]$, the number of codewords of C_{x} (resp., $C_{\mathrm{x}}^\perp$) of weight w, we have the following proposition.

Proposition IV.1: The code $C_{\mathrm{x}}^\perp$ is contained in $R(m-2, m)$; thus, we have $b_1 = b_2 = b_3 = 0$. The codewords of C_{x} which have weight 4 are the indicators of two-dimensional affine subspaces of $\boldsymbol{F}_2^m$.

Proof: This result comes from well-known properties of Reed–Muller codes: $R(1, m)^\perp = R(m-2, m)$ is the extended Hamming code and has minimum weight 4. The codewords of weight 4 have the form

$$\mathrm{y} = X^a + X^{a+b} + X^{a+c} + X^{a+b+c}, \quad \text{with } a, b, c \in \boldsymbol{F}_2^m,\ b \ne c \ne 0. \tag{20}$$

Their supports are two-dimensional affine subspaces (see [22, Ch. 13, Theorems 4 and 5]). Since $R(1, m) \subset C_{\mathrm{x}}$, then $C_{\mathrm{x}}^\perp \subset R(m-2, m)$, completing the proof. □

We focus here on the weight enumerators of cosets of $R(1, m)$ whose minimum weights are near the optimal value. Two values of λ are of most interest: $2^{m/2}$ for m even and $2^{(m-1)/2}$ for m odd, corresponding to the following kinds of cosets.

Definition IV.1: A coset of $R(1, m)$ is said to be *almost-optimal* if its minimum weight is greater than or equal to w_0, where $w_0 = 2^{m-1} - 2^{(m-1)/2}$ for odd m, and $w_0 = 2^{m-1} - 2^{m/2}$ for even m. It is said to be *three-valued* when it has exactly three nonzero weights.

Proposition IV.2: A coset of $R(1, m)$ is three-valued almost-optimal if and only if its weight distribution is

w	$2^{m-1} - 2^{(m-1)/2}$	2^{m-1}	$2^{m-1} + 2^{(m-1)/2}$
a_w	2^{m-1}	2^m	2^{m-1}

for odd m and

w	$2^{m-1} - 2^{m/2}$	2^{m-1}	$2^{m-1} + 2^{m/2}$
a_w	2^{m-2}	3.2^{m-1}	2^{m-2}

for even m.

Proof: Suppose that a coset $\mathrm{x}+R(1, m)$ has three weights only. Clearly, these weights lie in $\{2^{m-1}, \delta, 2^m - \delta\}$. Combining (18) and (16), we obtain

$$\sum_{w=\delta}^{2^{m-1}-1} (w-2^{m-1})^2 a_w = (2^{m-1}-\delta)^2 a_\delta = 2^{2m-2}.$$

Thus, $2^{m-1} - \delta$ is a power of 2. Assume that $\mathrm{x} + R(1, m)$ is almost-optimal. Then the only possibility for m odd is $\delta = 2^{m-1} - 2^{(m-1)/2}$. When m is even, the only possibility for the coset to have exactly three weights is $\delta = 2^{m-1} - 2^{m/2}$. □

Consider the notation of Theorem IV.1. By replacing C by C_{x} we obtain the following necessary condition on three-valued almost-optimal cosets.

Corollary IV.1: If the coset $\mathrm{x} + R(1, m)$ is almost-optimal, then we have

- if m is odd, then $b_4 \le \frac{1}{3}\,2^{m-2}\,(2^{m-1}-1)^2$;
- if m is even, then $b_4 \le \frac{1}{3}\,(2^{m-2}\,(2^{m-1}-1)^2 + 2^{2m-3})$.

In both cases, equality holds if and only if $\mathrm{x}+R(1, m)$ is three-valued almost-optimal.

Proof: We simply apply Theorem IV.1.

- If m is odd, we set $\lambda = 2^{(m-1)/2}$. As the coset is almost-optimal, $\delta \ge 2^{m-1} - 2^{(m-1)/2}$.

- If m is even, we set $\lambda = 2^{m/2}$. As the coset is almost-optimal, $\delta \geq 2^{m-1} - 2^{m/2}$(where δ is the minimum distance of C_{x}—i.e., the minimum weight of the coset). □

Remark IV.1:

1) By taking $\lambda = 2^{(m-2)/2}$ with m even, we obtain cosets whose minimum weight is $\rho(R(1,\ m))$ only. These cosets have two weights, $2^{m-1} \pm 2^{(m-2)/2}$, and correspond to the bent functions. Moreover,
$$b_4 = \tfrac{1}{3}\, 2^{m-2} \left((2^{m-1} - 1)^2 - 2^{m-2} \right).$$
2) It is quite easy to construct three-valued almost-optimal cosets. Although these cosets are not yet classified, they are completely known when $\mathrm{x} \in R(2,\ m)$ (see [22, Ch. 14] and a short presentation in Appendix I). We will give other examples in the next section.
3) Assume that $\mathrm{x} + R(1,\ m)$ is almost-optimal and that the upper bound on b_4 is not reached. Very little is known about these cosets and several hard open problems are involved, as the covering radius of $R(1,\ m)$ for m odd, or the covering radius of $R(1,\ m)$ restricted to codewords of weight 2^{m-1} for any m. Examples of such cosets can be found in [23] for $m = 5$ and in [12] for $m < 11$.

Open Problem IV.1: The dual of the code C_{x} is a subspace of $R(m-2,\ m)$ of codimension 1. How can a subspace containing *few codewords of weight* 4 be constructed?

B. Computing b_4

Let us denote by $\mathcal{W}$ the set of all affine subspaces of $\boldsymbol{F}_2^m$ of dimension 2 and by $\mathcal{W}_0$ the subset of $\mathcal{W}$ of all *linear subspaces* of $\boldsymbol{F}_2^m$. Recall that $C_{\mathrm{x}} = (\mathrm{x} + R(1,\ m)) \cup R(1,\ m), \mathrm{x} \notin R(1,\ m)$, and that b_4 is the number of codewords of weight 4 in $C_{\mathrm{x}}^{\perp}$. In this subsection, we want to be more explicit about the computation of b_4. We later apply our results to the cosets which are contained in $R(3,\ m)$.

The codewords of weight 4 in $C_{\mathrm{x}}^{\perp}$ are of type (20). These codewords have as support an element V of $\mathcal{W}$; they belong to $R(m-2,\ m)$ (i.e., $\mathcal{P}^2$). In this section, we will denote by y^V such a codeword. Let $a,\ b,\ c \in \boldsymbol{F}_2^m,\ b \neq c \neq 0$, and $V = \{a,\ a+b,\ a+c,\ a+b+c\}$. Then
$$\mathrm{y}^V = \sum_{v \in V} X^v = X^a (X^b + 1)(X^c + 1). \tag{21}$$
The next result is a direct application of Lemma III.1.

Proposition IV.3: For any $V \in \mathcal{W}_0$ we denote by $\mathrm{x}_i,\ 1 \leq i \leq 2^{m-2}$, the restrictions of x to the cosets of V. Then the number b_4 of codewords of weight 4 in $C_{\mathrm{x}}^{\perp}$ can be expressed as follows:
$$b_4 = \sum_{V \in \mathcal{W}_0} \frac{2^m - \mathrm{wt}\,(\mathrm{y}^V \mathrm{x})}{4}$$
where the codeword y^V is defined by (21). Moreover,
$$\mathrm{wt}\,(\mathrm{y}^V \mathrm{x}) = 4 \times \#\{i \,|\, \mathrm{wt}\,(\mathrm{x}_i) \text{ is odd}\}.$$

Proof: Let $V \in \mathcal{W}_0$. Since $\mathrm{y}^V \in R(m-2,\ m)$, $\mathrm{y}^V \in C_{\mathrm{x}}^{\perp}$ if and only if y^V is orthogonal to x—i.e., the weight of the restriction x_1 of x to V is even.

In accordance with Lemma III.1, we have
$$\mathrm{wt}\,(\mathrm{y}^V \mathrm{x}) = \#\{i \,|\, \mathrm{wt}\,(\mathrm{x}_i) \text{ is odd}\} \times 4$$
where the x_i's are the restrictions of x to the cosets of V. This implies that the number of cosets $h + V$ such that $\mathrm{y}^{h+V} \in C_{\mathrm{x}}^{\perp}$ is equal to $(2^m - \mathrm{wt}\,(\mathrm{y}^V \mathrm{x}))/4$. The value of b_4 is obtained by considering all $V \in \mathcal{W}_0$. □

When there are few possible values for $\mathrm{wt}\,(\mathrm{y}^V \mathrm{x})$, the expression of b_4 becomes simpler. It is especially the case when x is in $R(3,\ m)$.

Corollary IV.2: Let x be in $R(3,\ m) \backslash R(1,\ m)$. Let us define

- $N_0 = \#\{V \in \mathcal{W}_0 | \mathrm{y}^V \mathrm{x} = \mathbf{0}\}$
- $N_b = \#\{V \in \mathcal{W}_0 | \mathrm{wt}\,(\mathrm{y}^V \mathrm{x}) = 2^{m-1}\}$.

Then $b_4 = 2^{m-2} N_0 + 2^{m-3} N_b$. We have $N_b = 0$ when $\mathrm{x} \in R(2,\ m)$.

Proof: As $\mathrm{x} \in R(3,\ m)$ and $\mathrm{y}^V \in R(m-2,\ m)$, $\mathrm{y}^V \mathrm{x}$ is in $R(1,\ m)$—since
$$\mathcal{P}^2 \mathcal{P}^{m-3} = \mathcal{P}^{m-1}.$$
So $\mathrm{wt}\,(\mathrm{y}^V \mathrm{x})$ belongs to $\{0,\ 2^{m-1},\ 2^m\}$. When $\mathrm{x} \in R(2,\ m)$, we have $\mathrm{y}^V \mathrm{x} \in \{\mathbf{0},\ \mathbf{1}\}$ implying $N_b = 0$.

In accordance with Proposition IV.3, we obtain
$$b_4 = \sum_{\substack{V \in \mathcal{W}_0 \\ \mathrm{y}^V \mathrm{x} = 0}} 2^{m-2} + \sum_{\substack{V \in \mathcal{W}_0 \\ \mathrm{wt}\,(\mathrm{y}^V \mathrm{x}) = 2^{m-1}}} 2^{m-3} = N_0 2^{m-2} + N_b 2^{m-3}.$$
□

Remark IV.2: Note that the weight enumerators of the cosets $\mathrm{x} + R(1,\ m)$ with $\mathrm{x} \in R(2,\ m)$ are known. For cosets which are not contained in $R(2,\ m)$, the weight enumerators are generally not known. The study of such cosets contained in $R(3,\ m)$ is the *first* open problem. In this paper, we point out that these cosets have specific properties. However, it seems difficult to strengthen any conjecture.

Corollary IV.3: Let $R_{\mathrm{x}} = \mathrm{x} + R(1,\ m)$ with $\mathrm{x} \in R(3,\ m) \setminus R(1,\ m)$. Let
$$N_1 = \#\{V \in \mathcal{W}_0 \,|\, \mathrm{y}^V \mathrm{x} = \mathbf{1}\}$$
and N_0 has been defined in the previous corollary. If R_{x} is almost-optimal, we have

- if m is odd, then $N_1 - N_0 \geq \frac{2^{m-1}-1}{3}$.
- if m is even, then $N_0 - N_1 \leq \frac{2^{m-1}+1}{3}$.

In both cases, equality holds if and only if R_{x} is three-valued almost-optimal.

Proof: Recall that
$$\#\mathcal{W}_0 = \tfrac{1}{3}(2^m - 1)(2^{m-1} - 1) = N_0 + N_1 + N_b \tag{22}$$
implying
$$\begin{aligned} b_4 &= 2^{m-2} N_0 + 2^{m-3} N_b \\ &= 2^{m-3} \left(N_0 - N_1 + \frac{(2^m - 1)(2^{m-1} - 1)}{3} \right). \end{aligned}$$
Suppose that R_{x} is almost-optimal. According to Corollary IV.1 we obtain the expected bounds. □

For m even, it is easy to find cosets, defined as above, satisfying $N_0 - N_1 < (2^{m-1}+1)/3$ (see the next example). In the case where m is odd it is not so easy to find cosets satisfying $N_1 - N_0 > (2^{m-1}-1)/3$. Actually, the existence of such cosets is just proved by Canteaut in [24]; she exhibits almost-optimal cosets with five weights which are contained in $R(3, 9)$. These weights are $2^8 \pm 2^4$, $2^8 \pm 2^3$ and 2^8. However, the determination of the minimum weights of such cosets remains an open problem for $m \leq 13$ (see the end of Section V-A for more explanations).

Example IV.1: Let $m = 6$ and

$$f(x_1, \ldots, x_6) = x_1x_2x_3 + x_1x_2x_4 + x_1x_2 + x_3x_4 + x_5x_6.$$

The weight distribution of the coset $\Omega_f + R(1, 6)$ is

$$a_{24} = a_{40} = 8, \quad a_{28} = a_{36} = 32, \quad \text{and} \quad a_{32} = 48.$$

This coset is almost-optimal with five weights.

V. The Propagation Criterion and the Nonlinearity

We come back to the terminology of Boolean functions but we will always consider together a given function f of m variables and its associated binary codeword Ω_f. So we first fix the terminology for functions which generate a coset $\Omega_f + R(1, m)$ with a high minimum weight (see Definition IV.1 and the following proposition).

Definition V.1: The Boolean function f is said to be almost-optimal if its associated coset $\Omega_f + R(1, m)$ is almost-optimal or equivalently if

- $\mathcal{L}(f) \leq 2^{(m+2)/2}$, when m is even;
- $\mathcal{L}(f) \leq 2^{(m+1)/2}$, when m is odd.

The function f is said to be three-valued almost-optimal if its associated coset is three-valued almost-optimal—i.e., its Fourier spectrum is $\{0, \pm 2^{(m+2)/2}\}$ when m is even and $\{0, \pm 2^{(m+1)/2}\}$ when m is odd.

Recall the definition of the so-called $PC(\ell)$ property.

Definition V.2: Let $\boldsymbol{e} = (e_1, \ldots, e_m)$ be a basis of $\boldsymbol{F}_2^m$. Then f satisfies the *propagation criterion of order* ℓ (PC(ℓ)), *with respect to* $\boldsymbol{e}$ if, for any vector $a = (a_1, \ldots, a_m)$ in $\boldsymbol{F}_2^m$ such that $0 < \mathrm{wt}(a) \leq k$

$$D_a f, \; a = \sum_{i=1}^{m} a_i e_i$$

is balanced.

A. Bounds on the Sum-of-Squares Indicator

From now on, we focus on almost-optimal functions $f \in \mathcal{B}_m$, $f \notin R(1, m)$, $m \geq 3$. Notation is the same as in Theorem IV.1 and its proof: we consider the code

$$C_{\mathrm{x}} = (\mathrm{x} + R(1, m)) \cup R(1, m)$$

with $\mathrm{x} = \Omega_f$; a_w denotes the number of codewords of weight w in C_{x}, and $\mathcal{I}(\lambda)$ is defined by (13). Recall that the sum-of-squares indicator $\mathcal{V}(f)$ allows to measure the global avalanche criterion of f (see Section II-A, (6)). The next propositions are in fact corollaries of Theorem IV.1; our aim is to make explicit the link between two points of view (in terms of codewords and in terms of functions).

Lemma V.1: Let $\lambda \leq 2^{m-1}$ be any positive integer. Then

$$\mathcal{I}(\lambda) = 2^{m-4}(\mathcal{V}(f) - \lambda^2 2^{m+2}).$$

Thus, $\mathcal{I}(\lambda) \leq 0$ if and only if $\mathcal{V}(f) \leq \lambda^2 2^{m+2}$.

Proof: From (19), we have

$$\mathcal{I}(\lambda) = \sum_{w=\delta}^{2^{m-1}-1} (w - 2^{m-1})^2[(w - 2^{m-1})^2 - \lambda^2]a_w$$
$$= \frac{1}{2}\left[I_4 - \lambda^2 I_2\right].$$

According to (18) we obtain

$$I_4 = 2\sum_{w=1}^{2^{m-1}-1} (w - 2^{m-1})^4 a_w = \frac{1}{8}\sum_{\alpha \in \boldsymbol{F}_2^m} \mathcal{F}^4(f + \varphi_\alpha) \quad (23)$$

since $\mathcal{F}^4(f + \varphi_\alpha) = (2^m - 2w)^4$ where w is the weight of $\Omega_{f+\varphi_\alpha}$, and

$$a_w = |\{\alpha \in \boldsymbol{F}_2^m \,|\, \mathrm{wt}(\Omega_{f+\varphi_\alpha}) = w \text{ or } 2^m - w\}|.$$

Moreover, $I_2 = 2^{2m-1}$.

It follows that

$$\mathcal{I}(\lambda) = \frac{1}{16}\left[\sum_{\alpha \in \boldsymbol{F}_2^m} \mathcal{F}^4(f + \varphi_\alpha) - \lambda^2 2^{2m+2}\right].$$

Using (7), we deduce that

$$\mathcal{I}(\lambda) = 2^{m-4}\left[\mathcal{V}(f) - \lambda^2 2^{m+2}\right]$$

completing the proof. □

Proposition V.1: Let $f \in \mathcal{B}_m$ and $\mathrm{x} = \Omega_f$. Let b_w denote the number of codewords of weight w in $C_{\mathrm{x}}^{\perp}$. Then

$$b_4 = \frac{1}{48}\left(\mathcal{V}(f) + 2^{m+2}\left((2^{m-1}-1)^2 - 2^{m-1}\right)\right).$$

Proof: We simply consider together the formula given in the previous lemma and (14). So

$$\mathcal{V}(f) - \lambda^2 2^{m+2}$$
$$= 2^4\left(3b_4 - 2^{m-2}\left((2^{m-1}-1)^2 + (\lambda^2 - 2^{m-1})\right)\right).$$

Hence,

$$\mathcal{V}(f) = 2^4\left(3b_4 - 2^{m-2}\left((2^{m-1}-1)^2 - 2^{m-1}\right)\right)$$

completing the proof. □

Proposition V.2: Let m be a positive integer, $m \geq 3$, and $f \in \mathcal{B}_m$. Assume that f is almost-optimal. Then

- if m is odd then $\mathcal{V}(f) \leq 2^{2m+1}$ with equality if and only if f is three-valued almost-optimal;
- if m is even then $\mathcal{V}(f) \leq 2^{2m+2}$ with equality if and only if f is three-valued almost-optimal.

Proof: Since f is almost-optimal, the minimum weight δ of the coset $\Omega_f + R(1, m)$ satisfies $\delta \geq 2^{m-1} - 2^{(m-1)/2}$ for odd m and $\delta \geq 2^{m-1} - 2^{m/2}$ for even m. According to

Theorem IV.1, this implies $\mathcal{I}(2^{(m-1)/2}) \leq 0$ for odd m and $\mathcal{I}(2^{m/2}) \leq 0$ for even m.

From Lemma V.1, replacing λ by either $2^{(m-1)/2}$ or $2^{m/2}$ (depending on whether m is odd or even), we immediately deduce the expected bounds on $\mathcal{V}(f)$. □

Example V.1: There are many three-valued almost-optimal functions. The *almost-bent* functions provide such functions (see, for instance, [25]–[28]). Any three-valued almost-optimal partially bent function is linearly equivalent to (see Proposition II.6)

$$f(x_1, \ldots, x_m) = g(x_1, \ldots, x_{m-\ell}) + \varphi_\alpha(x_1, \ldots, x_m)$$

where g is bent and $\ell = 1$ for odd m and $\ell = 2$ for even m. For these functions, $\mathcal{L}(f) = 2^{(m+\ell)/2}$ and $\mathcal{V}(f) = 2^{2m+\ell}$.

It is easy to find almost-optimal functions such that $\mathcal{L}(f) = 2^{(m+\ell)/2}$ and $\mathcal{V}(f) < 2^{2m+\ell}$, where ℓ is defined as above. These functions have a good (generally not the best) nonlinearity but are not three-valued (see Example IV.1, a number of numerical results in [29], [12] and Proposition V.5).

It is not so easy to obtain almost-optimal functions such that $\mathcal{L}(f) < 2^{(m+\ell)/2}$ (implying $\mathcal{V}(f) < 2^{2m+\ell}$ according to Proposition V.2). The class of bent functions seems to be the only known large class. Numerical results are easily obtained for m even (see [30], [12]). When m is odd, the only known such functions are obtained from those given in [31] for $m = 15$.

Note that there exist non-almost-optimal functions f such that $\mathcal{V}(f) \leq 2^{2m+\ell}$.

Example V.2: For $m = 5$ one finds in [23] the function

$$f(x_1 \ldots, x_5) = x_1x_2x_3x_4x_5 + x_1x_2x_3 + x_1x_4x_5 + x_4x_5 + x_3x_5 + x_2x_4 + x_2x_3.$$

It generates a coset of $R(1, m)$ with weight distribution

$$a_{11} = a_{21} = 4, \quad a_{13} = a_{19} = 16, \quad \text{and} \quad a_{15} = a_{17} = 12$$

$a_w = 0$ otherwise. Thus, $\mathcal{L}(f) = 10$. Using (7), we obtain $\mathcal{V}(f) = 1904$ which is strictly less than $2^{11} = 2048$.

Let $f \in \mathcal{B}_m$ be a function of degree d. Set the notation

$$E_f = \{e \in \boldsymbol{F}_2^m \,|\, D_e f \text{ is balanced}\} \quad \text{and} \quad \overline{E_f} = \boldsymbol{F}_2^m \setminus E_f.$$

In [10, Proposition 14], we have stated the following relation between the cardinality of $\overline{E_f}$, denoted by $\#\overline{E_f}$, and the value of $\mathcal{V}(f)$.

Proposition V.3: Let $f \in \mathcal{B}_m$ be a function of degree d. Then

$$\mathcal{V}(f) \geq 2^{2m} + \left(\#\overline{E_f} - 1\right) 2^{2\lfloor \frac{m-2}{d-1} \rfloor + 4}.$$

This is of most interest for functions of degree 3. We obviously obtain from the previous result and from Proposition V.2 the following corollary.

Corollary V.1: Let $f \in \mathcal{B}_m$ a function of degree 3. So the following properties hold.

i) When m is even then $\mathcal{V}(f) \geq 2^{2m} + (\#\overline{E_f} - 1)2^{m+2}$. Thus, if $\mathcal{L}(f) \leq 2^{(m+2)/2}$ then $\#E_f \geq 2^{m-2} - 1$.

ii) When m is odd then $\mathcal{V}(f) \geq 2^{2m} + (\#\overline{E_f} - 1)2^{m+1}$. Thus, if $\mathcal{L}(f) \leq 2^{(m+1)/2}$ then $\#E_f \geq 2^{m-1} - 1$.

Therefore, we point out that for almost-optimal functions of degree 3, the rank of E_f must be *high*. Note that we call *rank* of E_f the dimension of the subspace generated by the elements of E_f (remark that $E_f \cup \{0\}$ is not, in general, a subspace).

Corollary V.2: An almost-optimal function of degree 3 is such that the rank of E_f is at least $m - 2$ for even m and at least $m - 1$ for odd m.

When m is odd, such a function is PC (1), unless $E_f \cup \{0\}$ is a subspace of codimension 1. In this case, f is three-valued almost-optimal.

Proof: If E_f is a set of rank k then its cardinality is at most $2^k - 1$ (E_f does not contain 0). Assume that f is almost-optimal. Clearly, Corollary V.1 provides the lower bounds $m-2$ (m even) and $m - 1$ (m odd) for k. When m is odd, k is either $m - 1$ or m.

Assume that m is odd. If $k = m - 1$ then $\#E_f = 2^{m-1} - 1$ (since $E_f \cup \{0\}$ is a subspace of codimension 1), implying $\mathcal{V}(f) = 2^{2m+1}$ thanks to Corollary V.1 and Proposition V.2. In accordance with Proposition V.2, f is three-valued almost-optimal; note that it can be proved by another way, using Theorem V.2 of Section V-C.

When $k = m$, it means that there exists a basis of $\boldsymbol{F}_2^m$, say $\boldsymbol{e} = (e_1, \ldots, e_m)$, such that $e_i \in E_f$ and $D_{e_i} f$ is balanced, for all i; so f is PC (1), with respect to $\boldsymbol{e}$. □

Note that it is very easy to construct almost-optimal functions of degree 3, which are three-valued. It is more difficult to construct such functions which are almost-optimal and not three-valued, especially when m is odd—as we indicated in other terms at the end of Section IV-B. Moreover, the general problem of the maximal nonlinearity of functions of degree 3 remains open for odd m.

It is known that, for any odd $m \leq 13$, all almost-optimal functions f of degree 3 satisfy $\mathcal{L}(f) = 2^{(m+1)/2}$ [32]. It has been recently proved by Canteaut that, for any odd $m \leq 7$, all almost-optimal functions of degree 3 are three-valued. For $m = 9$, she has proved that there is only one weight polynomial for almost-optimal non-three-valued cosets of $R(1, m)$ which are contained in $R(3, 9)$; moreover, she proves that such cosets exist [24].

Open Problem V.1: For odd m, $m > 13$, does there exist $f \in \mathcal{B}_m$ of degree 3 such that f is almost-optimal and $\mathcal{L}(f) < 2^{(m+1)/2}$?

B. Decompositions on Affine Subspaces of $\boldsymbol{F}_2^m$

We are going to study the restrictions of $f \in \mathcal{B}_m$ to any subspace W of $\boldsymbol{F}_2^m$. Lemma V.2 is derived from well-known properties of the Fourier transform.

Lemma V.2: Let f be a Boolean function of m variables and let V be a subspace of $\boldsymbol{F}_2^m$ of dimension k. Then we have, for any $\beta \in \boldsymbol{F}_2^m$

$$\sum_{\alpha \in V} \mathcal{F}^2(f + \varphi_{\alpha+\beta}) = 2^k \sum_{e \in V^\perp} (-1)^{\beta \cdot e} \mathcal{F}(D_e f).$$

Proof: According to (3), we have for any $\alpha \in \boldsymbol{F}_2^m$

$$\mathcal{F}^2(f+\varphi_\alpha) = \sum_{e\in\boldsymbol{F}_2^m} (-1)^{\alpha\cdot e}\mathcal{F}(D_e f).$$

We deduce that, for any $\beta \in \boldsymbol{F}_2^m$

$$\begin{aligned}\sum_{\alpha\in V} \mathcal{F}^2(f+\varphi_{\alpha+\beta}) &= \sum_{\alpha\in V}\sum_{e\in\boldsymbol{F}_2^m} (-1)^{(\alpha+\beta)\cdot e}\mathcal{F}(D_e f)\\ &= \sum_{e\in\boldsymbol{F}_2^m} (-1)^{\beta\cdot e}\mathcal{F}(D_e f)\sum_{\alpha\in V}(-1)^{\alpha\cdot e}\\ &= 2^k \sum_{e\in V^\perp} (-1)^{\beta\cdot e}\mathcal{F}(D_e f). \qquad \square\end{aligned}$$

Remark V.1: Note that for $V = \boldsymbol{F}_2^m$, Lemma V.2 provides the well-known formula of Parseval. When $V^\perp = \{0, a\}$, we get the following relation:

$$\sum_{\alpha\in V} \mathcal{F}^2(f+\varphi_{\alpha+\beta}) = 2^{m-1}\left(2^m + (-1)^{a\cdot\beta}\mathcal{F}(D_a f)\right),$$

for all $\beta \in \boldsymbol{F}_2^m$.

We need to define precisely the restrictions of any $f \in \mathcal{B}_m$ to a subspace W, of dimension k, and to the cosets of W. Let such a coset $W' = a + W$, $a \notin W$. The restriction of f to W' can be identified with $h \in \mathcal{B}_k$ such that $h(x) = f(a+x)$. This representation depends, in fact, on the choice of $a \in W'$ since for $b = a + u$, $u \in W$, we have $h'(x) = f(b+x) = h(u+x)$ (h' is a *translation* of h). However, in the context of our study, h and h' have the same properties. So when we say *the decomposition of* f (as defined below) we mean that, for a fixed W, the restrictions are chosen up to translations.

Definition V.3: Let W be a subspace of $\boldsymbol{F}_2^m$ of dimension k. *The decomposition of f with respect to W* is the sequence $\{h_a | a \in V\}$ where V is such that $\boldsymbol{F}_2^m$ is the direct sum of W and V and h_a is the Boolean function of k variables, from W to $\boldsymbol{F}_2$, defined by $h_a(x) = f(a+x)$ for any $x \in W$.

Theorem V.1: Let W be a subspace of $\boldsymbol{F}_2^m$ of dimension k and let $(h_a | a \in V)$ be the decomposition of f with respect to W. Then

$$\sum_{\alpha\in W^\perp} \mathcal{F}^2(f+\varphi_\alpha) = 2^{m-k}\sum_{a\in V}\mathcal{F}^2(h_a).$$

Proof: Consider Ω_f the associated codeword of f. We have

$$\Omega_f = \sum_{a\in V} X^a \Omega_{h_a}. \tag{24}$$

We obviously deduce

$$\mathrm{wt}(\Omega_f) = \sum_{a\in V}\mathrm{wt}(\Omega_{h_a}).$$

Note the extension of this property to $D_\beta f$, for any $\beta \in W$. Indeed, we have for such a β

$$\Omega_{D_\beta f} = \sum_{a\in V} X^a \Omega_{D_\beta h_a}.$$

Thus,

$$\begin{aligned}\mathcal{F}(D_\beta f) = 2^m - 2\mathrm{wt}(\Omega_{D_\beta f}) &= \sum_{a\in V}\left(2^{m-k} - 2\mathrm{wt}(\Omega_{D_\beta h_a})\right)\\ &= \sum_{a\in V}\mathcal{F}(D_\beta h_a).\end{aligned}$$

Set $L = \sum_{\alpha\in W^\perp}\mathcal{F}^2(f+\varphi_\alpha)$. According to Lemma V.2 and to the above formula, we have

$$\begin{aligned}L &= 2^{m-k}\sum_{\beta\in W}\mathcal{F}(D_\beta f) = 2^{m-k}\sum_{\beta\in W}\left(\sum_{a\in V}\mathcal{F}(D_\beta h_a)\right)\\ &= 2^{m-k}\sum_{a\in V}\left(\sum_{\beta\in W}\mathcal{F}(D_\beta h_a)\right) = 2^{m-k}\sum_{a\in V}\mathcal{F}^2(h_a)\end{aligned}$$

according to (4). $\square$

Corollary V.3: Let $(h_a | a \in V)$ be the decomposition of f with respect to the k-dimensional subspace W. Then

$$\sum_{a\in V}\mathcal{F}^2(h_a) \le \mathcal{L}^2(f).$$

Moreover, $\mathcal{L}(h_a) \le \mathcal{L}(f)$, for all $a \in V$.

Proof: According to Theorem V.1 and since $|W^\perp| = 2^{m-k}$, we obviously deduce

$$2^{m-k}\sum_{a\in V}\mathcal{F}^2(h_a) \le 2^{m-k}\mathcal{L}^2(f)$$

implying $\mathcal{F}^2(h_a) \le \mathcal{L}^2(f)$ for every a. Moreover, this property holds if we replace h_a by $h_a + \ell$, where ℓ is any linear function of $\mathcal{B}_k$—considering the decomposition of $f + \varphi_\beta$, for some $\beta \in \boldsymbol{F}_2^m$, instead of f. Hence, $\mathcal{L}(h_a) \le \mathcal{L}(f)$ for all a, completing the proof. $\square$

Remark V.2: We have

$$\begin{aligned}&\mathcal{N}(f) - \mathcal{N}(h_a)\\ &\quad = 2^{m-1} - 2^{k-1} + (\mathcal{L}(h_a) - \mathcal{L}(f))/2 \le 2^{m-1} - 2^{k-1}\end{aligned}$$

since $\mathcal{L}(h_a) - \mathcal{L}(f) \le 0$. This upper bound on $\mathcal{N}(f) - \mathcal{N}(h_a)$ was already proved by Zheng *et al.* in [8]. The authors noticed that when m is odd, $k = (m+1)/2$ and h_a is an affine function, then $\mathcal{N}(f) \le 2^{m-1} - 2^{(m-1)/2}$.

Notice that, when m is even, $k = m/2$ and h_a affine, we find again the covering radius of $R(1, m)$.

The previous results provide the exact connection between the nonlinearity of f and the nonlinearity of each element of any decomposition of f—"any" means "with respect to W, for any W." The well-known conjecture of Dobbertin has to be placed in this context. In [30], he introduced the notion of *normal* function for even m. A function $f \in \mathcal{B}_m$ is said to be normal if it is constant on at least one $m/2$-dimensional flat. He proposed the next conjecture.

Conjecture. *Any bent function is normal.*

The link between the nonlinearity of a function and the nonlinearity of each element of its decomposition has several consequences. For instance, when f is almost-optimal, any function h of any decomposition of f is such that $\mathcal{L}(h) \le 2^{(m+1)/2}$ for odd m, and $\mathcal{L}(h) \le 2^{(m+2)/2}$ for even m. This notably leads to the following property.

Proposition V.4: Assume that m is odd. Suppose that W has codimension 1. For simplicity, we denote by (h_1, h_2) the decomposition of f with respect to W.

If h_1 (or h_2) is partially bent and not bent then $\mathcal{L}(f) \geq 2^{(m+1)/2}$.

Proof: If h_1 is partially bent then $\mathcal{L}(h_1) = 2^\rho$, where ρ is an integer such that $\rho \geq (m-1)/2$, with equality if and only if h_1 is bent.

If h_1 is not bent we have, in accordance with Corollary V.3

$$\mathcal{L}(f) \geq \mathcal{L}(h_1) \geq 2^{(m+1)/2}$$

completing the proof. □

Remark V.3: The function with five variables given in Example V.6

$$g(x_1, \ldots, x_5) = x_3x_5 + x_2x_4 + x_1x_2x_3 + x_2x_3x_4x_5$$

is almost-optimal (not three-valued) of degree 4. It satisfies the hypothesis of Proposition V.4, since one element of its decomposition with respect to the hyperplane

$$\{(x_1, \ldots, x_5) \in \boldsymbol{F}_2^m, x_2 = 0\}$$

is quadratic

$$g(x_1, \ldots, x_5) = (1 + x_2)x_3x_5 + x_2(x_4 + x_3x_5 + x_1x_3 + x_3x_4x_5).$$

This proves that the class of such functions is interesting.

Example V.3: It is very easy to construct a function f satisfying the hypothesis of Proposition V.4, with algebraic normal form equal, up to equivalence, to

$$f(x_1, \ldots, x_m) = g(x_1, \ldots, x_{m-1}) + x_m h(x_1, \ldots, x_{m-1}) \tag{25}$$

where g has degree 2.

Let $m = 7$. The functions

$$f(x_1, \ldots, x_7) = x_1x_2 + x_3x_4 + x_7h(x_1, \ldots, x_6)$$

where h is any function in $\mathcal{B}_6$, satisfy $\mathcal{L}(f) \geq 2^{(7+1)/2}$. Indeed, it is well known that $\mathcal{L}(x_1x_2 + x_3x_4) = 2^4$ (see Appendix I and [22, Ch. 15, Sec. 2]).

C. Derivatives on Subspaces of Large Dimensions

Now we are considering the cases where the derivatives D_ef of a given function f are balanced for any $e \neq 0$ belonging to a subspace of codimension 1 or 2. This allows us to obtain a new characterization of bent functions and of some three-valued almost-optimal functions. We first fix notation.

Recall that φ_α, $\alpha \in \boldsymbol{F}_2^m$, denotes the linear function $x \longmapsto \alpha \cdot x$. We denote by H_α the kernel of φ_α

$$H_\alpha = \{x \in \boldsymbol{F}_2^m,\ \varphi_\alpha(x) = 0\}.$$

We denote by $\overline{H}_\alpha$ the affine subspace $\boldsymbol{F}_2^m \setminus H_\alpha$. Clearly, φ_α is the characteristic function of $\overline{H}_\alpha$.

Lemma V.3: Let $\alpha \in \boldsymbol{F}_2^m$ and φ_α the associated linear function with kernel H_α. We have

$$\mathcal{F}^2(f + \varphi_\alpha) = \sum_{e \in H_\alpha} \mathcal{F}(D_ef) - \sum_{e \in \overline{H}_\alpha} \mathcal{F}(D_ef) \tag{26}$$

$$\mathcal{F}^2(f) + \mathcal{F}^2(f + \varphi_\alpha) = 2 \sum_{e \in H_\alpha} \mathcal{F}(D_ef) \tag{27}$$

$$\mathcal{F}^2(f) - \mathcal{F}^2(f + \varphi_\alpha) = 2 \sum_{e \in \overline{H}_\alpha} \mathcal{F}(D_ef) \tag{28}$$

$$\mathcal{F}^2(f) = \mathcal{F}^2(f + \varphi_\alpha) \Leftrightarrow \sum_{e \in \overline{H}_\alpha} \mathcal{F}(D_ef) = 0 \tag{29}$$

$$\mathcal{F}^2(f)\mathcal{F}^2(f + \varphi_\alpha) = \left(\sum_{e \in H_\alpha} \mathcal{F}(D_ef) \right)^2 - \left(\sum_{e \in \overline{H}_\alpha} \mathcal{F}(D_ef) \right)^2. \tag{30}$$

Proof: Relations (3) and (4) can be rewritten

$$\mathcal{F}^2(f + \varphi_\alpha) = \sum_{e \in H_\alpha} \mathcal{F}(D_ef) - \sum_{e \in \overline{H}_\alpha} \mathcal{F}(D_ef)$$

which is exactly (26), and

$$\mathcal{F}^2(f) = \sum_{e \in H_\alpha} \mathcal{F}(D_ef) + \sum_{e \in \overline{H}_\alpha} \mathcal{F}(D_ef).$$

Formulas (27), (28), and (30) are obtained by combining the above relations. Formula (28) obviously implies (29). Note that (27) can be directly obtained from Lemma V.2 ($k = 1$). □

Lemma V.4: Let m be a positive integer, $m \geq 3$, and $f \in \mathcal{B}_m$. Define, for any $\alpha \in \boldsymbol{F}_2^m$, the property $(\mathcal{H}_\alpha)$: the function D_ef is balanced for every nonzero element e of H_α. If f satisfies $(\mathcal{H}_\alpha)$ for some α, then

$$\mathcal{F}^2(f + \varphi_\beta) + \mathcal{F}^2(f + \varphi_{\beta+\alpha}) = 2^{m+1}$$

for all $\beta \in \boldsymbol{F}_2^m$.

Proof: Since D_ef is balanced if and only if $\mathcal{F}(D_ef) = 0$, $(\mathcal{H}_\alpha)$ implies, in accordance with (27)

$$\mathcal{F}^2(f) + \mathcal{F}^2(f + \varphi_\alpha) = 2\mathcal{F}(D_0f) = 2^{m+1}.$$

Moreover, this property holds for any $f + \varphi_\beta$, since any function $D_e\varphi_\beta$ is constant, implying that $D_e(f + \varphi_\beta)$ is balanced as soon as D_ef is balanced. □

Theorem V.2: Let m be an odd integer, $m \geq 3$, $\alpha \in \boldsymbol{F}_2^m$, $\alpha \neq 0$, and $f \in \mathcal{B}_m$. Then the following properties are equivalent:

i) f satisfies $(\mathcal{H}_\alpha)$;

ii) f is three-valued almost-optimal and
$$\mathcal{F}^2(f + \varphi_\beta) \neq \mathcal{F}^2(f + \varphi_{\beta+\alpha})$$
for all $\beta \in \boldsymbol{F}_2^m$;

iii) both restrictions of f to H_α and $\overline{H}_\alpha$ are bent.

Proof: i) ⇒ ii). Lemma V.4 implies that

$$\mathcal{F}^2(f + \varphi_\beta) + \mathcal{F}^2(f + \varphi_{\beta+\alpha}) = 2^{m+1}$$

for all $\beta \in \boldsymbol{F}_2^m$. If there exists $\beta \in \boldsymbol{F}_2^m$ such that

$$\mathcal{F}^2(f + \varphi_\beta) = \mathcal{F}^2(f + \varphi_{\alpha+\beta})$$

then we obtain $\mathcal{F}^2(f + \varphi_\beta) = 2^m$ where 2^m is not a square, a contradiction.

Moreover, applying Lemma B.1 (in Appendix II), we deduce that $\mathcal{F}^2(f + \varphi_\beta) \in \{0, 2^{m+1}\}$, for all β. So f is three-valued almost-optimal.

ii) ⇒ iii). Let us denote by (h_1, h_2) the decomposition of f with respect to H_α. From Theorem V.1, we have

$$\mathcal{F}^2(f) + \mathcal{F}^2(f + \varphi_\alpha) = 2\left(\mathcal{F}^2(h_1) + \mathcal{F}^2(h_2)\right). \tag{31}$$

Since the Fourier spectrum of f is $\{0, \pm 2^{(m+1)/2}\}$ and $\mathcal{F}^2(f) \neq \mathcal{F}^2(f+\varphi_\alpha)$, we obtain $\mathcal{F}^2(h_1)+\mathcal{F}^2(h_2) = 2^m$, implying (see Lemma B.1)

$$\mathcal{F}^2(h_1) = \mathcal{F}^2(h_2) = 2^{m-1}.$$

This property holds for $f+\varphi_\beta$, for any β. Note that the decomposition of $f+\varphi_\beta$ with respect to H_α, when β ranges over $\boldsymbol{F}_2^m$, is $(h_1+\ell_1, h_2+\ell_2)$ where ℓ_i is any affine or constant function and where $\ell_1+\ell_2$ is constant. This proves that the Fourier spectrum of each h_i is $\{\pm 2^{m/2}\}$; thus h_i is a bent function of $m-1$ variables.

iii) $\Rightarrow$ i). Since h_i is bent, then $D_e h_i$ is balanced for any nonzero $e \in H_\alpha$; but

$$\mathcal{F}(D_e f) = \mathcal{F}(D_e h_1) + \mathcal{F}(D_e h_2)$$

for any such e. So we obtain $\mathcal{F}(D_e f) = 0$ for all such e. Hence f satisfies $(\mathcal{H}_\alpha)$. □

Remark V.4: It is important to notice that when f is three-valued almost-optimal (m odd) we have for any α

$$\mathcal{F}^2(f)+\mathcal{F}^2(f+\varphi_\alpha) \in \{0, 2^{m+1}, 2^{m+2}\}$$

according to Lemma B.1. Thus, the values occurring in the Fourier spectrum of h_1 (resp., h_2) are always contained in $\{0, \pm 2^{(m-1)/2}, \pm 2^{(m+1)/2}\}$. This means that h_1 and h_2 are both almost-optimal and this is true for any α—then for any corresponding decomposition of f.

Theorem V.3: Let m be an even integer, $m \geq 4$, and let $f \in \mathcal{B}_m$. Then the following properties are equivalent:

i) there is $\alpha \in \boldsymbol{F}_2^m$ such that f satisfies $(\mathcal{H}_\alpha)$;

ii) f is bent;

iii) f satisfies $(\mathcal{H}_\alpha)$ for all $\alpha \in \boldsymbol{F}_2^m \setminus \{0\}$;

iv) for any α, the decomposition (h_1, h_2) of f with respect to H_α satisfies: h_1 and h_2 are three-valued almost-optimal and for any linear Boolean function ℓ of $\mathcal{B}_{m-1}$, we have

$$\mathcal{F}^2(h_1+\ell) \neq \mathcal{F}^2(h_2+\ell)$$

(i.e., $\mathcal{F}^2(h_1+\ell) = 2^m$ if and only if $\mathcal{F}^2(h_2+\ell) = 0$).

Proof: Recall that a Boolean function f is bent if and only if $D_e f$ is balanced for all $e \neq 0$. Hence: ii)$\Leftrightarrow$ iii) and ii) $\Rightarrow$ i).

Assume that f satisfies $(\mathcal{H}_\alpha)$ for some α. Then for all β we have from Lemma V.4

$$\mathcal{F}^2(f+\varphi_\beta)+\mathcal{F}^2(f+\varphi_{\beta+\alpha}) = 2^{m+1}.$$

This implies, from Lemma B.1

$$\mathcal{F}^2(f+\varphi_\beta) = \mathcal{F}^2(f+\varphi_{\beta+\alpha}) = 2^m$$

completing the proof of i) $\Leftrightarrow$ ii).

Assuming that f is bent, we fix $\alpha \in \boldsymbol{F}_2^m$ and we denote by (h_1, h_2) the decomposition of f with respect to H_α. As in the previous proof, we obtain (31) which implies here (by using Lemma B.1), $\mathcal{F}^2(h_i) \in \{0, 2^m\}$ and $\mathcal{F}^2(h_1) \neq \mathcal{F}^2(h_2)$. This property holds if we consider $f+\varphi_\beta$ instead of f in (31)

$$\mathcal{F}^2(f+\varphi_\beta)+\mathcal{F}^2(f+\varphi_{\alpha+\beta}) = 2(\mathcal{F}^2(h_1+\ell)+\mathcal{F}^2(h_2+\ell'))$$

where $\ell, \ell' \in \mathcal{B}_{m-1}$, ℓ is a linear function (which can be 0), and ℓ' is either ℓ or $1+\ell$. Thus, $\mathcal{F}^2(h_2+\ell') = \mathcal{F}^2(h_2+\ell)$ and then $\mathcal{F}^2(h_1+\ell) \neq \mathcal{F}^2(h_2+\ell)$ completing the proof of iv).

Conversely, if iv) is satisfied then

$$\mathcal{F}^2(f+\varphi_\beta)+\mathcal{F}^2(f+\varphi_{\alpha+\beta}) = 2^{m+1}$$

for all β, implying that f is bent, completing the proof of ii) $\Leftrightarrow$ iv). □

Remark V.5: Note that the previous theorem is of interest for effective purpose. For checking that a function f is bent it is sufficient to compute the $\mathcal{F}(D_e f)$ for e in some hyperplane.

Example V.4: On the other hand, Property iv) provides some constructions: for every bent function f and every $\alpha \in \boldsymbol{F}_2^m$, $\alpha \neq 0$, both restrictions of f to H_α and $\overline{H}_\alpha$ are three-valued almost-optimal. For instance, choose f in class $\mathcal{PS}_{ap}$ (cf. [33]): $\boldsymbol{F}_2^m$ is identified, as a vector space, with $\boldsymbol{F}_{2^{m/2}} \times \boldsymbol{F}_{2^{m/2}}$ (i.e., the elements of $\boldsymbol{F}_2^m$ are considered as ordered pairs (x, y) where x and y belong to the finite field $\boldsymbol{F}_{2^{m/2}}$) and f is defined as $f(x, y) = g(\frac{x}{y})$, with $\frac{x}{0} = 0$, where g is any balanced Boolean function on $\boldsymbol{F}_{2^{m/2}}$. We do not know how to prove directly (i.e., without using Theorem V.3) that the restrictions of such a function to any hyperplane are three-valued almost-optimal.

We study now the more general case where a function f has balanced derivatives $D_a f$ for all nonzero a of W, a subspace of $\boldsymbol{F}_2^m$ of codimension 2. First note that, with the notation of Section V-B, we obtain, by applying Lemma V.2 and Theorem V.1

$$\sum_{\alpha \in W^\perp} \mathcal{F}^2(f+\varphi_\alpha) = 4 \sum_{e \in W} \mathcal{F}(D_e f) = 4 \sum_{i=1}^{4} \mathcal{F}^2(h_i) \quad (32)$$

where $(h_1, \ldots, h_4)$ is the decomposition of f with respect to W as described at the beginning of Section V-B. These formulas hold when f is replaced by $f+\varphi_\beta$, for any $\beta \in \boldsymbol{F}_2^m$.

Theorem V.4: Let m be any positive integer, $m \geq 3$, and $f \in \mathcal{B}_m$. Assume that there exists a linear subspace $W \subset \boldsymbol{F}_2^m$ of codimension 2 such that $D_a f$ is balanced for any nonzero $a \in W$. Let $(h_1, \ldots, h_4)$ be the decomposition of f with respect to W.

- If m is odd then f is three-valued almost-optimal and every h_i is three-valued almost-optimal.
- If m is even, then either f is bent or $\mathcal{L}(f) = 2^{(m+2)/2}$ and the values occurring in the Fourier spectrum of f belong to $\{0, \pm 2^{m/2}, \pm 2^{(m+2)/2}\}$. Moreover, all the h_i have the same Fourier spectrum: either all the h_i are bent, either all the h_i are three-valued almost-optimal, or the h_i have the same Fourier spectrum with values $\{0, \pm 2^{(m-2)/2}, \pm 2^{m/2}\}$. If all the h_i are three-valued almost-optimal then f is bent.

Proof: Since $\mathcal{F}(D_a f) = 0$ for any nonzero a in W, we have from (32), for any β

$$\sum_{\alpha \in W^\perp} \mathcal{F}^2(f+\varphi_{\alpha+\beta}) = 4\mathcal{F}(D_0 f) = 2^{m+2}.$$

Since $W^\perp$ has cardinality 4, we deduce from Lemma B.2 (in Appendix II) that the Fourier spectrum of f is $\{0, \pm 2^{(m+1)/2}\}$ when m is odd; the values occurring in this Fourier spectrum belong to $\{0, \pm 2^{m/2}, \pm 2^{(m+2)/2}\}$ when m is even. Hence, f is either three-valued almost-optimal (m odd), either bent or such

that $\mathcal{L}(f) = 2^{(m+2)/2}$, and the values of its Fourier transform belong to $\{0, \pm 2^{m/2}, \pm 2^{(m+2)/2}\}$.

Consider now the decomposition of f, say $(h_1, \ldots, h_4)$, with respect to W. We have from (32) again

$$\sum_{i=1}^{4} \mathcal{F}^2(h_i) = 2^m \tag{33}$$

and this property holds for any $f + \varphi_\beta$ and its decomposition, which implies that the values occurring in the Fourier spectrum of each h_i are (by applying Lemma B.2)

- if m is odd, $\{0, \pm 2^{(m-1)/2}\}$—i.e., h_i is three-valued almost-optimal;
- if m is even, either $\{\pm 2^{(m-2)/2}\}$ (i.e., h_i is bent) or contained in $\{0, \pm 2^{(m-2)/2}, \pm 2^{m/2}\}$—with $\mathcal{L}(h_i) = 2^{m/2}$.

According to Lemma B.2, the sum in (33) for even m is either $2^{m-2} \times 4$ or $2^m + 0 \times 3$. If one h_i is bent this sum is always $2^{m-2} \times 4$ implying that all h_i are bent too.

Similarly, if one h_i is three-valued almost-optimal, the values of its Fourier transform are in $\{0, \pm 2^{m/2}\}$. Since the value 2^{m-2} never appears, the sum in (33) is always $2^m + 0 \times 3$ implying that this property holds for all h_i. Moreover, for any β

$$\mathcal{F}(f + \varphi_\beta) = \sum_{i=0}^{4} \mathcal{F}(h_i + \ell) = \pm 2^{m/2} + 0 \times 3 = \pm 2^{m/2}$$

for some ℓ; so f is bent. Now suppose that the h_i are neither bent nor three-valued almost-optimal; then the values appearing in their Fourier spectra are 0, $\pm 2^{(m-2)/2}$, and $\pm 2^{m/2}$. We know that the number of times value $\pm 2^{(m-2)/2}$ occurs is the same for each h_i (by using (33) and Lemma B.2 as above), and Parseval's relation settles the case of the two other magnitudes. □

Note that there exist some functions f such that all h_i are three-valued almost-optimal and f is not.

Example V.5: For $m = 7$, we consider

$$\begin{aligned} f(x_1, \ldots, x_7) = {} & x_6 x_7 h_1(x_1, \ldots, x_5) \\ & + (1 + x_6) x_7 h_2(x_1, \ldots, x_5) \\ & + x_6 (1 + x_7) h_3(x_1, \ldots, x_5) \\ & + (1 + x_6)(1 + x_7) h_4(x_1, \ldots, x_5) \end{aligned}$$

where

$$\begin{aligned} h_1(x_1, \ldots, x_5) &= x_1 x_2 x_3 + x_1 x_4 + x_2 x_5 \\ h_2(x_1, \ldots, x_5) &= x_1 x_2 + x_3 x_4 \\ h_3(x_1, \ldots, x_5) &= x_2 x_3 + x_4 x_5 \\ h_4(x_1, \ldots, x_5) &= x_1 x_2 x_3 + x_1 x_4 x_5 + x_2 x_3 + x_2 x_4 + x_3 x_5. \end{aligned}$$

Although all functions h_i are three-valued almost-optimal, f is not: the coefficients $\mathcal{F}(f + \varphi_\alpha)$ belong to

$$\{0, \pm 8, \pm 16, \pm 24, \pm 32, \pm 40\}.$$

Moreover, $D_a f$ is balanced for 23 values of $a \in \boldsymbol{F}_2^7$.

Remark V.6: Take any bent function f on $\boldsymbol{F}_2^m$ (m even), any $(m-2)$-dimensional subspace W of $\boldsymbol{F}_2^m$ and any $a \in \boldsymbol{F}_2^m$. Then the Boolean function $g = f + 1_{a+W}$, where 1_{a+W} denotes the indicator of the flat $a + W$, satisfies the hypothesis of Theorem V.4. Indeed, we have for every $b \in W$: $D_b g = D_b f$ since $a + W$ is invariant under the translation by vector b.

It is clear that the set of functions which satisfy the hypothesis of Theorem V.4 (m even) contains all bent functions and also some functions whose Fourier spectrum is $\{0, \pm 2^{(m+2)/2}\}$. But we have also the following.

Proposition V.5: For every even $m \geq 6$, there exists $f \in \mathcal{B}_m$ satisfying the hypothesis of Theorem V.4, whose Fourier transform takes on exactly the three magnitudes 0, $2^{(m+2)/2}$, and $2^{m/2}$.

Proof: Let $m = 2t$; we identify the elements of $\boldsymbol{F}_2^m$ with the ordered pairs (x, y) where $x = (x_1, \ldots, x_t)$ and $y = (y_1, \ldots, y_t)$. Choose $g \in \mathcal{B}_m$ in Maiorana–McFarland class of bent functions in the form

$$g(x, y) = x \cdot y + k(y)$$

where k is some function in $\mathcal{B}_t$. Set $W = \{(x, y) \,|\, x_1 = x_2 = 0\}$ and

$$f(x, y) = x_1 x_2 + g(x, y) = x_1 x_2 + x \cdot y + k(y).$$

As remarked above, for any $e \in W$ we have $D_e f = D_e g$. Since g is bent then $D_e f$ is balanced. Now remark that

$$f(x, y) = (x_1 + y_2)(x_2 + y_1) + y_1 y_2 + \sum_{i=3}^{t} x_i y_i + k(y).$$

Thus, f is linearly equivalent to the function

$$x_1 x_2 + y_1 y_2 + \sum_{i=3}^{t} x_i y_i + k(y).$$

We see by exchanging x_2 and y_1 that f is linearly equivalent to the function

$$f'(x, y) = x \cdot y + k(x_2, y_2, y_3, \ldots, y_t).$$

It is a simple matter to check that, if $m \geq 6$, there exists a function $k(y)$ such that $\mathcal{F}^2(f' + \varphi_a(x) + \varphi_b(y))$ takes at least once each value of $\{0, 2^{m-2}, 2^m\}$—where $\varphi_a(x) = a \cdot x$ and $\varphi_b(x) = b \cdot y$ in $\mathcal{B}_t$. Take for instance $k(y) = y_1 y_2 y_3$. Then

$$\begin{aligned} & f'(x, y) + \varphi_a(x) + \varphi_b(y) \\ & \quad = \left(\sum_{i=1,3,\ldots,t} x_i (y_i + a_i) \right) + x_2 y_2 (1 + y_3) + a_2 x_2 + b \cdot y. \end{aligned}$$

Since

$$\sum_{x_1, x_3, \ldots, x_t \in \boldsymbol{F}_2} (-1)^{\sum_{i=1,3,\ldots,t} x_i (y_i + a_i)} \neq 0$$

if and only if $y_i = a_i$ for every $i = 1, 3, \ldots, t$, we deduce:

$$\begin{aligned} & \mathcal{F}(f' + \varphi_a(x) + \varphi_b(y)) \\ & \quad = \pm 2^{t-1} \sum_{x_2, y_2 \in \boldsymbol{F}_2} (-1)^{x_2 y_2 (1 + a_3) + a_2 x_2 + b_2 y_2}. \end{aligned}$$

If $a_3 = 0$, then we obtain $\pm 2^t$; if $a_3 = 1$, $a_2 = 0$, and $b_2 = 0$, we obtain $\pm 2^{t+1}$; and if $a_3 = 1$, $a_2 \neq 0$, or $b_2 \neq 0$ we obtain 0. The proof is complete. □

Remark V.7: There exist three-valued almost-optimal functions with m variables, m odd, which do not satisfy the hypotheses of Theorem V.4. For example, for $m = 7$, the function

$$f(x_1, \ldots, x_7) = x_2x_3 + x_4x_6 + x_5x_7 + x_1x_6x_7 + x_5x_6x_7 + x_2x_3x_6x_7 + x_4x_5x_6x_7$$

is three-valued almost-optimal. It has exactly 14 nonbalanced derivatives, which are all D_ef for

$$e \in \langle e_1, e_2, e_3\rangle \cup \langle e_1, e_4, e_5\rangle.$$

Open Problem V.2: Find some general property for a function f such that D_af is balanced when $a \in W$, $a \neq 0$, where W has codimension 3.

D. The Nonbalanced Derivatives

On the other hand, we consider the set of nonbalanced derivatives. Recall that, for $f \in \mathcal{B}_m$, E_f is the set $\{e \in \boldsymbol{F}_2^m | \mathcal{F}(D_ef) = 0\}$ and $\overline{E_f}$ is the complementary set $\boldsymbol{F}_2^m \setminus E_f$. In this section, we consider the rank of $\overline{E_f}$. For clarity, we first indicate an obvious property.

Lemma V.5: Let r be the rank of $\overline{E_f}$. Then $r < m$ means that there is a subspace V of dimension r such that $a + V$ is contained in E_f for all $a \notin V$.

It is natural to first consider the small values of r. As a direct application of our previous results, we are able to characterize the functions which correspond to the cases $r \leq 2$.

Corollary V.4: Let m be an odd integer, $m \geq 3$, $f \in \mathcal{B}_m$, and $e \in \boldsymbol{F}_2^m$. Then the following properties are equivalent:

- f is almost-optimal and e is a linear structure of f;
- f is three-valued almost-optimal and e is a linear structure of f;
- $\overline{E_f} = \{0, e\}$.

Proof: If f has a linear structure then $\mathcal{V}(f) \geq 2^{2m+1}$ (see Lemma II.1). Suppose that, moreover, f is almost-optimal. In accordance with Proposition V.2, the only possibility is $\mathcal{V}(f) = 2^{2m+1}$ which means (when f is almost-optimal) that f is three-valued almost-optimal. Since $\mathcal{F}^2(D_0f) = \mathcal{F}^2(D_ef) = 2^{2m}$, we deduce

$$\mathcal{V}(f) = \sum_{a \in \{0, e\}} \mathcal{F}^2(D_af)$$

providing $\mathcal{F}^2(D_af) = 0$ for $a \notin \{0, e\}$, according to (6).

Assume now that $\overline{E_f} = \{0, e\}$. Clearly, the set $E_f \cup \{0\}$ contains a subspace of codimension 1. So we apply Theorem V.2 and deduce that f is three-valued almost-optimal. Since

$$\mathcal{F}^2(f) = \mathcal{F}(D_0f) + \mathcal{F}(D_ef)$$

then $\mathcal{F}(D_ef) = \pm 2^m$, completing the proof. □

Corollary V.5: Let m be an even integer, $m \geq 4$, and $f \in \mathcal{B}_m$. Let V be some linear space of dimension 2. Then the following properties are equivalent:

- f is almost-optimal and any $e \in V$ is a linear structure of f;
- f is three-valued almost-optimal and any $e \in V$ is a linear structure of f;
- $\overline{E_f} = V$.

Proof: We proceed as in the previous proof. If V is a linear space for f, then $\mathcal{V}(f) \geq 2^{2m+2}$. If, moreover, f is almost-optimal then $\mathcal{V}(f) = 2^{2m+2}$ which means that f is three-valued almost-optimal. Now compute the sum-of-squares indicator

$$\mathcal{V}(f) = \sum_{a \in V} \mathcal{F}^2(D_af) + \sum_{a \notin V} \mathcal{F}^2(D_af) = 2^{2m+2} + \sum_{a \notin V} \mathcal{F}^2(D_af)$$

providing $\overline{E_f} = V$.

Conversely, assume that $\overline{E_f} = V$. Thus, $E_f \cup \{0\}$ contains a subspace of codimension 2, say W. In accordance with Theorem V.4, $\mathcal{L}(f) = 2^{(m+2)/2}$ and the values of the Fourier transform of f lie in $\{0, \pm 2^{m/2}, \pm 2^{(m+2)/2}\}$ (f cannot be bent). We can assume that $\mathcal{F}^2(f) = 2^{m+2}$. By using (27), taking $H_\alpha = W \cup (a + W)$ with $a \in V$, $a \neq 0$, we obtain

$$\mathcal{F}^2(f) + \mathcal{F}^2(f + \varphi_\alpha) = 2(\mathcal{F}(D_0f) + \mathcal{F}(D_af)) = 2^{m+1} + 2\mathcal{F}(D_af) \quad (34)$$

where $0 < |\mathcal{F}(D_af)| \leq 2^m$. Then $\mathcal{F}(D_af) = \pm 2^m$. Note that this property holds for any $a \in V$. So we have proved that f is almost-optimal and that it has any $a \in V$ as linear structure, completing the proof. □

Remark V.8: Note that we are not able to give the Fourier spectrum of any almost-optimal function which has a linear space of dimension 1 when m is even. Actually, this problem is equivalent to the determination of Fourier spectrum of almost-optimal functions of $\mathcal{B}_{m-1}$ (see the next example).

Example V.6: There exist almost-optimal (non-three-valued) functions of degree 4 for odd $m \geq 5$. For instance, for $m = 5$, the function

$$g(x_1, \ldots, x_5) = x_3x_5 + x_2x_4 + x_1x_2x_3 + x_2x_3x_4x_5$$

is given in [23]. Its Fourier transform takes all the values in $\{0, \pm 4, \pm 8\}$. Note that in the decomposition of g with respect to the subspace defined by $x_3 = 0$ is x_2x_4, a quadratic component of g (see Proposition V.4). Moreover, one can check that $\overline{E_g}$ is a subspace of dimension 3.

Now consider the function of six variables

$$f(x_1, \ldots, x_6) = g(x_1, \ldots, x_5) + x_6.$$

It is clear that $(0, \ldots, 0, 1)$ is a linear structure of f and it is easy to check that the set of values appearing in the Fourier spectrum of f is $\{0, \pm 8, \pm 16\}$.

The previous corollaries were partially proved in [34] where the authors study the cases $\#\overline{E_f} = 1, 2, \ldots, 6$. Generally, it seems difficult to characterize f such that $\overline{E_f}$ is a linear space of dimension k for some k (see [10]). When the rank of $\overline{E_f}$ is 3, we can give the next property but cannot describe the case $\#\overline{E_f} = 8$—examples are easily obtained (see Example V.6).

Corollary V.6: Let $f \in \mathcal{B}_m$ and assume that the rank of $\overline{E_f}$ is 3.

If $\#\overline{E_f} < 8$ then E_f contains all nonzero elements of some subspace of codimension 2. So Theorem V.4 can be applied.

Proof: Set $E'_f = E_f \cup \{0\}$. Assume that the cardinality of $\overline{E_f}$ is strictly less than 8. Let (e_1, e_2, e_3) be linearly independent in $\overline{E_f}$; by completing, we have a basis $(e_1, \ldots, e_m)$ of $\boldsymbol{F}_2^m$ such that $W = \langle e_4, \ldots, e_m\rangle$ is contained in E'_f. But there is some a, a linear combination of (e_1, e_2, e_3), which is in E'_f. So the subspace $W \cup (a+W)$, of codimension 2 is contained in E'_f. □

Note that, for any t, there exist some functions f such that $\overline{E_f}$ has rank t (such functions can be constructed recursively, by taking partially bent functions). Hence, this property does not seem to be significant. It nevertheless induces some simplifications on the decompositions of the function, as shown in the next theorem.

Theorem V.5: Suppose that $\overline{E_f}$ is contained in W, a subspace of dimension t, $1 < t < m$. Considering notation of Theorem V.1, let $\{h_a | a \in V\}$ be the decomposition of f with respect to W, then we have

i) $\sum_{a\in V} \mathcal{F}^2(h_a) = \mathcal{F}^2(f)$. Moreover,

$$\mathcal{L}^2(f) \le \sum_{a\in V} \mathcal{L}^2(h_a);$$

ii) $\mathcal{F}^2(f+\varphi_\alpha) = \mathcal{F}^2(f)$ for any $\alpha \in W^\perp$ and the Fourier spectrum of f cannot have more than 2^t magnitudes.

Proof: Let $\alpha \in W^\perp$. Any $e \in \overline{H}_\alpha$ does not belong to W, since $e \cdot \alpha = 1$. It follows that $D_e f$ is balanced for all $e \in \overline{H}_\alpha$. We deduce from Lemma V.3 that $\mathcal{F}^2(f+\varphi_\alpha) = \mathcal{F}^2(f)$. Thus, applying Theorem V.1, we obtain

$$\sum_{\alpha\in W^\perp} \mathcal{F}^2(f+\varphi_\alpha) = 2^{m-t}\mathcal{F}^2(f) = 2^{m-t}\sum_{a\in V} \mathcal{F}^2(h_a).$$

It follows that

$$\sum_{a\in V} \mathcal{F}^2(h_a) = \mathcal{F}^2(f).$$

This property holds for any $f+\varphi_\beta, \beta \in \boldsymbol{F}_2^m$, since $D_e(f+\varphi_\beta)$ is balanced as soon as $D_e f$ is balanced. The upper bound on $\mathcal{L}^2(f)$ is obviously deduced. Take any $\beta \in \boldsymbol{F}_2^m$; then we obtain, as above

$$\mathcal{F}^2(f+\varphi_{\beta+\alpha}) = \mathcal{F}^2(f+\varphi_\beta)$$

for all $\alpha \in W^\perp$. This implies that the Fourier spectrum of f cannot have more than 2^t magnitudes. □

Several corollaries can be deduced. We study, for instance, the case where W has codimension 1.

Corollary V.7: Assume that $f \in \mathcal{B}_m$ is such that $\overline{E_f} \subset H_\alpha$, some subspace of codimension 1. Denote by (h_1, h_2) the decomposition of f with respect to H_α. Then we have

i) for any $\beta \in \boldsymbol{F}_2^m$

$$\mathcal{F}^2(f+\varphi_\beta) = \max(\mathcal{F}^2(h_1+\ell), \mathcal{F}^2(h_2+\ell))$$

where $(h_1+\ell, h_2+\ell+\varepsilon)$ is the decomposition of $f+\varphi_\beta$, ℓ is a linear function in $\mathcal{B}_{m-1}$ and ε is constant;

ii) for every linear function ℓ in $\mathcal{B}_{m-1}$, at least one term in the pair $(\mathcal{F}(h_1+\ell), \mathcal{F}(h_2+\ell))$ is zero;

iii)

$$\mathcal{L}(f) = \max(\mathcal{L}(h_1), \mathcal{L}(h_2))$$
$$\mathcal{N}(f) = 2^{m-2} + \min(\mathcal{N}(h_1), \mathcal{N}(h_2))$$

and

$$\mathcal{V}(f) = \mathcal{V}(h_1) + \mathcal{V}(h_2).$$

Proof: Since $\mathcal{F}(f) = \mathcal{F}(h_1) + \mathcal{F}(h_2)$, we have, according to Theorem V.5

$$\mathcal{F}^2(f) = \mathcal{F}^2(h_1) + \mathcal{F}^2(h_2) = (\mathcal{F}(h_1) + \mathcal{F}(h_2))^2.$$

Thus, $\mathcal{F}(h_1)\mathcal{F}(h_2) = 0$ providing either $\mathcal{F}(h_1) = 0$ or $\mathcal{F}(h_2) = 0$—where both can be zero. This property holds when we consider $f+\varphi_\beta$ for any $\beta \in \boldsymbol{F}_2^m$ and the decomposition of $f+\varphi_\beta$ which is actually of the form $(h_1+\ell, h_2+\ell+\varepsilon)$, where ℓ is linear and ε is constant. We obviously have that $\mathcal{F}^2(h_2+\ell+\varepsilon) = \mathcal{F}^2(h_2+\ell)$. Then i) and ii) are clearly proved and the values of $\mathcal{L}(f)$ and $\mathcal{N}(f)$ given in iii) are easily deduced.

Now we compute $\mathcal{V}(f)$, by using (7) and the previous properties. We denote by L_{m-1} the set of all linear functions in $\mathcal{B}_{m-1}$. We then deduce

$$\begin{aligned}
2^m\mathcal{V}(f) &= \sum_{\beta\in\boldsymbol{F}_2^m} \mathcal{F}^4(f+\varphi_\beta) \\
&= \sum_{\ell\in L_{m-1}, \varepsilon\in\boldsymbol{F}_2} (\mathcal{F}^2(h_1+\ell) + \mathcal{F}^2(h_2+\ell+\varepsilon))^2 \\
&= 2\sum_{\ell\in L_{m-1}} (\mathcal{F}^2(h_1+\ell) + \mathcal{F}^2(h_2+\ell))^2 \\
&= 2\sum_{\ell\in L_{m-1}} (\mathcal{F}^4(h_1+\ell) + \mathcal{F}^4(h_2+\ell)) \\
&= 2\left(2^{m-1}\mathcal{V}(h_1) + 2^{m-1}\mathcal{V}(h_2)\right) \\
&= 2^m\left(\mathcal{V}(h_1) + \mathcal{V}(h_2)\right)
\end{aligned}$$

completing the proof. □

Example V.7: Let $f \in \mathcal{B}_m$ such that $D_a f$ is a linear nonconstant function, for some a. So it is clear that $D_e D_a f$ is constant for any e.

Set $D_a f = \varphi_\alpha$ and recall that H_α denotes the kernel of φ_α. It is clear that $D_e D_a f = 1$ if and only if $e \notin H_\alpha$. Thus, according to Proposition II.5, $D_e f$ is balanced for any $e \notin H_\alpha$. This implies that $\overline{E_f} \subset H_\alpha$. It is very easy to construct such a function f. For instance,

$$f(x_1, \ldots, x_5) = x_1x_2x_3 + x_1x_3x_4 + x_2x_5$$

is such that $D_a f$, $a = (0, 0, 0, 0, 1)$, is equal to x_2.

Remark V.9: For any bent function f we have $\overline{E_f} = \{0\}$. Hence, f satisfies the hypothesis of Theorem V.5 for any W. For such a function, Theorem V.5 i), becomes

$$\sum_{a\in V} \mathcal{F}^2(h_a) = 2^m \quad \text{implying} \quad 2^m \le \sum_{a\in V} \mathcal{L}^2(h_a).$$

Note that the property on the left holds for any $f+\varphi_\beta$ and the corresponding decomposition. If V has dimension 1 or 2 we can apply Lemmas B.1 and B.2. So we obtain again some results given by Theorem V.3 and V.4.

Appendix I

We briefly recall some properties of quadratic functions. More can be found in [22, Ch. 15] and [15]. In this appendix, f denotes a Boolean function of degree 2 of m variables. The *associated symplectic form* of f is the mapping from $(\boldsymbol{F}_2^m)^2$ to $\boldsymbol{F}_2$

$$\Psi(u, v) = f(0) + f(u) + f(v) + f(u+v)$$

where $(u, v) \in (\boldsymbol{F}_2^m)^2$. The *kernel* of Ψ is defined as follows:

$$\mathcal{E}_f = \{u \in \boldsymbol{F}_2^m | \forall v \in \boldsymbol{F}_2^m \colon \Psi(u, v) = 0\}.$$

The set $\mathcal{E}_f$ is a $\boldsymbol{F}_2$-subspace of $\boldsymbol{F}_2^m$ of dimension $m-2h$, where $2h$ is the *rank* of Ψ. This rank satisfies

- $1 \le h \le m/2$ for even m, and
- $1 \le h \le (m-1)/2$ for odd m.

Obviously, $\mathcal{E}_{f+\ell} = \mathcal{E}_f$ for any linear function ℓ. The Fourier spectrum of f (and thus the weight distribution of the corresponding coset $\Omega_f + R(1, m)$) only depends on h (cf. [22, p. 441]). For such a coset, the weights are $\{2^{m-1}, 2^{m-1} \pm 2^{m-h-1}\}$ and the corresponding numbers of codewords $\{2^{m+1}-2^{2h+1}, 2^{2h}\}$.

So the quadratic functions are three-valued unless $h = m/2$ for even m. In this case, the function is bent and its Fourier spectrum is $\{\pm 2^{m/2}\}$.

Proposition A.1: An element a is in $\mathcal{E}_f$ if and only if the function $D_a f$ is constant. The subspace $\mathcal{E}_f$ is the linear space of f.

Moreover, f is balanced if and only if there is $a \in \mathcal{E}_f$ such that $D_a f = 1$.

Proof: Note that $D_a f$ is constant if and only if

$$f(v) + f(a+v) = \varepsilon$$

for all v, where ε denotes a constant—either 0 or 1. But $a \in \mathcal{E}_f$ means

$$f(0) + f(a) = f(v) + f(a+v) \qquad \forall v$$

or, equivalently, $D_a f(v) = D_a f(0)$, for all v. This proves the first sentence of the proposition.

If $D_a f = 1$ for some a then f is balanced (see Proposition II.5). Conversely, suppose that f is balanced. Denote by τ the dimension of $\mathcal{E}_f$. Recall that $\tau = m-2h$ and that the number of 0's in the Fourier spectrum of f is equal to $2^{m+1} - 2^{2h+1}$. Note that f cannot be bent, so that the dimension of $\mathcal{E}_f$ is at least 1. We assume that for any $a \in \mathcal{E}_f$, $a \neq 0$, $D_a f = 0$ and we are going to prove that this is impossible. Define the subspace

$$B = \{\varphi_b \,|\, \mathcal{E}_f \subset \mathrm{Ker}\varphi_b\}$$

of the space of linear functions $\{\varphi_b \,|\, b \in \boldsymbol{F}_2^m\}$ (where $\varphi_0 = 0$). The number of hyperplanes of $\boldsymbol{F}_2^m$ containing $\mathcal{E}_f$ is equal to $2^{m-\tau}-1$. Thus, the cardinality of B is $2^{m-\tau}$ (by adding φ_0). We then have $2^m - 2^{m-\tau}$ functions φ_u, $u \neq 0$, such that $\varphi_u(a) = 1$ for some $a \in \mathcal{E}_f$. But for such φ_u, we have

$$D_a(f+\varphi_u) = D_a f + 1 = 1$$

implying that $f + \varphi_u$ is balanced. Therefore, $f + \varphi_u + 1$ is balanced too providing at all $2^{m+1} - 2^{2h+1}$ zero values in the Fourier spectrum of f. We have proved that any balanced function $f + \varphi_u$ is such that $u \neq 0$ (since $u \notin B$). This contradicts that f itself is balanced. □

Appendix II

Lemma B.1: Let m be an integer, $m \ge 0$, and let X and Y be two integers. Then the condition $X^2 + Y^2 = 2^{m+1}$ implies

- if m is even then $X^2 = Y^2 = 2^m$;
- if m is odd then $X^2 = 2^{m+1}$ and $Y = 0$ or vice versa.

Proof: This lemma can be proved by induction, as it is shown in [34], but this result was first stated and proven by Jacobi in 1828. His proof relies on the fact that the number of solutions $(X_1, \ldots, X_k) \in \mathbb{Z}^k$ of the equation

$$X_1^2 + \cdots + X_k^2 = N \tag{35}$$

is exactly the coefficient c_N of x^N in the expansion of θ^k, where

$$\theta = \sum_{i=-\infty}^{+\infty} x^{i^2} = 1 + 2x + 2x^4 + 2x^9 + \cdots.$$

In 1828, Jacobi proved that (see [35] and [36])

$$\theta^2 = 1 + 4\sum_{N=1}^{+\infty} \left(\sum_{d|N,\, d\equiv 1 \bmod 4} 1 - \sum_{d|N,\, d\equiv 3 \bmod 4} 1 \right) x^N$$

which means that the number of solutions of (35) satisfies

- if there exists a divisor d of N, $d \equiv 3 \bmod 4$, which occurs in N to an odd power, then $c_N = 0$;
- else

$$c_N = 4\left(\sum_{d|N,\, d\equiv 1 \bmod 4} 1 - \sum_{d|N,\, d\equiv 3 \bmod 4} 1 \right).$$

Then we have $c_{2^{m+1}} = 4$; this means that for our case ($N = 2^{m+1}$ and $X_i \ge 0$ for all i), the only solutions are the ones presented in the lemma. □

Lemma B.2: Let m be an integer, $m \ge 1$, and let X, Y, Z and T be four integers. Then the condition

$$X^2 + Y^2 + Z^2 + T^2 = 2^{m+2}$$

implies

- if m is even, then either $X^2 = Y^2 = Z^2 = T^2 = 2^m$ or $X^2 = 2^{m+2}$ and $Y^2 = Z^2 = T^2 = 0$;
- if m is odd, then $X^2 = Y^2 = 2^{m+1}$ and $Z = T = 0$.

Proof: In the same way, this result can be obtained by induction, but was stated by Jacobi in 1828, since we have

$$\theta^4 = 1 + 8 \sum_{s\ge 1,\, 4\not| s} \frac{sx^s}{1-x^s}$$

where s runs through all positive integers which are not multiples of 4. Then we have $c_{2^{m+2}} = 24$; Jacobi then proved that for our case, the solutions presented in the lemma are the only ones. □

ACKNOWLEDGMENT

The authors wish to thank one anonymous referee for many useful comments which greatly improved the manuscript.

REFERENCES

[1] T. Kasami, "Weight distributions of Bose–Chaudhuri–Hocquenghem codes," in *Proc. Conf. Combinatorial Mathematics and its Applications*. Chapel Hill, NC: Univ. North Carolina Press, 1968, pp. 335–357.

[2] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, pp. 656–715, 1949.

[3] M. Matsui, "Linear cryptanalysis method for DES cipher," in *Advances in Cryptology—EUROCRYPT'93 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1993, vol. 765, pp. 386–397.

[4] A. Canteaut and M. Trabbia, "Improved fast correlation attacks using parity-check equations of weight 4 and 5," in *Advances in Cryptology—EUROCRYPT 2000 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2000, vol. 1807, pp. 573–588.

[5] A. F. Webster and S. E. Tavares, "On the design of S-boxes," in *Advances in Cryptology—CRYPTO'85 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1985, vol. 219, pp. 523–534.

[6] B. Preneel, W. V. Leekwijck, L. V. Linden, R. Govaerts, and J. Vandewalle, "Propagation characteristics of Boolean functions," in *Advances in Cryptology—EUROCRYPT'90 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1991, vol. 437, pp. 155–165.

[7] B. Preneel, R. Govaerts, and J. Vandewalle, "Boolean functions satisfying higher order propagation criteria," in *Advances in Cryptology—EUROCRYPT'91 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1992, vol. 547, pp. 141–152.

[8] Y. Zheng, X.-M. Zhang, and H. Imai, "Restriction, terms and nonlinearity of Boolean functions," *Theor. Comput. Sci.*, vol. 226, no. 1–2, pp. 207–223, 1999.

[9] X.-M. Zhang and Y. Zheng, "GAC—The criterion for global avalanche characterics of cryptographic functions," *J. Univ. Comput. Sci.*, vol. 1, no. 5, pp. 320–337, 1995.

[10] A. Canteaut, C. Carlet, P. Charpin, and C. Fontaine, "Propagation characteristics and correlation-immunity of highly nonlinear Boolean functions," in *Advances in Cryptology—EUROCRYPT 2000 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2000, vol. 1807, pp. 507–522.

[11] O. S. Rothaus, "On bent functions," *J. Combin. Theory Ser. A*, vol. 20, pp. 300–305, 1976.

[12] C. Fontaine, "On some cosets of the first-order Reed–Muller code with high minimum weight," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1237–1243, May 1999.

[13] T. Helleseth, T. Kløve, and J. Mykkeltveit, "On the covering radius of binary codes," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 627–628, Sept. 1978.

[14] J. H. Evertse, "Linear structures in block ciphers," in *Advances in Cryptology—EUROCRYPT' 87 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1988, vol. 304, pp. 249–266.

[15] C. Carlet, "Codes de Reed–Muller, codes de Kerdock et de Preparata," Ph.D. dissertation, Univ. Paris 6, Paris, France, 1990.

[16] C. Carlet, "Partially-bent functions," *Des., Codes Cryptogr.*, no. 3, pp. 135–145, 1993.

[17] E. F. Assmus and J. Key, "Polynomials codes and finite geometry," in *Handbook of Coding Theory—Part 2: Connections*, V. S. Pless, W. C. Huffman, and R. A. Brualdi, Eds. Amsterdam, The Netherlands: Elesevier, 1998, ch. 16, pp. 1269–1343.

[18] P. Charpin, "Codes cycliques étendus invariants sous le groupe affine,", Thése d'Etat, Univ. Paris 7, Paris, France, 1987. LITP 87-6.

[19] P. Charpin, "Open problems on cyclic codes," in *Handbook of Coding Theory—Part 1*, V. S. Pless, W. C. Huffman, and R. A. Brualdi, Eds. Amsterdam, The Netherlands: Elsevier, 1998, ch. 11, pp. 963–1063.

[20] S. D. Berman, "On the theory of group codes," *Kibernetica*, vol. 1, no. 1, pp. 31–39, 1967.

[21] V. Pless, "Power moment identities on weight distributions in error-correcting codes," *Inform. Contr.*, vol. 3, pp. 147–152, 1963.

[22] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.

[23] E. R. Berlekamp and L. R. Welch, "Weight distributions of the cosets of the $(32, 6)$ Reed–Muller code," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 203–207, Jan. 1972.

[24] A. Canteaut, "On the weight distributions of optimal cosets of the first-order Reed–Muller codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 407–413, Jan. 2001.

[25] A. Canteaut, P. Charpin, and H. Dobbertin, "Binary m-sequences with three-valued crosscorrelation: A proof of Welch conjecture," *IEEE Trans. Inform. Theory*, vol. 46, pp. 4–8, Jan. 2000.

[26] ——, "Weight divisibility of cyclic codes, highly nonlinear functions on GF(2^m) and crosscorrelation of maximum-length sequences," *SIAM J. Discr. Math.*, vol. 13, no. 1, pp. 105–138, 2000.

[27] H. Dobbertin, "Another proof of Kasami's Theorem," *Des., Codes Cryptogr.*, vol. 17, no. 1/3, pp. 177–180, 1999.

[28] C. Carlet, P. Charpin, and V. Zinoviev, "Codes, bent functions and permutations suitable for DES-like cryptosystems," *Des., Codes Cryptogr.*, vol. 15, no. 2, pp. 125–156, 1998.

[29] E. Filiol and C. Fontaine, "Highly nonlinear balanced Boolean functions with a good correlation-immunity," in *Advances in Cryptology—EUROCRYPT'98 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1998, vol. 1403, pp. 475–488.

[30] H. Dobbertin, "Construction of bent functions and balanced Boolean functions with high nonlinearity," in *Fast Software Encryption (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1994, vol. 1008, pp. 61–74.

[31] N. J. Patterson and D. H. Wiedemann, "The covering radius of the $[2^{15}, 16]$ Reed–Muller code is at least 16276," *IEEE Trans. Inform. Theory*, vol. IT-36, no. 2, p. 443, 1983.

[32] X.-D. Hou, "On the covering radius of $R(1,\ m)$ in $R(3,\ m)$," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 1035–1037, 1996.

[33] J. F. Dillon, "Elementary Hadamard Difference sets," Ph.D. dissertation, University of Maryland, 1974.

[34] X.-M. Zhang and Y. Zheng, "Characterizing the structures of cryptographic functions satisfying the propagation criterion for almost all vectors," *Designs, Codes and Cryptography*, vol. 7, no. 1, pp. 111–134, 1996.

[35] C. G. J. Jacobi, "Correspondance mathématique entre Legendre et Jacobi," *J. Reine Angew. Math.*, no. 80, pp. 205–279, 1875.

[36] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 3rd ed. London, U.K.: Clarendon, 1954.

[37] C. Carlet, "Two new classes of bent functions," in *Advances in Cryptology—EUROCRYPT'93 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1994, vol. 765, pp. 77–101.

[38] ——, "On the propagation criterion of degree ℓ and order k," in *Advances in Cryptology—EUROCRYPT'98 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1998, vol. 1403, pp. 462–474.

[39] ——, "On cryptographic propagation criteria for Boolean functions," *Inform. Comput.*, no. 151, pp. 32–56, 1999.

[40] T. Kasami, "The weight enumerators for several classes of subcodes of the second order binary Reed–Muller codes," *Inform. Contr.*, vol. 18, pp. 369–394, 1971.

[41] S. Maitra and P. Sarkar, "Highly nonlinear resilient functions optimizing Siegenthaler's inequality," in *Advances in Cryptology—CRYPTO'99 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1999, vol. 1666, pp. 198–215.

[42] R. J. McEliece, "Weight congruence for p-ary cyclic codes," *Discr. Math.*, vol. 3, pp. 177–1925, 1972.

[43] T. Siegenthaler, "Correlation-immunity of nonlinear combining functions for cryptographic applications," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 776–780, Sept. 1984.

[44] G. Xiao and J. L. Massey, "A spectral characterization of correlation-immune combining functions," *IEEE Trans. Inform. Theory*, vol. 34, pp. 569–571, May 1988.

Bibliographie

[ABD+99] D. Augot, J.M. Boucqueau, J.-F. Delaigle, C. Fontaine, and E. Goray. Secure Delivery of Images over Open Networks. *Proceedings of the IEEE*, 87(7) :1251–1266, 1999. numÃľro spécial "Identification and protection of multimedia information", article invité.

[Abd11] W. Abdul. *Robust Multichannel Perceptual Color Image Watermarking and Private Anonymous Fingerprinting.* PhD thesis, Université de Poitiers, 2011.

[ABF11] D. Augot, M. Barbier, and C. Fontaine. Ensuring message embedding in wet paper steganography. In *11th IMA Conference on Cryptography and Coding*, Lecture Notes in Computer Science. Springer-Verlag, 2011. to appear.

[ABTD+06] D. Azemard, A. Benjelloun-Touimi, C. Delpha, C. Duhamel, J.-B. Fischer, C. Fontaine, C. Giraud, A. Le Guyader, P. Martin, and M. Milhau. Secured diffusion of music on mobile : an end-to-end approach. In *Taiwanese-French conference on Information Technology, TFIT'06.* INRIA, 2006. invited paper.

[AFD98] D. Augot, C. Fontaine, and J.-F. Delaigle. DHWM : a scheme for managing watermarking keys in the Aquarelle multimedia distributed system. In *European Symposium On Research In Computer Security - ESORICS 98*, volume 1485 of *Lecture Notes in Computer Science*, pages 241–255. Springer-Verlag, 1998.

[AFI+04] G. Ars, J.-C. Faugère, H. Imai, M. Kawazoe, and M. Sugita. Comparison between XL and Gröbner basis algorithms. 338-353. In *Advances in Cryptology – ASIACRYPT'04*, number 3329 in Lecture Notes in Computer Science, pages 338–353. Springer-Verlag, 2004.

[AGC10] W. Abdul, P. Gaborit, and P. Carré. Private anonymous fingerprinting for color images in the wavelet domain. In *IS&T/SPIE International Symposium on Electronic Imaging 2010 - Media Forensics and Security II*, volume 7541 of *Proceedings of the SPIE.* SPIE, 2010.

[AKS06] A. Adelsbach, S. Katzenbeisser, and A.-R. Sadeghi. A computational model for watermark robustness. In *Information Hiding, 8th*

International Workshop - IH 2006, volume 4437 of *Lecture Notes in Computer Science*, pages 145–160. Springer-Verlag, 2006.

[Ann97] F.S. Annexstein. Generating de Bruijn sequences : an efficient implementation. *IEEE Transactions on Computers*, 46(2) :198–200, 1997.

[AP98] R. Anderson and F.A.P. Petitcolas. On the limits of steganography. *IEEE Journal of Selected Areas in Communications*, 16(4) :474–481, 1998.

[ARS06] A. Adelsbach, M. Rohe, and A.-R. Sadeghi. Efficient implementation of zero-knowledge proofs for watermark detection in multimedia data. In *Transactions on Data Hiding and Multimedia Security I*, number 4300 in Lecture Notes in Computer Science, pages 73–103. Springer-Verlag, 2006.

[AT09] E Amiri and G. Tardos. High rate fingerprinting codes and the fingerprinting capacity. In *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms - SODA 2009*, pages 336–345. SIAM, 2009.

[BA08] J. Barbier and S. Alt. Practical insecurity for effective steganalysis. In *Information Hiding, 10th International Workshop - IH 2008*, volume 5284 of *Lecture Notes in Computer Science*, pages 195–208. Springer-Verlag, 2008.

[Bar98] A. Barg. Complexity issues in coding theory. In V.S. Pless, W.C. Huffman, and R.A. Brualdi, editors, *Handbook of coding theory*, volume 1, chapter 7. Elsevier, 1998.

[Bar11] M. Barbier. *Décodage en liste et application à la sécurité de l'information*. PhD thesis, Ecole Polytechnique, 2011.

[BBCS06] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai. Superposition coding for side-information channels. *IEEE Transactions on Information Theory*, 52(5) :1872–1889, 2006.

[BBF03] M. Barni, F. Bartolini, and T. Furon. A general framework for robust watermarking security. *Signal Processing*, 83(10) :2069–2084, October 2003. Special issue on Security of Data Hiding Technologies, invited paper.

[BBK03] A. Barg, G.R. Blakley, and G.A. Kabatianski. Digital fingerprinting codes : problem statements, constructions, identification of traitors. *IEEE Transactions on Information Theory*, 49(4) :852–865, 2003.

[BC05] M. Backes and C. Cachin. Public-key steganography with active attacks. In *2nd Theory of Cryptography Conference (TCC)*, volume 3378 of *Lecture Notes in Computer Science*, pages 210–226. Springer-Verlag, 2005.

[BC06a] P. Bas and F. Cayre. 1-14. In *Information Hiding, 8th International Workshop - IH 2006*, volume 4437 of *Lecture Notes in Computer Science*. Springer-Verlag, 2006.

[BC06b] P. Bas and F. Cayre. Achieving subspace or key security for woa using natural or circular watermarking. In *Proc. of the ACM Multimedia and Security Workshop 2006*, pages 80–88. ACM, 2006.

[BCE+01] A. Barg, G. Cohen, S. Encheva, G.A. Kabatianski, and G. Zèmor. A hypergraph approach to the identifying parent property : the case of multiple parents. *SIAM Journal of Discrete Mathematics*, 14(3) :423–431, 2001.

[BCG11] J. Bahi, J.-F. Couchot, and C. Guyeux. Steganography : a class of algorithms having secure properties. In *Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing – IIH-MSP 2011*, pages 209–212. IEEE Computer Society Press, 2011.

[BCG12] J. Bahi, J.-F. Couchot, and C. Guyeux. Steganography : a class of secure and robust algorithms. *The Computer Journal*, 2012.

[BD07] P. Bas and G. Doërr. Practical security analysis of dirty paper trellis watermarking. In *Information Hiding, 9th International Workshop - IH 2007*, volume 4567 of *Lecture Notes in Computer Science*, pages 174–188. Springer-Verlag, 2007.

[BDF01] F. Bao, R.H. Deng, and P. Feng. An efficient and practical Scheme for Privacy Protection in the E-Commerce of Digital Goods. In *ICISC 2000*, volume 2015 of *LNCS*, pages 162–170. Springer-Verlag, 2001.

[Bén06] V. Bénony. *Étude et conception de systèmes de chiffrement à flot dans le contexte d'architectures matérielles fortement contraintes.* PhD thesis, Université des Sciences et Technologies de Lille, 2006.

[BF08] J. Bierbrauer and J. Fridrich. Constructing good covering codes for applications in steganography. In *Transactions on Data Hiding and Multimedia Security III*, number 4920 in Lecture Notes in Computer Science, pages 1–22. Springer-Verlag, 2008.

[BFP11] P. Bas, T. Filler, and T. Pevný. "break our steganographic system" - the ins and outs of organizing boss. In *13th Information Hiding – IH'11*, Lecture Notes in Computer Science. Springer-Verlag, 2011. to appear.

[BG10] J. Bahi and C. Guyeux. A new chaos-based watermarking algorithm. In *International Conference on Security and Cryptography – SECRYPT'2010*, pages 455–458. SciTePress, 2010.

[BGL01] R.P. Brent, S. Gao, and A.G.B. Lauder. Random Krylov spaces over finite fields. *SIAM Journal on Discrete Mathematics*, 16 :276–287, 2001.

[BH05] P. Bas and J. Hurri. Security of dm quantization watermarking schemes : a practical study for digital images. In *Digital Watermarking, 4th International Workshop - IWDW'05*, volume 3710 of *Lecture Notes in Computer Science*, pages 186–200. Springer-Verlag, 2005.

[BHP08] O. Billet and D. Hieu Phan. Efficient traitor tracing from collusion secure codes. In *ICITS 2008*, volume 5155 of *Lecture Notes in Computer Science*, pages 171–182. Springer-Verlag, 2008.

[BHP09] O. Billet and D. Hieu Phan. Traitors collaborating in public : Pirates 2.0. In *Advances in Cryptology - EUROCRYPT'2009*, volume 5479 of *Lecture Notes in Computer Science*, pages 189–205. Springer-Verlag, 2009.

[Bie97] B. Biehl, I. et Meyer. Protocols for collusion-secure asymmetric fingerprinting. In *STACS'97*, volume 1200 of *Lecture Notes in Computer Science*, pages 399–412. Springer-Verlag, 1997.

[Bie98] J. Bierbrauer. On Crandall's problem. Personal communication, 1998. `http://www.ws.binghamton.edu/fridrich/covcodes.pdf`.

[Bie01] J. Bierbrauer. On Crandall's problem. Personal communication, 2001. `http://www.ws.binghamton.edu/fridrich/covcodes.pdf`.

[BK04] A. Barg and G. Kabatianski. A class of I.P.P. codes with efficient identification. *Journal of Complexity*, 20(2) :137–147, 2004.

[BM08] J. Barbier and E. Mayer. Non-malleable schemes resisting adaptive adversaries. In *Digital Watermarking, 8th International Workshop - IWDW'08*, volume 5450 of *Lecture Notes in Computer Science*, pages 240–253. Springer-Verlag, 2008. Best Paper Award.

[BMP86] G. Blakley, C. Meadows, and G. Purdy. Fingerprinting long forgiving messages. In *Advances in Cryptology - CRYPTO'85*, volume 218 of *Lecture Notes in Computer Science*, pages 180–189. Springer-Verlag, 1986.

[Böh09] R. Böhme. An epistemological approach to steganography. In *Information Hiding, 11th International Workshop - IH 2009*, volume 5806 of *Lecture Notes in Computer Science*, pages 15–30. Springer-Verlag, 2009.

[Böh10] R. Böhme. *Advanced Statistical Steganalysis*. Information Security and Cryptography. Springer-Verlag, 2010.

[BPW07] Y. Bo, L. Piyuan, and Z. Wenzheng. An efficient anonymous fingerprinting protocol. In *International Conference on Computational Intelligence and Security – ICCIS 2006*, volume 4456 of *Lecture Notes in Computer Science*, pages 824–832. Springer-Verlag, 2007.

[BRWF05] V. Bénony, F. Recher, E. Wegrzynowski, and C. Fontaine. Cryptanalysis of a particular case of Klimov-Shamir pseudo-random genera tor. In *SEquences and Their Applications – SETA 2004, Revised Selected Papers*, volume 3486 of *Lecture Notes in Computer Science*, pages 313–322. Springer-Verlag, 2005.

[BS95] D. Boneh and J. Shaw. Collusion-secure fingerprinting for digital data. In *Advances in Cryptology - CRYPTO'95*, volume 963 of *Lecture Notes in Computer Science*, pages 452–564. Springer-Verlag, 1995.

[BS98] D. Boneh and J. Shaw. Collusion-secure fingerprinting for digital data. *IEEE Transactions on Information Theory*, 44(5) :1897–1905, 1998.

[BT08] O. Blayer and T. Tassa. Improved versions of Tardos' fingerprinting scheme. *Designs, Codes and Cryptography*, 48 :79–103, 2008.

[BTFF+06] A. Benjelloun-Touimi, J.-B. Fischer, C. Fontaine, C. Giraud, and M. Milhau. Enhanced Security Architecture for Music Distribution on Mobile. In *European Symposium On Research In Computer Security – ESORICS 2006*, volume 4189 of *Lecture Notes in Computer Science*, pages 97–109. Springer-Verlag, 2006.

[BW04] R. Böhme and A. Westfeld. Exploiting preserved statistics for steganalysis. In *Information Hiding, 6th International Workshop - IH 2004*, volume 3220 of *Lecture Notes in Computer Science*, pages 82–96. Springer-Verlag, 2004.

[BW09] P. Bas and A. Westfeld. Two key estimation techniques for the broken-arrows watermarking scheme. In *ACM Multimedia & Security'09*. ACM Press, 2009.

[Cac98] C. Cachin. An information-theoretic model for steganography. In *Information Hiding, 2nd International Workshop - IH 1998*, volume 1525 of *Lecture Notes in Computer Science*, pages 306–318. Springer-Verlag, 1998.

[Cac04] C. Cachin. An information-theoretic model for steganography. *Information and Computation*, 192(1) :41–56, 2004.

[Cam00] J. Camenisch. Efficient anonymous fingerprinting with group signatures. In *Advances in Cryptology – ASIACRYPT'00*, volume 1976 of *Lecture Notes in Computer Science*, pages 415–428. Springer-Verlag, 2000.

[CB08] F. Cayre and P. Bas. Kerckhoffs-based embedding security classes for WOA data-hiding. *IEEE Transactions on Information Forensics and Security*, 3(1) :1–15, 2008.

[CCCF00a] A. Canteaut, C. Carlet, P. Charpin, and C. Fontaine. Fourier Spectrum of Optimal Boolean Functions via Kasami's Identities. In *Proceedings 2000 IEEE International Symposium on Information Theory*, page 183. IEEE, 2000.

[CCCF00b] A. Canteaut, C. Carlet, P. Charpin, and C. Fontaine. Propagation characteristics and correlation-immunity of highly nonlinear Boolean functions. In *Advances in Cryptology - EUROCRYPT'2000*, volume 1807 of *Lecture Notes in Computer Science*, pages 507–522, 2000.

[CCCF01] A. Canteaut, C. Carlet, P. Charpin, and C. Fontaine. On cryptographic properties of the cosets of R(1,m). *IEEE Transactions on Information Theory*, 47(4) :1494–1513, May 2001.

[CDF06] I. Cox, G. Doërr, and T. Furon. Watermarking is not cryptography. In *Digital Watermarking, 4th International Workshop - IWDW'05*, volume 4283 of *Lecture Notes in Computer Science*, pages 1–15. Springer-Verlag, 2006.

[CFF05a] F. Cayre, C. Fontaine, and T. Furon. A theoretical study of watermarking security. In *IEEE International Symposium on Information Theory 2005*. IEEE, 2005.

[CFF05b] F. Cayre, C. Fontaine, and T. Furon. Watermarking Attack : Security of WSS Techniques. In *International Workshop on Digital Watermarking – IWDW 2004*, volume 3304 of *Lecture Notes in Computer Science*, pages 171–183. Springer-Verlag, 2005. Best Paper Award.

[CFF05c] F. Cayre, C. Fontaine, and T. Furon. Watermarking security, part one : theory. In *IS&T/SPIE International Symposium on Electronic Imaging 2005 - Security, Steganography, and Watermarking of Multimedia Contents VII*, volume 5681 of *Proceedings of the SPIE*, pages 746–757. SPIE, 2005.

[CFF05d] F. Cayre, C. Fontaine, and T. Furon. Watermarking security, part two : practice. In *IS&T/SPIE International Symposium on Electronic Imaging 2005 - Security, Steganography, and Watermarking of Multimedia Contents VII*, volume 5681 of *Proceedings of the SPIE*, pages 758–768. SPIE, 2005.

[CFF05e] F. Cayre, C. Fontaine, and T. Furon. Watermarking Security : Theory and Practice. *IEEE Transactions on Signal Processing*, 53(10) :3976–3987, 2005. numéro spécial "Supplement on Secure Media III".

[CFF07] F. Cayre, C. Fontaine, and T. Furon. *Digital Audio Watermarking Techniques and Technologies : Applications and Benchmarking*, chapter Watermarking Security. Idea Group Publishing, 2007.

[CFF09] A. Charpentier, C. Fontaine, and T. Furon. Décodage EM du code de Tardos pour le fingerprinting. In *XIIe colloque GRETSI*, 2009.

[CFF10] A. Charpentier, C. Fontaine, and T. Furon. Décodage EM du code de Tardos pour le fingerprinting. *Traitement du signal*, 27(2) :127–147, 2010.

[CFFC11] A. Charpentier, C. Fontaine, T. Furon, and I. Cox. An asymmetric fingerprinting scheme based on tardos codes. In *13th Information Hiding – IH'11*, Lecture Notes in Computer Science. Springer-Verlag, 2011. to appear.

[CFG08] F. Cérou, T. Furon, and A. Guyader. Experimental Assessment of the Reliability for Watermarking and Fingerprinting Schemes. *EURASIP Journal on Information Security*, 2008 :Article ID 414962, 2008.

[CFNP00] B. Chor, A. Fiat, M. Naor, and B. Pinkas. Tracing traitors. *IEEE Transactions on Information Theory*, 46(3) :893–910, 2000.

[CFS01] N. Courtois, M. Finiasz, and N. Sendrier. How to achieve a McEliece-based digital signature scheme. In *Advances in Cryptology – ASIACRYPT 2001*, volume 2248 of *Lecture Notes in Computer Science*. Springer-Verlag, 2001.

[CJ98] I. Cox and J.-P. Linnartz. Some general methods for tampering with watermarks. *IEEE Journal on Selected Areas in Communications*, 16(4) :587–93, 1998. Special issue on copyright and privacy protection.

[CKLS97] I.J. Cox, J. Killian, T. Leighton, and T. Shamoon. Secure Spread Spectrum Watermarking for Multimedia. *IEEE Transactions on Image Processing*, 6(12) :1673–1687, 1997.

[CL97] I. Cox and J.-P. Linnartz. Public watermarks and resistance to tampering. In *IEEE ICIP'97*, pages 3–6. IEEE, 1997.

[CLdV] A. Canteaut and F. Lévy-dit Véhel. La cryptologie moderne. revue Armement. `http://www-roc.inria.fr/secret/Anne.Canteaut/crypto_moderne.pdf`.

[CM03a] R. Chandramouli and N. Memon. Steganography capacity : A steganalysis perspective. In *IS&T/SPIE International Symposium on Eletronic Imaging 2003 - Security and Watermarking of Multimedia Contents V*, volume 5020 of *Proceedings of the SPIE*, pages 173–177. SPIE, 2003.

[CM03b] N. Courtois and W. Meier. Algebraic attacks on stream ciphers with linear feedback. In *Advances in Cryptology – EUROCRYPT'03*, number 2656 in Lecture Notes in Computer Science, pages 345–359. Springer-Verlag, 2003.

[CMB01] I. Cox, M. Miller, and J. Bloom. *Digital watermarking*. Morgan Kaufmann Publishers, 2001. ISBN 1-55860-714-5.

[CMB+08] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker. *Digital Watermarking and Steganography*. Morgan Kaufmann Publishers, 2nd edition, 2008. ISBN-13 978-0123725851.

[CMM99] I.J. Cox, M.L. Miller, and A.L. McKellips. Watermarking as communications with side information. *Proceedings of the IEEE*, 87(7) :1127–1141, 1999. special issue on.

[CNBM97] S. Craver, N. Memon, B.-L. Yeo, and M.M. Yeung. On the invertibility of invisible watermarking technique. In *ICIP 1997*, pages 540–543. IEEE, 1997.

[Cos83] M.H.M. Costa. Writing on dirty papers. *IEEE Transactions on Information Theory*, 29(3) :439–441, 1983.

[Cou02] N. Courtois. Higher order correlation attacks, XL algorithm and cryptanalysis of TOYOCRYPT. In *International Conference on Information Security and Cryptology, ICISC 2002*, number 2587 in Lecture Notes in Computer Science, pages 182–199. Springer-Verlag, 2002.

[Cou03] N. Courtois. Fast algebraic attacks on stream ciphers with linear feedback. In *Advances in Cryptology – CRYPTO'03*, number 2729 in Lecture Notes in Computer Science, pages 177–194. Springer-Verlag, 2003.

[Cou05] N. Courtois. Algebraic attacks on combiners with memory and several outputs. In *International Conference on Information Security and Cryptology, ICISC 2004*, number 3506 in Lecture Notes in Computer Science. Springer-Verlag, 2005.

[CPFPG05] P. Comesaña, L. Pérez-Freire, and F. Pérez-González. The return of the sensitivity attack. In *Digital Watermarking, 4th International Workshop - IWDW'05*, volume 3710 of *Lecture Notes in Computer Science*. Springer-Verlag, 2005.

[CPG05] P. Comesaña and F. Pérez-Gonzáles. Fundamentals of data hiding security and their application to spread-spectrum analysis. In *Information Hiding, 7th International Workshop - IH 2005*, volume 3727 of *Lecture Notes in Computer Science*, pages 146–160. Springer-Verlag, 2005.

[CPG07a] P. Comesaña and F. Pérez-Gonzáles. On the capacity of stegosystems. In *Proc. of the ACM Multimedia and Security Workshop 2007*, pages 15–24. ACM, 2007.

[CPG07b] P. Comesaña and F. Pérez-González. Breaking the bows watermarking system : Key guessing and sensitivity attacks. *EURASIP Journal on Information Security*, 2007, 2007. Article ID 25308.

[Cra98] R. Crandall. Some notes on steganography. Posted on steganography mailing list, 1998. `http://os.inf.tu-dresden.de/~westfeld/crandall.pdf`.

[Cra99] S. Craver. Zero Knowledge watermark detection. In *Information Hiding 1999*, volume 1768 of *Lecture Notes in Computer Science*, pages 101–116. Springer-Verlag, 1999.

[CT00] A. Canteaut and M. Trabbia. Improved fast correlation attacks using parity-check equations of weight 4 and 5. In *Advances in Cryptology - EUROCRYPT'2000*, number 1807 in Lecture Notes in Computer Science, pages 573–588, 2000.

[CT05] C.K. Chu and W.G. Tzeng. Efficient k-out-of-n oblivious transfer schemes with adaptive and non-adaptive queries. In *Public Key Cryptography – PKC 2005*, volume 3386 of *LNCS*, pages 172–183. Springer-Verlag, 2005.

[CW99a] B. Chen and G. Wornell. An information-theoretic approach to the design of robust digital watermarking systems. In *IEEE ICASSP'99*. IEEE, 1999.

[CW99b] B. Chen and G.W. Wornell. Achievable performance of digital watermarking systems. In *IEEE International Conference on Multimedia Computing and Systems 1999*, pages 13–18. IEEE, 1999.

[CW01] B. Chen and G. Wornell. Quantization Index Modulation : A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory*, 47(4) :1423–1443, 2001.

[CWH04] Q. Cheng, Y. Wang, and T. Huang. Performance analysis and error exponents of asymmetric watermarking systems. *Signal Processing*, 84(8) :1429–1445, 2004.

[CXFF09] A. Charpentier, F. Xie, C. Fontaine, and T. Furon. Expectation Maximisation decoding of Tardos probabilistic fingerprinting code. In *IS&T/SPIE International Symposium on Electronic Imaging 2009 - Media Forensics and Security XI*, volume 7254 of *Proceedings of the SPIE*. SPIE, 2009.

[CZF+11] R. Cogranne, C. Zitzmann, L. Fillatre, F. Retraint, I. Nikiforov, and P. Cornu. A cover image model for reliable steganalysis. In *13th Information Hiding – IH'11*, Lecture Notes in Computer Science. Springer-Verlag, 2011. to appear.

[DBPP09] M. Deng, T. Bianchi, A. Piva, and B. Preneel. An efficient Buyer-Seller watermarking protocol based on composite signal representation. In *ACM MM&Sec'09*, pages 9–18. ACM, 2009.

[DD04a] G. Doërr and J.-L. Dugelay. Danger of low-dimensional watermarking subspaces. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2004*. IEEE, 2004.

[DD04b] G. Doërr and J.-L. Dugelay. Security pitfalls of frame-by-frame approaches to video watermarking. *IEEE Transactions on Signal Processing*, 52(10) :2955–2964, 2004.

[DDGM97] J.-F. Delaigle, C. De Vleeschouwer, F. Goffin, and B. Macq. Low Cost Watermarking Based on a Human Visual Model. In *Multimedia Applications, Services and Techniques - ECMAST '97*, volume 1242 of *Lecture Notes in Computer Science*, pages 153–168. Springer-Verlag, 1997.

[DDVM98] J.-F. Delaigle, C. De Vleeschouwer, and B.M. Macq. A psychovisual approach for digital picture watermarking. *Journal of Electronic Imaging*, 7(3) :628–640, 1998.

[DF99] J. Domingo-Ferrer. Anonymous fingerprinting based on committed oblivious transfer. In *International Workshop on Practice and Theory in Public Key Cryptography – PKC'99*, volume 1560 of *Lecture Notes in Computer Science*, pages 43–52. Springer-Verlag, 1999.

[DFF+01] J. Dittmann, N. Fates, C. Fontaine, F.A. Petitcolas, F. Raynal, M. Steinebach, and C. Seibel. Stirmark benchmark : audio watermarking attacks. In *International Conference on Information Technology : Coding and Computing, ITCC 2001*. IEEE computer society press, 2001. Special Session in Multimedia Security and Watermarking Applications.

[DGF+08] C. Dikici, C. Guillemot, C. Fontaine, K. Idrissi, and A. Baskurt. Dirty Paper Coding with Partial State Information. In *IEEE International Symposium on Image/Video Communications over fixed and mobile networks, ISIVC 2008*, 2008.

[DH76] W. Diffie and M.E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, IT-22(6) :644–654, 1976.

[DKM05] T. Das, H.J. Kim, and S. Maitra. Security evaluation of generalized patchwork algorithm from cryptanalytic viewpoint. In *KES 2005*, volume 1, pages 1240–1247, 2005.

[DLGP11] M. Desoubeaux, G. Le Guelvouit, and W. Puech. Probabilistic fingerprinting codes used to detect traitor zero-bit watermark. In *IS&T/SPIE International Symposium on Electronic Imaging 2011 - Media Watermarking, Security and Forensics III*, volume 7880 of *Proceedings of the SPIE*. SPIE, 2011.

[DM02] T. Das and S. Maitra. Cryptanalysis of optimal differential energy watermarking (dew) and a modified robust scheme. In *Progress*

in Cryptology - INDOCRYPT 2002, 3rd International Conference on Cryptology in India, volume 2551 of *Lecture Notes in Computer Science*, pages 135–148. Springer-Verlag, 2002.

[DM04] T. Das and S. Maitra. Cryptanalysis of "wavelet tree quantization" watermarking scheme. In *Distributed Computing - IWDC 2004*, volume 3326 of *Lecture Notes in Computer Science*, pages 75–96. Springer-Verlag, 2004.

[DM06] T. Das and S. Maitra. Analysis of the "wavelet tree quantization" watermarking strategy and a modified robust scheme. *Multimedia Systems*, 12(2) :151–163, 2006.

[DMM05] T.K. Das, S. Maitra, and J. Mitra. Cryptanalysis of optimal differential energy watermarking (DEW) and a modified robust scheme. *IEEE Transactions on Signal Processing*, 53(2) :768–775, 2005. numï£¡ro spï£¡cial "Supplement on Secure Media II".

[DMZ06] T.K. Das, S. Maitra, and J. Zhou. Cryptanalysis of chu's DCT based watermarking scheme. *IEEE Transactions on Multimedia*, 8(3) :629–632, 2006.

[DZ05] T. Das and J. Zhou. Cryptanalysis of barni et al. watermarking scheme. In *Progress in Cryptology - INDOCRYPT 2005, 6th International Conference on Cryptology in India*, volume 3797 of *Lecture Notes in Computer Science*, pages 197–208. Springer-Verlag, 2005.

[DZM05] T. Das, J. Zhou, and S. Maitra. Cryptanalysis of a wavelet based watermarking scheme. In *Digital Watermarking, 2nd International Workshop - IWDW'04*, volume 3304 of *Lecture Notes in Computer Science*, pages 192–203. Springer-Verlag, 2005.

[EBTG03] J. Eggers, R. Baüml, R. Tzschoppe, and B. Girod. Scalar Costa Scheme for Information Embedding. *IEEE Transactions on Signal Processing*, 51(4) :1003–1019, 2003. numï£¡ro spï£¡cial "signal processing for data hiding in digital media and secure content delivery".

[ECM07a] M. El Choubassi and P. Moulin. Noniterative algorithms for sensitivity analysis attacks. *IEEE Transactions on Information Forensics and Security*, 2(2) :113–126, 2007.

[ECM07b] M. El Choubassi and P. Moulin. Sensitivity analysis attacks against randomized detectors. In *IEEE ICIP'07*, volume 2, pages 129–132. IEEE, 2007.

[ECM09a] M. El Choubassi and P. Moulin. Joint detection-estimation games for sensitivity analysis attacks. In *IS&T/SPIE International Symposium on Electronic Imaging 2009 - Media Forensics and Security*, volume 7254 of *Proceedings of the SPIE*. SPIE, 2009.

[ECM09b] M. El Choubassi and P. Moulin. On reliability and security of randomized detectors against sensitivity analysis attacks. *IEEE Transactions on Information Forensics and Security*, 4(3) :273–283, 2009.

[EG00] J. Eggers and B. Girod. Public key watermarking by eigenvectors of linear transforms. In *EUSIPCO'00*, 2000.

[Ett98] J.M. Ettinger. Steganalysis and game equilibria. In *Information Hiding, 2nd International Workshop - IH 1998*, volume 1525 of *Lecture Notes in Computer Science*, pages 319–328. Springer-Verlag, 1998.

[FB08] T. Furon and P. Bas. Broken Arrows. *EURASIP Journal on Information Security*, 2008 :Article ID 597040, 2008.

[FD00a] T. Furon and P. Duhamel. An asymmetric public detection watermarking technique. In *Information Hiding, 3rd International Workshop - IH 1999*, volume 1768 of *Lecture Notes in Computer Science*, pages 88–100. Springer-Verlag, 2000.

[FD00b] T. Furon and P. Duhamel. Robustness of an asymmetric technique. In *IEEE ICIP'00*, volume 3, pages 21–24. IEEE, 2000.

[FD03] T. Furon and P. Duhamel. An Asymmetric Watermarking Method. *IEEE Transactions on Signal Processing*, 51(4) :981–995, 2003. numéro spécial "signal processing for data hiding in digital media and secure content delivery".

[FDD+08] C. Fontaine, C. Delpha, P. Duhamel, A. Benjelloun Touimi, M. Milhau, A. Le Guyader, C. Giraud, and D. Martin. An end-to-end security architecture for multimedia content distribution on mobile phones. *ISAST Transactions on Communication and Networking*, 2(1) :81–91, 2008.

[FF98] E. Filiol and C. Fontaine. Highly nonlinear balanced Boolean functions with a good correlation-immunity. In *Advances in Cryptology - EUROCRYPT'98*, volume 1403 of *Lecture Notes in Computer Science*, pages 475–488. Springer-Verlag, 1998.

[FF01] E. Filiol and C. Fontaine. A new fast stream cipher design : COS ciphers. In *8th IMA Conference on Cryptography and Coding*, volume 2260 of *Lecture Notes in Computer Science*, pages 85–98. Springer-Verlag, 2001.

[FF07] J. Fridrich and T. Filler. Practical methods for minimizing embedding impact in steganography. In *IS&T/SPIE International Symposium on Electronic Imaging 2007 - Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505 of *Proceedings of the SPIE*. SPIE, 2007.

[FF09a] T. Filler and J. Fridrich. Complete characterization of perfectly secure stego-systems with mutually inedpendent embedding operation. In *IEEE ICASSP'09*, pages 1429–1432. IEEE, 2009.

[FF09b] T. Filler and J. Fridrich. Wet ZZW construction for steganography. In *IEEE International Workshop on Information Forensics and Security - WIFS 2009*, pages 131–135, 2009.

[FF10a] T. Filler and J. Fridrich. Gibbs construction in steganography. *IEEE Transactions on Information Forensics and Security*, 5(4) :705–720, 2010.

[FF10b] T. Filler and J. Fridrich. Minimizing additive distortion functions with non-binary embedding operation in steganography. In *IEEE International Workshop on Information Forensics and Security - WIFS 2010*. IEEE, 2010.

[FF11] T. Filler and J. Fridrich. Design of adaptive steganographic schemes for digital images. In *IS&T/SPIE International Symposium on Electronic Imaging 2011 - Media Watermarking, Security and Forensics III*, volume 7880 of *Proceedings of the SPIE*. SPIE, 2011.

[FFJ04] E. Filiol, C. Fontaine, and S. Josse. The COSvd ciphers. In *SASC : the State of the Art of Stream Ciphers*, 2004.

[FFV01] E. Filiol, C. Fontaine, and D. Vianne. A new fast block cipher design : COS ciphers. In *IEEE International Symposium on Information Theory 2001*, page 138. IEEE, 2001.

[FG03] J. Fridrich and M Goljan. Digital image steganography using stochastic modulation. In *IS&T/SPIE International Symposium on Eletronic Imaging 2003 - Security and Watermarking of Multimedia Contents V*, volume 5020 of *Proceedings of the SPIE*, pages 191–202. SPIE, 2003.

[FG04] J. Fridrich and D. Goljan, M nd Soukal. Perturbed quantization steganograhy using wet paper codes. In *ACM MM&Sec'04*, pages 4–15. ACM, 2004.

[FG07a] C. Fontaine and F. Galand. A survey of homomorphic encryption for non-specialists. *EURASIP Journal on Information Security*, 2007 :Article ID 13801, 2007. special issue "Signal Processing in the Encrypted Domain" ; freely downloadable at http ://www.hindawi.com/GetArticle.aspx ?doi=10.1155/2007/13801.

[FG07b] C. Fontaine and F. Galand. How can Reed-Solomon codes improve steganographic schemes. In *9th Information Hiding – IH'07*, volume 4567 of *Lecture Notes in Computer Science*, pages 130–144. Springer-Verlag, 2007.

[FG09] C. Fontaine and F. Galand. How Reed-Solomon Codes Can Improve Steganographic Schemes. *EURASIP Journal on Information Security*, 2009 :Article ID 274845, 2009. special issue "Secure Steganography in Multimedia Content" ; freely downloadable at http ://www.hindawi.com/GetArticle.aspx ?doi=10.1155/2009/274845.

[FGB11] N. Friot, C. Guyeux, and J. Bahi. Chaotic iterations for steganography – stego-security and chaos-security. In *International Conference on Security and Cryptography – SECRYPT'2011*, 2011.

[FGC08] T. Furon, A. Guyader, and F. Cérou. On the Design and Optimization of Tardos Probabilistic Fingerprinting Codes. In *Information Hiding, 10th International Workshop - IH 2008*, volume 5284 of *Lecture Notes in Computer Science*, pages 341–356. Springer-Verlag, 2008.

[FGD01a] J. Fridrich, M. Goljan, and R. Du. Detecting LSB steganography in color and gray-scale images. *Magazine of IEEE multimedia*, 8(4) :22–28, 2001. Special issue on security.

[FGD01b] J. Fridrich, M Goljan, and R. Du. Steganalysis based on JPEG compatibility. In *SPIE Multimedia Systems and Applications IV*, volume 4518 of *Proceedings of the SPIE*, pages 275–280. SPIE, 2001.

[FGH02] J. Fridrich, M. Goljan, and D. Hogea. Steganalysis of jpeg images : breaking the f5 algorithm. In *Information Hiding, 5th International Workshop - IH 2002*, volume 2578 of *Lecture Notes in Computer Science*, pages 310–323, 2002.

[FGHS03] J. Fridrich, M. Goljan, D. Hogea, and D. Soukal. Quantitative steganalysis of digital images : estimating the secret message length. *ACM Multimedia Systems Journal*, 9(3) :288–302, 2003.

[FGKH11] J. Fridrich, M. Goljan, J. Kodovský, and V. Holub. Steganalysis of content-adaptive steganography in spatial domain. In *13th Information Hiding – IH'11*, Lecture Notes in Computer Science. Springer-Verlag, 2011. to appear.

[FGLS05] J. Fridrich, M. Goljan, P. Lisoněk, and D. Soukal. Writing on Wet Paper. *IEEE Transactions on Signal Processing*, 53(10) :3923–3935, 2005. special issue "Supplement on Secure Media III".

[FGS05a] J. Fridrich, M. Goljan, and D. Soukal. Efficient wet paper codes. In *Information Hiding, 7th International Workshop - IH 2005*, volume 3727 of *Lecture Notes in Computer Science*, pages 204–218. Springer-Verlag, 2005.

[FGS05b] J. Fridrich, M. Goljan, and D. Soukal. Perturbed Quantization Steganography. *ACM Multimedia and Security Journal*, 11(2) :98–107, 2005.

[FGS06] J. Fridrich, M. Goljan, and D. Soukal. Wet paper Codes with Improved Embedding Efficiency. *IEEE Transactions on Information Forensics and Security*, 1(1) :102–110, 2006.

[FJF10] T. Filler, J. Judas, and J. Fridrich. Minimizing embedding impact in steganography using trellis-coded quantization. In *IS&T/SPIE*

International Symposium on Electronic Imaging 2010 - Media Forensics and Security II, volume 7541 of *Proceedings of the SPIE*. SPIE, 2010.

[FJF11] T. Filler, J. Judas, and J. Fridrich. Minimizing additive distortion in steganography unsing syndrome-trellis codes. *IEEE Transactions on Information Forensics and Security*, 2011.

[FKF09] T. Filler, A. Ker, and J. Fridrich. The square root law of steganographic capacity for markov covers. In *IS&T/SPIE International Symposium on Electronic Imaging 2009 - Media Forensics and Security*, volume 7254 of *Proceedings of the SPIE*. SPIE, 2009.

[FKGH11] J. Fridrich, J. Kodovský, M. Goljan, and V. Holub. Breaking hugo - the process discovery. In *13th Information Hiding – IH'11*, Lecture Notes in Computer Science. Springer-Verlag, 2011. to appear.

[FL07] J. Fridrich and P. Lisoněk. Grid Colorings in Steganography. *IEEE Transactions on Information Theory*, 53(4) :1547–1549, 2007.

[FLS06] J. Fridrich, P. Lisoněk, and D. Soukal. On steganographic embedding efficiency. In *Information Hiding, 8th International Workshop - IH 2006*, volume 4437 of *Lecture Notes in Computer Science*, pages 282–296. Springer-Verlag, 2006.

[FMD00] T. Furon, N. Moreau, and P. Duhamel. Audio public key watermarking technique. In *IEEE ICASSP'00*, volume 4, pages 1959–1962. IEEE, 2000.

[Fon98] C. Fontaine. *Contribution à la recherche de fonctions booléennes hautement non-linéaires, et au marquage d'images numériques en vue de la protection des droits d'auteur*. PhD thesis, Université de Paris 6, 1998.

[Fon99] C. Fontaine. On some cosets of the first-order Reed-Muller code with high minimum weight. *IEEE Transactions on Information Theory*, 45(4) :1237–1243, 1999.

[FPF09a] T. Furon and L. Pérez-Freire. EM decoding of Tardos traitor tracing codes. In *ACM Multimedia & Security'09*, pages 99–106. ACM, 2009.

[FPF09b] T. Furon and L. Pérez-Freire. Worst case attacks against binary probabilistic traitor tracing codes. In *IEEE International Workshop on Information Forensics and Security - WIFS 2009*. IEEE, 2009.

[FPFGC09] T. Furon, L. Pérez-Freire, A. Guyader, and F. Cérou. Estimating the Minimal Length of Tardos Code. In *Information Hiding, 11th International Workshop - IH 2009*, volume 5806 of *Lecture Notes in Computer Science*, pages 176–190. Springer-Verlag, 2009.

[FR02] C. Fontaine and F. Raynal. About the links between cryptography and information hiding. In *IS&T/SPIE International Symposium on Electronic Imaging 2002 - Security and Watermarking of Multimedia Contents IV*, volume 4675 of *Proceedings of the SPIE*, pages 269–280. SPIE, 2002.

[Fra02] E. Franz. Steganography preserving statistical properties. In *Information Hiding, 5th International Workshop - IH 2002*, volume 2578 of *Lecture Notes in Computer Science*, pages 278–294. Springer-Verlag, 2002.

[Fre82] H. Fredricksen. A survey of full length nonlinear shift register cycle algorithms. *SIAM Review*, 24(2) :195–221, 1982.

[Fri09a] J. Fridrich. Asymptotic behavior of the ZZW embedding construction. *IEEE Transactions on Information Forensics and Security*, 4(1) :151–153, 2009.

[Fri09b] J. Fridrich. *Steganography in Digital Media : Principles, Algorithms, and Applications.* Cambridge University Press, 2009.

[FS04a] M. Fernandez and M. Soriano. Identification of traitors in algebraic-geometric traceability codes. *IEEE Transactions on Signal Processing*, 52(10) :3073–3077, 2004. numï£¡ro spï£¡cial "Supplement on Secure Media".

[FS04b] M. Fernandez and M. Soriano. Identification of traitors using a trellis. In *ICICS 2004*, volume 3269 of *Lecture Notes in Computer Science*, pages 211–222. Springer-Verlag, 2004.

[FS04c] M. Fernandez and M. Soriano. Soft-decision tracing in fingerprinted multimedia content. *IEEE Multimedia*, 11(2) :38–46, 2004.

[FS05] E. Franz and A. Schneidewind. Pre-processing for adding noise steganograhy. In *Information Hiding, 7th International Workshop - IH 2005*, volume 3727 of *Lecture Notes in Computer Science*, pages 189–203. Springer-Verlag, 2005.

[FS06] J. Fridrich and D. Soukal. Matrix embedding for large payloads. *IEEE Transactions on Information Forensics and Security*, 1(3) :390–394, 2006.

[FVD01] T. Furon, I. Venturini, and P. Duhamel. An unified approach of asymmetric watermarking schemes. In *IS&T/SPIE International Symposium on Eletronic Imaging 2001 - Security and Watermarking of Multimedia Contents III*, volume 4518 of *Proceedings of the SPIE*. SPIE, 2001.

[Gal05] F. Galand. Practical Construction Against Theoretical Approach in Fingerprinting. In *IEEE International Symposium on Information Theory – ISIT 2005*. IEEE, 2005.

[GFB10] C. Guyeux, N. Friot, and J. Bahi. Chaotic iterations versus spread-spectrum : chaos and stego-security. In *Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing – IIH-MSP 2010*, pages 208–211. IEEE Computer Society Press, 2010.

[GFD02] P. Guillon, T. Furon, and P. Duhamel. Applied public-key steganography. In *IS&T/SPIE International Symposium on Electronic Imaging 2002 - Security and Watermarking of Multimedia Contents IV*, volume 4675 of *Proceedings of the SPIE*, pages 38–49. SPIE, 2002.

[GFF+06] C. Giraud, J.-B. Fischer, C. Fontaine, A. Benjelloun-Touimi, M. Milhau, and B. Prady. Dispositif de restitution d'un contenu numérique, entité électronique sécurisée comprenant ces éléments et procédé de restitution d'un contenu numérique. Patent request 0651089, March 2006.

[GFH06] M. Goljan, J. Fridrich, and T. Holotyak. New blind steganalysis and its implications. In *IS&T/SPIE International Symposium on Electronic Imaging 2006 - Security, Steganography, and Watermarking of Multimedia Contents VIII*, volume 6072 of *Proceedings of the SPIE*, pages 1–13. SPIE, 2006.

[GG05] B.M. Gammel and R. Göttfert. Linear filtering of nonlinear shift register sequences. In *Workshop on Coding and Cryptography, WCC 2005*, pages 117–126, 2005.

[GH07] M. Green and S. Hohenberger. Blind identity-based encryption and simulatable oblivious transfer. In *Advances in Crypotology – ASIACRYPT 2007*, volume 4833 of *LNCS*, pages 265–282. Springer-Verlag, 2007.

[GK03] F. Galand and G. Kabatiansky. Information hiding by coverings. In *Proc. ITW 2003*, pages 151–154, 2003.

[GK09] F. Galand and G. Kabatiansky. Coverings, centered codes, and combinatorial steganography. *Problems of Information Transmission*, 45(3) :289–297, 2009.

[GK11] G. Gül and F. and Kurugollu. A new methodology in steganalysis : Breaking highly undetactable steganograpy (hugo). In *13th Information Hiding – IH'11*, Lecture Notes in Computer Science. Springer-Verlag, 2011. to appear.

[GP80] S.I. Gel'fand and M.S. Pinsker. Coding for channel with random parameters. *Problems of Control and Information Theory*, 9(1) :19–31, 1980.

[GP00] H.J. Guth and B. Pfitzmann. Error- and collusion-secure fingerprinting for digital data. In *Information Hiding, 3rd International Work-*

shop - IH 1999, volume 1768 of *Lecture Notes in Computer Science*, pages 134–145. Springer-Verlag, 2000.

[HC05] H.F. Huang and C.C. Chang. A new design for efficient t-out-n oblivious transfer scheme. In *19th International Conference on Advanced Information Networking and Applications (AINA'05)*. IEEE Computer Society, 2005.

[HG97] F. Hartung and B. Girod. Fast public-key watermarking of compressed video. In *IEEE ICIP'97*, volume 1, pages 528–531. IEEE, 1997.

[HHI06] M. Hagiwara, G. Hanaoka, and H. Imai. A short random fingerprinting code against a small number of pirates. In *AAECC 2006*, volume 3857 of *Lecture Notes in Computer Science*, pages 193–202. Springer-Verlag, 2006.

[HJJ02] G. Hachez and Quisquater J.-J. Which directions for asymmetric watermarking ? In *EUSIPCO'02*, 2002.

[HLvA02] N.J. Hopper, J. Langford, and L. von Ahn. Provably secure steganography. In *Advances in Cryptology - CRYPTO'02*, number 2442 in Lecture Notes in Computer Science. Springer-Verlag, 2002.

[HLvA09] N.J. Hopper, J. Langford, and L. von Ahn. Provably secure steganography. *IEEE Transactions on Computers*, 58(5) :662–676, 2009.

[HLW06] C. Hundt, M. Liskiewicz, and U. Wölfel. Provably secure steganography and the complexity of sampling. In *ISAAC 2006*, volume 4317 of *Lecture Notes in Computer Science*, pages 754–763. Springer-Verlag, 2006.

[HM08] Y.-W. Huang and P. Moulin. Universal fingerprinting : Capacity and random-coding exponents. In *IEEE International Symposium on Information Theory, ISIT 2009*, pages 220–224. IEEE, 2008.

[HM09] Y.-W. Huang and P. Moulin. Saddle-point solution of the fingerprinting capacity game under the marking assumption. In *IEEE International Symposium on Information Theory, ISIT 2009*, pages 2256–2260. IEEE, 2009.

[HM10] Y.-W. Huang and P. Moulin. Maximin optimality of the arcsine fingerprinting distribution and the interleaving attack for large coalitions. In *IEEE International Workshop on Information Forensics and Security - WIFS 2010*. IEEE, 2010.

[HP03] W.C. Huffman and V. Pless. *Fundamentals of Error-Correcting Codes*. Cambridge University Press, 2003.

[HR04] P. Hawkes and G.G. Rose. Rewrinting variables : the complexity of fast algebraic attacks on stream ciphers. In *Advances in Cryptology*

– *CRYPTO'04*, number 3152 in Lecture Notes in Computer Science, pages 390–406. Springer-Verlag, 2004.

[Hua90] Y. Huang. A new algorithm for the generation of binary de Bruijn sequences. *Journal of Algorithms*, 11 :44–51, 1990.

[HvLLT98] H.D.L. Hollmann, J.H. van Lint, J.-P. Linnartz, and L.M.G.M. Tolhuizen. On codes with identifiable parent property. *Journal of Combinatorial Theory*, 82 :121–133, 1998.

[HW05] S. He and M. Wu. Performance study on multimedia fingerprinting employing traceability codes. In *Digital Watermarking, 4th International Workshop - IWDW'05*, volume 3710 of *Lecture Notes in Computer Science*, pages 84–96. Springer-Verlag, 2005.

[HW06] S. He and M. Wu. Joint Coding and Embedding Techniques for Multimedia Fingerprinting. *IEEE Transactions on Information Forensics and Security*, 1(2) :231–247, 2006.

[HW07] S. He and M. Wu. Collusion-Resistant Video Fingerprinting for Large User Group. *IEEE Transactions on Information Forensics and Security*, 2(4) :697–709, 2007.

[HYF08] K. Haramura, M. Yoshida, and T. Fujiwara. Anonymous fingerprinting for predelivery of contents. In *Information Security and Cryptology – ICISC 2008*, volume 5461 of *Lecture Notes in Computer Science*, pages 134–151. Springer-Verlag, 2008.

[JJ02] F. Jönnson and T. Johansson. A fast correlation attack on LILI-128. *Information Processing Letters*, 81(3) :127–132, 2002.

[Kal01] T. Kalker. Considerations on watermarking seccurity. In *Proc of the IEEE Multimedia Signal Processing workshop*, pages 201–206, Cannes, France, October 2001.

[KDR06] Y. Kim, Z. Duric, and D. Richards. Modified matrix encoding technique for minimal distortion steganography. In *Information Hiding, 8th International Workshop - IH 2006*, volume 4437 of *Lecture Notes in Computer Science*. Springer-Verlag, 2006.

[Ker83a] A. Kerckhoffs. La cryptographie militaire (part i). *Journal des sciences militaires*, 9(1) :5–38, 1883. in French, available in English at `http://www.petitcolas.net/fabien/kerckhoffs/`.

[Ker83b] A. Kerckhoffs. La cryptographie militaire (part ii). *Journal des sciences militaires*, 9(2) :161–191, 1883. in French, available in English at `http://www.petitcolas.net/fabien/kerckhoffs/`.

[Ker05] A. Ker. A general framework for structural analysis of lsb replacement. In *Information Hiding, 7th International Workshop - IH 2005*, volume 3727 of *Lecture Notes in Computer Science*, pages 296–311. Springer-Verlag, 2005.

[Ker07a] A. Ker. A capacity result for batch steganography. *IEEE Signal Processing Letters*, 14(8) :525–528, 2007.

[Ker07b] A. Ker. A fusion of maximum likelihood and structural steganalysis. In *Information Hiding, 9th International Workshop - IH 2007*, volume 4567 of *Lecture Notes in Computer Science*, pages 204–219. Springer-Verlag, 2007.

[Ker07c] A. Ker. The ultimate steganalysis benchmark ? In *Proc. of the ACM Multimedia and Security Workshop 2007*, pages 141–147. ACM, 2007.

[Ker08a] A. Ker. Locating steganographic payload via ws residuals. In *Proc. of the ACM Multimedia and Security Workshop 2008*, pages 27–31. ACM, 2008.

[Ker08b] A. Ker. Steganographic strategies for a square distortion function. In *IS&T/SPIE International Symposium on Electronic Imaging 2008 - Security, Forensics,Steganography and Watermarking of Multimedia Contents X*, volume 6819 of *Proceedings of the SPIE*. SPIE, 2008.

[Ker09a] A. Ker. Estimating the information theoretic optilmal stego noise. In *Digital Watermarking, 8th International Workshop - IWDW'09*, volume 5703 of *Lecture Notes in Computer Science*, pages 184–198. Springer-Verlag, 2009.

[Ker09b] A. Ker. The square root law requires a linear key. In *Proc. of the ACM Multimedia and Security Workshop 2009*, pages 85–92. ACM, 2009.

[Ker10a] A. Ker. The square root law does not require a linear key. In *Proc. of the ACM Multimedia and Security Workshop 2010*, pages 213–223. ACM, 2010.

[Ker10b] A. Ker. The square root law in stegosystems with imperfect information. In *Information Hiding, 12th International Workshop - IH 2010*, volume 6387 of *Lecture Notes in Computer Science*. Springer-Verlag, 2010.

[KF08] J. Kodovský and J. Fridrich. On completeness of feature spaces in blind steganalysis. In *Proc. of the ACM Multimedia and Security Workshop 2008*, pages 123–132. ACM, 2008.

[KF10] J. Kodovský and J. Fridrich. Quantitative structural steganalysis of jsteg. *IEEE Transactions on Information Forensics and Security*, 5(4) :681–693, 2010.

[KF11] J. Kodovský and J. Fridrich. Steganalysis in high dimensions : Fusing classifiers built on random spaces. In *IS&T/SPIE International Symposium on Electronic Imaging 2011 - Media Watermarking, Security and Forensics III*, volume 7880 of *Proceedings of the SPIE*. SPIE, 2011.

[KFP07] J. Kodovský, J. Fridrich, and T. Pevný. Statistically undetectable jpeg steganography : Dead ends, challenges, and opportunities. In *Proc. of the ACM Multimedia and Security Workshop 2007*, pages 3–14. ACM, 2007.

[KH07] T. Kitagawa and M. Hagiwara. A short random fingerprinting code and its tracing algorithm. In *IEICE and SITA Joint Conference on Information Theory*, 2007.

[KHN+08] T. Kitagawa, M. Hagiwara, K. Nuida, H. Watanabe, and H. Imai. A Group Testing Based Deterministic Tracing Algorithm for a Short Random Fingerprint Code. In *International Symposium on Information Theory and its Applications - ISITA 2008*, 2008.

[KLvD98] T. Kalker, J.-P. Linnartz, and M. van Dijk. Watermark estimation through detector analysis. In *IEEE ICIP'98*, volume 1, pages 425–429. IEEE, 1998.

[KP99] S. Katzenbeisser and F.A.P. Petitcolas. *Information hiding techniques for steganography and digital watermarking.* Artech House Publishers, 1999. ISBN 1-58053-035-4.

[KP02] S. Katzenbeisser and F. Petitcolas. On defining security in steganographic systems. In *IS&T/SPIE International Symposium on Electronic Imaging 2002 - Security and Watermarking of Multimedia Contents IV*, volume 4675 of *Proceedings of the SPIE*, pages 50–56. SPIE, 2002.

[KP11] A. Ker and T. Pevny. A new paradigm for steganalysis via clustering. In *IS&T/SPIE International Symposium on Electronic Imaging 2011 - Media Watermarking, Security and Forensics III*, volume 7880 of *Proceedings of the SPIE.* SPIE, 2011.

[KS02] A. Klimov and A. Shamir. A New Class of Invertible Mappings. In *CHESS 2002*, number 2523 in Lecture Notes in Computer Science, pages 470–483. Springer-Verlag, 2002.

[KS04a] A. Klimov and A. Shamir. Cryptocraphic applications of T-functions. In *Selected Areas in Cryptography – SAC 2003*, number 3006 in Lecture Notes in Computer Science, pages 248–261. Springer-Verlag, 2004.

[KS04b] A. Klimov and A. Shamir. New Cryptographic Primitives Based on Multiword T-functions. In *Fast Software Encryption – FSE 2004*, number 3017 in Lecture Notes in Computer Science, pages 1–15. Springer-Verlag, 2004. invited talk.

[KSA00] M. Kutter, S. Voloshynovskiy, and A. Herrigel. Watermark copy attack. In *IS&T/SPIE International Symposium on Eletronic Imaging 2000 - Security and Watermarking of Multimedia Contents II*, volume 3971 of *Proceedings of the SPIE.* SPIE, 2000.

[Kur10] M. Kuribayashi. On the implementation of spread spectrum fingerprinting in asymmetric cryptographic protocol. *EURASIP Journal on Information Security*, 2010, 2010.

[KvCS07] S. Katzenbeisser, B. Škorić, M. Celik, and A.-R. Sadeghi. Combining Tardos fingerprinting codes and fingercasting. In *Information Hiding, 9th International Workshop - IH 2007*, volume 4567 of *Lecture Notes in Computer Science*, pages 294–310. Springer-Verlag, 2007.

[KZ95] E. Koch and J. Zhao. Towards Robust and Hidden Image Copyright Labeling. In *IEEE Workshop on Nonlinear Signal and Image Processing*, pages 452–455, 1995.

[LF02] S. Lyu and H. Farid. Detecting hidden messages using higher-order statistics and support vector machines. In *Information Hiding, 5th International Workshop - IH 2002*, volume 2578 of *Lecture Notes in Computer Science*, pages 340–354. Springer-Verlag, 2002.

[LF06] S. Lyu and H. Farid. Steganalysis using higher-order image statistics. *IEEE Transactions on Information Forensics and Security*, 1(1) :111–119, 2006.

[LvD98] J.-P. Linnartz and M. van Dijk. Analysis of the Sensitivity Attack Against Electronic Watermarks in Images. In *Information Hiding, 2nd International Workshop - IH 1998*, volume 1525 of *Lecture Notes in Computer Science*, pages 258–272. Springer-Verlag, 1998.

[LYTC04] C.-L. Lei, P.-L. Yu, P.-L. Tsai, and M.-H. Chan. An efficient and anonymous buyer-seller watermarking protocol. *IEEE Transactions on Image Processing*, 13(12) :1618–1626, 2004.

[Mas69] J.L. Massey. Shift-register synthesis and BCH decoding. *IEEE Transactions on Information Theory*, IT-15 :122–127, 1969.

[MB] C. Munuera and M. Barbier. Wet paper codes and the dual distance in steganography. soumis.

[MBCM09a] B. Mathon, P. Bas, F. Cayre, and B. Macq. Comparison of secure spread-spectrum modulations applied to still image watermarking. *Annals of Telecommunications*, 64 :801–813, 2009.

[MBCM09b] B. Mathon, P. Bas, F. Cayre, and B. Macq. Optimization of natural watermarking using transportation theory. In *ACM Multimedia & Security'09*. ACM Press, 2009.

[MBCM10a] B. Mathon, P. Bas, F. Cayre, and B. Macq. Considering security and robustness constraints for watermark-based tardos fingerprinting. In *IEEE International Workshop on Multimedia Signal Processing – MMSP'10*, pages 46–51. IEEE, 2010. Top 10% Paper Award.

[MBCM10b] B. Mathon, P. Bas, F. Cayre, and B. Macq. Security and robustness constraints for spread-spectrum tardos fingerprinting. In *IEEE International Workshop on Information Forensics and Security - WIFS 2010*. IEEE, 2010.

[MBCPG08] B. Mathon, P. Bas, F. Cayre, and F. Pérez-González. Distortion optimization of model-based secure embedding schemes for data-hiding. In *Information Hiding, 10th International Workshop - IH 2008*, volume 5284 of *Lecture Notes in Computer Science*. Springer-Verlag, 2008.

[MCB00] M. Miller, I. Cox, and J. Bloom. Informed embedding : exploiting image and detector information during watermark insertion. In *IEEE ICIP'00*, volume 3, pages 1–4, 2000.

[MCB07] B. Mathon, F. Cayre, and P. Bas. Practical performance analysis of secure modulations for woa spread-spectrum based image watermarking. In *Proc. of the ACM Multimedia and Security Workshop 2007*. ACM, 2007.

[McL84] A. McLoughlin. The complexity of computing the covering radius of a code. *IEEE Transactions on Information Theory*, 30(6) :800–804, 1984.

[MDC04] M.L. Miller, G.J. Doërr, and I.J. Cox. Applying Informed Coding and Embedding to Design a Robust, High capacity Watermark. *IEEE Transactions on Image Processing*, 13(6) :792–807, 2004.

[MF04] H. Malvar and D. Florencio. Improved spread spectrum : a new modulation technique for robust watermarking. *IEEE Transactions on Signal Processing*, 51(4) :898–905, 2004.

[MF11a] P. Meerwald and Furon. Towards joint tardos decoding : The 'don quixote' algorithm. In *13th Information Hiding – IH'11*, Lecture Notes in Computer Science. Springer-Verlag, 2011. to appear.

[MF11b] P. Meerwald and T. Furon. Group testing meets trator tracing. In *ICASSP'11*, pages 4204–4207. IEEE, 2011.

[MI03] P. Moulin and A. Ivanović. The zero-rate spread-spectrum watermarking game. *IEEE Transactions on Signal Processing*, 51(4) :1098–1117, 2003.

[Mie06] J. Mielikainen. Lsb matching revisited. *IEEE Signal Processing Letters*, 13(5) :285–287, 2006.

[Mit99] T. Mittelholzer. An information-theoretic approach to steganography and watermarking. In *Information Hiding 1999*, volume 1768 of *Lecture Notes in Computer Science*, pages 1–16. Springer-Verlag, 1999.

[MK05] P. Moulin and R. Koetter. Data-hiding codes. *Proceedings of the IEEE*, 93(12) :2083–2126, 2005.

[MO03] P. Moulin and J. O'Sullivan. Information-theoretic analysis of information hiding. *IEEE Transactions on Information Theory*, 49(3) :563–593, 2003.

[Mou01] P. Moulin. The role of information theory in watermarking and its application to image watermarking. *Signal Processing*, 81 :1121–1139, 2001.

[MS88] W. Meier and O. Staffelbach. Fast Correlation attacks on stream ciphers. In C.G. Günther, editor, *Advances in Cryptology - EUROCRYPT'88*, number 330 in Lecture Notes in Computer Science, pages 301–314. Springer-Verlag, 1988.

[MvOV97] A.J. Menezes, P.C. van Orschot, and S.A. Vanstone. *Handbook of applied cryptography*. CRC Press, 1997. accessible on line for free at `http://www.cacr.math.uwaterloo.ca/hac/`.

[MW01] N. Memon and P.W. Wong. A buyer-seller watermarking protocol. *IEEE Transactions on Image Processing*, 10(4) :643–649, 2001.

[MW04] P. Moulin and Y. Wang. New results on steganographic capacity. In *Conference on Information Sciences and Systems*, pages 813–818, 2004.

[MW07] P. Moulin and Y Wang. Capacity and random-coding exponents for channel coding with side information. *IEEE Transactions on Information Theory*, 53(4) :1326–1347, 2007.

[NFH+07] K. Nuida, S. Fujitsu, M. Hagiwara, T. Kitagawa, H. Watanabe, K. Ogawa, and H. Imai. An improvement of Tardos's collusion-secure fingerprinting codes with very short lengths. In *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes - AAECC 17*, volume 4851 of *Lecture Notes in Computer Science*, pages 80–89. Springer-Verlag, 2007.

[NFH+09] K. Nuida, S. Fujitsu, M. Hagiwara, T. Kitagawa, H. Watanabe, K. Ogawa, and H. Imai. An Improvement of Discrete Tardos Fingerprinting Codes. *Designs, Codes and Cryptography*, 52(3) :339–362, 2009.

[NHWI07] K. Nuida, M. Hagiwara, H. Watanabe, and H. Imai. Optimization of Tardos's fingerprinting codes in a viewpoint of memory amount. In *Information Hiding, 9th International Workshop - IH 2007*, volume 4567 of *Lecture Notes in Computer Science*, pages 279–293. Springer-Verlag, 2007.

[NP99] M. Naor and B. Pinkas. Oblivious transfer with adaptive queries. In *Advances in Cryptology – CRYPTO'99*, volume 1666 of *LNCS*, pages 791–791. Springer-Verlag, 1999.

[Nui10] K. Nuida. Short collusion-secure fingerprint codes against three pirates. In *Information Hiding, 12th International Workshop - IH 2010*, volume 6387 of *Lecture Notes in Computer Science*, pages 86–102. Springer-Verlag, 2010.

[OMS10] M.B. Ould Medeni and El Mamoun Souidi. A steganography schema and error-correcting codes. *Journal of Theoretical and Applied Information Technology*, 18(1) :42–47, 2010.

[OP98] J.J.K. O'Ruanaidh and T. Pun. Rotation, scale and translation invariant spread sspectrum digital image watermarking. *Signal Processing*, 66(3) :303–317, 1998.

[PB07] A. Piva and M. Barni. The first bows contest : Break our watermarking system. In *IS&T/SPIE International Symposium on Electronic Imaging 2007 - Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505 of *Proceedings of the SPIE*. SPIE, 2007.

[Pev11] T. Pevny. Detecting messages of unknown length. In *IS&T/SPIE International Symposium on Electronic Imaging 2011 - Media Watermarking, Security and Forensics III*, volume 7880 of *Proceedings of the SPIE*. SPIE, 2011.

[PF04] F. Petitcolas and C. Fontaine. *Tatouage de documents audiovisuels numériques*, chapter Nouveaux outils pour l'évaluation des algorithmes de tatouage. Hermès-Lavoisier, 2004.

[PF07] T. Pevný and J. Fridrich. Merging markov and dct features for multiclass jpeg steganalysis. In *IS&T/SPIE International Symposium on Electronic Imaging 2007 - Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505 of *Proceedings of the SPIE*. SPIE, 2007.

[PFB10] T. Pevný, T. Filler, and P. Bas. Using High-Dimensional Image Models to Perform Highly Undetectable Steganography. In *Information Hiding, 12th International Workshop - IH 2010*, volume 6387 of *Lecture Notes in Computer Science*. Springer-Verlag, 2010.

[PFCPG05] L. Pérez-Freire, P. Comesaña, and F. Pérez-González. Information-theoretic analysis of security in side-informed data hiding. In *Information Hiding, 7th International Workshop - IH 2005*, volume 3727 of *Lecture Notes in Computer Science*, pages 131–145. Springer-Verlag, 2005.

[PFCTPPG06] L. Pérez-Freire, P. Comesña, J.R. Troncoso-Pastoriza, and F. Pérez-González. Watermarking security : a survey. In *Transactions on Data Hiding and Multimedia Security I*, number 4300 in Lecture Notes in Computer Science, pages 41–72. Springer-Verlag, 2006.

[PFK09] T. Pevny, J. Fridrich, and A. Ker. From blind to quantitative steganalysis. In *IS&T/SPIE International Symposium on Electronic Imaging 2009 - Media Forensics and Security*, volume 7254 of *Proceedings of the SPIE*. SPIE, 2009.

[PFPG05] L. Pérez-Freire and F. Pérez-González. Spread-spectrum vs. quantization-based data hiding : misconceptions and implications. In *IS&T/SPIE International Symposium on Electronic Imaging 2005 - Security, Steganography, and Watermarking of Multimedia Contents VII*, volume 5681 of *Proceedings of the SPIE*, pages 341–352. SPIE, 2005.

[PFPG07] L. Pérez-Freire and F. Pérez-González. Exploiting security holes in lattice data hiding. In *Information Hiding, 9th International Workshop - IH 2007*, volume 4567 of *Lecture Notes in Computer Science*, pages 159–173. Springer-Verlag, 2007.

[PFPG08] L. Pérez-Freire and F. Pérez-González. Security of lattice-based data hiding against the watermarked-only attack. *IEEE Transactions on Information Forensics and Security*, 3(4) :593–610, 2008.

[PFPG09] L. Pérez-Freire and F. Pérez-González. Spread-spectrum watermarking security. *IEEE Transactions on Information Forensics and Security*, 4(1) :2–24, 2009.

[PFPGFC06] L. Pérez-Freire, F. Pérez-Gonzáles, T. Furon, and P. Comesaña. Security of lattice-based data hiding against the known message attack. *IEEE Transactions on Information Forensics and Security*, 1(4) :421–439, 2006.

[PHB98] V.S. Pless, W.C. Huffman, and R.A. Brualdi, editors. *Handbook of coding theory*. Elsevier, 1998.

[PL03] S. Pateux and G. Le Guelvouit. Practical watermarking scheme based on wide spread spectrum and game theory. *Signal Processing : Image Communication*, 18 :283–296, 2003.

[PL05] R.C.-W. Phan and H.-C. Ling. Flaws in generic watermarking protocols based on zero-knowledge proofs. In *Digital Watermarking, 2nd International Workshop - IWDW'04*, volume 3304 of *Lecture Notes in Computer Science*, pages 184–191. Springer-Verlag, 2005.

[PP99] S. Pereira and T. Pun. Fast robust Template Matching for Affine resistance image Watermarking. In *International Workshop on Information Hiding*, volume 1768, pages 200–210. Lecture Notes in Computer Science, 1999.

[PR01] J. Picard and A. Robert. Neural network functions for public key watermarking. In *Information Hiding, 4th International Workshop - IH 2001*, volume 2137 of *Lecture Notes in Computer Science*, pages 142–156. Springer-Verlag, 2001.

[Pro01] N. Provos. Defending against statistical steganalysis. In *Usenix security symposium*, 2001.

[PS96] B. Pfitzmann and M. Schunter. Asymmetric fingerprinting. In *Advances in Cryptology - EUROCRYPT'96*, volume 1070 of *Lecture Notes in Computer Science*, pages 84–95. Springer-Verlag, 1996.

[PSR+01] F.A. Petitcolas, M. Steinebach, F. Raynal, J. Dittmann, C. Fontaine, and N. Fates. Public automated web-based evaluation service for watermarking schemes : Stirkark benchmark. In *IS&T/SPIE International Symposium on Electronic Imaging 2001- Security and Watermarking of Multimedia Contents III*, volume 4314 of *Proceedings of the SPIE*, pages 575–584. SPIE, 2001.

[PSS03] C. Peikert, A. Shelat, and A. Smith. Lower bounds for collusion-secure fingerprinting. In *Proceedings of the fourteenth annual ACM-SIAM Symposium On Discrete Algorithms – SODA 2003*, pages 472–479. SIAM, 2003.

[PW97a] B. Pfitzmann and M. Waidner. Anonymous fingerprinting. In *Advances in Cryptology - EUROCRYPT'97*, volume 1233 of *Lecture Notes in Computer Science*, pages 88–102. Springer-Verlag, 1997.

[PW97b] B. Pfitzmann and M. Waidner. Asymmetric fingerprinting for larger collusions. In *ACM Conference on Computer and Communication Security – CCS'97*, pages 151–160. ACM, 1997.

[Rab81] M. Rabin. How to exchange secrets by oblivious transfer. Technical report, Technical Report TR-81, Harvard Aiken Computation Laboratory, 1981, 1981.

[RDB+10] A. Rial, M. Deng, T. Bianchi, A. Piva, and B. Preneel. A provably secure anonymous buyerâĂŞseller watermarking protocol. *IEEE Transactions on Information Forensics and Security*, 5(4) :920–931, 2010.

[RPF01] F. Raynal, F. Petitcolas, and C. Fontaine. Évaluation automatique des méthodes de tatouage. *Traitement du signal*, 18(4) :271–282, 2001. numéro spécial "tatouage et sécurité de l'information", http ://documents.irevues.inist.fr/handle/2042/2184.

[RPR09] H. Rifà-Pous and J. Rifà. Product perfect codes and steganography. *Digital Signal Processing*, 19(4) :764–769, 2009.

[RR10] J. Rifà and L. Ronquillo. Product perfect z2z4-linear codes in steganography. In *International Symposium on Information Theory and its Applications - ISITA 2010*, 2010.

[Sal05] Sallee. Model-based methods for steganography and steganalysis. *International Journal of Image Graphics*, 5(1) :167–190, 2005.

[SBM03] A. Somekh-Baruch and N. Merhav. On the error exponent and capacity games of private watermarking systems. *IEEE Transactions on Information Theory*, 49(3) :537–562, 2003.

[SBM04] A. Somekh-Baruch and N. Merhav. On the capacity game of public watermarking systems. *IEEE Transactions on Information Theory*, 50(3) :511–524, 2004.

[SBM05] A. Somekh-Baruch and N. Merhav. On The Capacity Game of Private FingerPrinting Systems Under Collusion Attacks. *IEEE Transactions on Information Theory*, 51(3) :884–899, 2005.

[SCC06] Y.Q. Shi, C. Chen, and W. Chen. A markov process based approach to effective attacking jpeg steganography. In *Information Hiding, 8th International Workshop - IH 2006*, volume 4437 of *Lecture Notes in Computer Science*, pages 249–264. Springer-Verlag, 2006.

[Sch04] H.G. Schaathun. Binary collusion-secure codes : comparison and improvements. Technical report, University of Bergen (Norway), Department of Informatics, 2004.

[Sch08a] H.G. Schaathun. Attack analysis for he & wu's joint watermarking/fingerprinting scheme. In *Digital Watermarking, 4th International Workshop - IWDW'07*, volume 5041 of *Lecture Notes in Computer Science*, pages 45–59. Springer-Verlag, 2008.

[Sch08b] H.G. Schaathun. On error-correcting fingerprinting codes for use with watermarking. *Multimedia Systems*, 13 :331–344, 2008.

[SD99] J. Smith and C. Dodge. Developments in steganography. In *Information Hiding 1999*, volume 1768 of *Lecture Notes in Computer Science*, pages 77–87. Springer-Verlag, 1999.

[SFC05] M. Soriano, M. Fernandez, and J. Cotrina. Fingerprinting schemes. Identifying the guilty sources using side information. In *Digital Watermarking, 4th International Workshop - IWDW'05*, volume 3710 of *Lecture Notes in Computer Science*, pages 231–243. Springer-Verlag, 2005.

[Sha48] C.E. Shannon. A mathematical theory of communications. *Bell System Technical Journal*, 27 :379–423, 623–656, 1948.

[Sha49] C.E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28 :656–715, 1949.

[Sha58] C.E. Shannon. Channels with side information at the transmitter. *IBM journal of Research and Development*, 2(4) :289–293, 1958.

[SHHD01] G. Silvestre, N. Hurley, G. Hanau, and W. Dowling. Informed audio watermarking using digtial chaotic signals. In *IEEE ICASSP'01*, volume 3, pages 1361–1364. IEEE, 2001.

[Sie85] T. Siegenthaler. Decrypting a class of stream ciphers using ciphertext only. *IEEE Transactions on Computers*, 34(1) :81–84, 1985.

[Sim84] G.J. Simmons. The prisoners' problem and the subliminal channel. In *Advances in Cryptology – CRYPTO'83*, pages 51–67. Plenum Press, 1984.

[ŠKC08] B. Škorić, S. Katzenbeisser, and M.U. Celik. Symmetric Tardos fingerprinting codes for arbitrary alphabet sizes. *Designs, Codes and Cryptography*, 46(2) :137–166, 2008.

[ŠKSC09] B. Škorić, S. Katzenbeisser, H.G. Schaathun, and M. Celik. Tardos Fingerprinting codes in the combined digit model. In *IEEE International Workshop on Information Forensics and Security - WIFS 2009*, pages 41–45. IEEE, 2009.

[ŠKSC11] B. Škorić, S. Katzenbeisser, H.G. Schaathun, and M. Celik. Tardos Fingerprinting codes in the combined digit model. *IEEE Transactions on Information Forensics and Security*, 6(3) :906, 2011.

[SKZ09] V. Sachnev, H.J. Kim, and R. Zhang. Less detectable jpeg steganography method based on heuristic optimization and BCH syndrom coding. In *ACM Multimedia & Security'09*, pages 131–139. ACM Press, 2009.

[SNW02] R. Safavi-Naini and Y. Wang. Collusion-secure q-ary fingerprinting for perceptual content. In *Security and Privacy in Digital Rights Management - SPDRM'01*, volume 2320 of *Lecture Notes in Computer Science*, pages 57–75. Springer-Verlag, 2002.

[SPR+01] M. Steinebach, F. Petitcolas, F. Raynal, J. Dittmann, C. Fontaine, C. Seibel, and N. Fatès. Stirmark benchmark : audio watermarking attacks. In *International Conference on Information Technology : Coding and Computing, ITCC 2001*. IEEE computer society press, 2001. Special Session in Multimedia Security and Watermarking Applications.

[SŠ11] A. Simone and B. Škorić. Asymptotically false-positive-maximizing attack on non-binary tardos codes. In *13th Information Hiding – IH'11*, Lecture Notes in Computer Science. Springer-Verlag, 2011. to appear.

[SSM+06] K. Solanki, K. Sullivan, U. Madhow, B. Manjunath, and S. Chandrasekaran. Provably secure steganography : Achieving zero k-l divergence using statistical restoration. In *IEEE International Conference on Image Processing – ICIP 2006*, pages 125–128. IEEE, 2006.

[SSW01] J.R. Staddon, D.R. Stinson, and R. Wei. Combinatorial properties of frameproof and traceability codes. *IEEE Transactions on Information Theory*, 47(3) :1042–1049, 2001.

[SSW03] A. Silverberg, J.R. Staddon, and J. Walker. Application of list decoding to tracing traitors. *IEEE Transactions on Information Theory*, 49(5) :1312–1318, 2003.

[ST01] J. Stern and J.-P. Tillich. Automatic detection of a watermarked document using a private key. In *Information Hiding, 4th International Workshop - IH 2001*, volume 2137 of *Lecture Notes in Computer Science*, pages 258–272. Springer-Verlag, 2001.

[ŠVCT08] B. Škorić, T.U. Vladimirova, M.U. Celik, and J. Talstra. Tardos Fingerprinting is Better Than We Thought. *IEEE Transactions on Information Theory*, 54(8) :3663–3676, 2008.

[SvTW00] D.R. Stinson, T. van Trung, and R. Wei. Secure frameproof codes, key distribution patterns, group testing algorithms and related structures. *Journal of Statistical Planning and Inference*, 86 :595–617, 2000.

[SW98] D.R. Stinson and R. Wei. Combinatorial properties and construction of traceability schemes and frameproof codes. *SIAM Journal on Discrete Mathematics*, 11(1) :41–53, 1998.

[SW06] D. Schönfeld and A. Winkler. Embedding with syndrome coding based on BCH codes. In *Proc. of the ACM Multimedia and Security Workshop 2006*, pages 214–223. ACM, 2006.

[SW07] D. Schönfeld and A. Winkler. Reducing the Complexity of Syndrome Coding for Embedding. In *Information Hiding, 10th International Workshop - IH 2007*, volume 4567 of *Lecture Notes in Computer Science*, pages 145–158. Springer-Verlag, 2007.

[Tar03] G. Tardos. Optimal probabilistic fingerprint codes. In *ACM STOC'03*, pages 116–125. ACM, 2003.

[Tar08] G. Tardos. Optimal probabilistic fingerprint codes, journal of the acm 55 (2008) (2), art. 10, 24pp. *Journal of the ACM*, 55(2), 2008. Article Nb. 10.

[TBHK03] R. Tzschoppe, R. Bäuml, J.B. Huber, and A. Kaup. Steganographic system based on higher-order statistics. In *IS&T/SPIE International Symposium on Eletronic Imaging 2003 - Security and Watermarking of Multimedia Contents V*, volume 5020 of *Proceedings of the SPIE*, pages 147–154. SPIE, 2003.

[THC05] J. Tzeng, W.-L. Hwang, and I.-L. Chern. An asymetric subspace watermarking method for copyright protection. *IEEE Transactions on Signal Processing*, 53(2) :784–792, 2005. numéro spécial "Supplement on Secure Media II".

[Tie73] A. Tietäväinen. On the nonexistence of perfect codes over finite fields. *SIAM Journal on Applied Mathematics*, 24 :88–96, 1973.

[TPPG06] J.R. Troncoso-Pastoriza and F. Pérez-González. Zero-knowledge watermark detector robust to sensitivity attacks. In *Proc. of the ACM Multimedia and Security Workshop 2006*, pages 97–107. ACM, 2006.

[TWWL03] W. Trappe, M. Wu, Z.J. Wang, and K.J.R. Liu. Anti-collusion fingerprinting for multimedia. *IEEE Transactions on Signal Processing*, 51(4) :1069–1087, 2003. special issue on signal processing for data hiding in digital media and secure content delivery.

[vAH04] L. von Ahn and N. Hopper. Public-key steganography. In *Advances in Cryptology - EUROCRYPT'2004*, volume 3027 of *Lecture Notes in Computer Science*, pages 323–341, 2004.

[Var97] A. Vardy. The intractability of computing the minimum distance of a code. *IEEE Transactions on Information Theory*, 43(6) :1757–1766, 1997.

[Vau06] S. Vaudenay. *A classical introduction to Cryptography : Applications for Communications Security.* Springer-Verlag, 2006.

[VJ05] R. Venkatesan and M. Jakubowski. Randomized detection for spread-spectrum watermarking : Defending against sensitivity and other attacks. In *IEEE ICASSP'05*, 2005.

[VSTS99] R. Van Schyndel, A. Tirkel, and I. Svalbe. Key independent watermark detection. In *IEEE International Conference on Multimedia Computing and Systems*, 1999.

[vT05] H. van Tilborg, editor. *Encyclopedia of Cryptography and Security.* Springer-Verlag, 2005.

[Wag83] N.R. Wagner. Fingerprinting. In *IEEE Symposium on Security and Privacy*, pages 18–22. IEEE, 1983.

[WB02] H. Wu and F. Bao. Cryptanalysis of stream cipher COS (2, 128) mode I. In *Australian Conference on Information Security and Privacy, ACISP 2002*, number 2384 in Lecture Notes in Computer Science, pages 154–158. Springer-Verlag, 2002.

[WBSS04] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment : From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4) :600–612, 2004.

[Wes01] A. Westfeld. F5 – a steganographic algorithm : high capacity despite better steganalysis. In *Information Hiding, 4th International Workshop - IH 2001*, volume 2137 of *Lecture Notes in Computer Science*, pages 289–302. Springer-Verlag, 2001.

[Wes06] A. Westfeld. Lessons from the bows contest. In *Proc. of the ACM Multimedia and Security Workshop 2006*, pages 208–213. ACM, 2006.

[Wes08] A. Westfeld. A Regression-Based Restoration Technique for Automated Watermark Removal. In *ACM Multimedia & Security'08*. ACM, 2008.

[Wes09] A. Westfeld. Fast determination of sensitivity in the presence of countermeasures in bows-2. In *Information Hiding, 11th International Workshop - IH 2009*, volume 5806 of *Lecture Notes in Computer Science*, pages 89–101. Springer-Verlag, 2009.

[WM04] J. Wang and P. Moulin. Steganalysis of block-structured stegotext. In *IS&T/SPIE International Symposium on Electronic Imaging 2005 - Security, Steganography, and Watermarking of Multimedia Contents VI*, volume 5681 of *Proceedings of the SPIE*, pages 477–488. SPIE, 2004.

[WM07] Y. Wang and P. Moulin. Optimized feature extraction for learning-based image steganalysis. *IEEE Transactions on Information Forensics and Security*, 2(1) :31–45, 2007.

[WM08] Y. Wang and P. Moulin. Perfectly secure steganography : capacity, error exponents, and code constructions. *IEEE Transactions on Information Theory*, 54(6) :2706–2722, 2008.

[WP99] A. Westfeld and A. Pfitzmann. Attacks on steganographic systems. In *Information Hiding, 3rd International Workshop - IH 1999*, volume 1768 of *Lecture Notes in Computer Science*, pages 61–76. Springer-Verlag, 1999.

[WvD05] F.M.J. Willems and M. van Dijk. Capacity and codes for embedding information in gray-scale signals. *IEEE Transactions on Information Theory*, 51(3) :1209–1214, 2005.

[WWZ+03] Z.J. Wang, M. Wu, H. Zhao, W. Trappe, and K.J.R. Liu. zha. In *IEEE ICASSP'03*, pages 724–727. IEEE, 2003.

[WWZ+05] Z.J. Wang, M. Wu, H. Zhao, W. Trappe, and K.J.R. Liu. Anti-collusion forensics of multimedia fingerprinting using orthogonal modulation. *IEEE Transactions on Image Processing*, 14(6) :804–821, 2005.

[WZ76] A. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22(1) :1–10, 1976.

[WZW03] Q.H. Wu, J.H. Zhang, and Y.M. Wang. Practical t-out-n oblivious transfer and its applications. In *Information and Communications Security*, volume 2936 of *LNCS*, pages 226–237. Springer-Verlag, 2003.

[XFF08] F. Xie, T. Furon, and C. Fontaine. On-Off keying modulation and Tardos fingerprinting. In *ACM Multimedia & Security'08*. ACM, 2008.

[XFF09] F. Xie, C. Fontaine, and T. Furon. Un schéma complet de traçage de documents multimédia reposant sur des versions amÃľliorées des codes de tardos et de la technique de tatouage Broken Arrows. In *XIIe colloque GRETSI*, 2009.

[XFF10a] F. Xie, T. Furon, and C. Fontaine. Better security levels for 'Broken Arrows'. In *IS&T/SPIE International Symposium on Electronic Imaging 2010 - Media Forensics and Security XII*, volume 7541 of *Proceedings of the SPIE*. SPIE, 2010.

[XFF10b] F. Xie, T. Furon, and C. Fontaine. Towards Robust and Secure Watermarking. In *ACM Multimedia & Security'10*, pages 153–159. ACM, 2010.

[Xie10] F. Xie. *Tatouage sûr et robuste appliqué au traçage de documents multimedia*. PhD thesis, Université de Rennes 1, 2010.

[XSG+05] G. Xuan, Y.Q. Shi, J. Gao, D. Zou, C. Yang, Z.Z.P. Chai, C. Chen, and W. Chen. Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions. In *Information Hiding, 7th International Workshop - IH 2005*, volume 3727 of *Lecture Notes in Computer Science*, pages 262–277. Springer-Verlag, 2005.

[Yac01] Y. Yacobi. Improved boneh-shaw content fingerprinting. In *Topics in Cryptology - CT-RSA 2001, The Cryptographer's Track at RSA Conference 2001*, volume 2020 of *Lecture Notes in Computer Science*, pages 378–391. Springer-Verlag, 2001.

[YJXPLML05] H. Yan-Jun, M. Xiao-Ping, D. Lin-Ming, and G. Li. The effect of fidelity measure functions on the capacity of digital watermarks. In *DRMTICS 2005*, pages 113–122, 2005.

[YJXPLML08] H. Yan-Jun, M. Xiao-Ping, D. Lin-Ming, and G. Li. A computation model for capacity and robustness of robust image watermarking scheme in spatial domain. In *International Conference on Intelligent Information Hiding and Multimedia Signal Processing – IIHMSP'08*, pages 1154–1157. IEEE, 2008.

[YL05] C.-M. Yu and C.-S. Lu. Robust non-interactive zero-knowledge watermarking scheme against cheating prover. In *Proc. of the ACM Multimedia and Security Workshop 2005*, pages 103–110. ACM, 2005.

[ZCR+11] C. Zitzmann, R. Cogranne, F. Retraint, I. Nikiforov, L. Fillatre, and P. Cornu. Statistical decision methods in hidden information detection. In *13th Information Hiding – IH'11*, Lecture Notes in Computer Science. Springer-Verlag, 2011. to appear.

[ZFK+98] J. Zöllner, H. Federrath, H. Klimant, A. Pfitzmann, R. Piotraschke, A. Westfeld, G. Wicke, and G. Wolf. Modeling the security of stega-

nographic systems. In *Information Hiding, 2nd International Workshop - IH 1998*, volume 1525 of *Lecture Notes in Computer Science*, pages 344–354. Springer-Verlag, 1998.

[ZFZ05] Y. Zhu, D. Feng, and W. Zou. Collusion secure convolutional spread spectrum fingerprinting. In *Digital Watermarking, 4th International Workshop - IWDW'05*, volume 3710 of *Lecture Notes in Computer Science*, pages 67–83. Springer-Verlag, 2005.

[ZK95] J. Zhao and E. Koch. Embedding Robust labels into images for copyright protection. In *International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies - KnowRight'95*, Schriftenreihe der ï£¡sterreichischen Computer Gesellschaft, Band 82, pages 242–251. Oldenbourg, 1995.

[ZL73] V.A. Zinov'ev and V.K. Leont'ev. The nonexistence of perfect codes over galois fields. *Problems of Control and Information Theory*, 2(2) :123–132, 1973.

[ZL08] W. Zhang and S. Li. A coding problem in steganography. *Designs, Codes and Cryptography*, 46(1) :67–81, 2008.

[ZLWY10] W. Zhang, J. Liu, X. Wang, and N. Yu. Generalization and analysis of the paper folding method for steganography. *IEEE Transactions on Information Forensics and Security*, 5(4) :694–704, 2010.

[ZP03] T. Zhang and X. Ping. A fast and effective steganalytic technique against jsteg-like algorithms. In *ACM Symposium on Applied Computing*, pages 307–311, 2003.

[ZSK09] R. Zhang, V. Sachnev, and H.J. Kim. Fast BCH syndrome coding for steganography. In *Information Hiding*, volume 5806 of *Lecture Notes in Computer Science*, pages 48–58. Springer-Verlag, 2009.

[ZW06] X. Zhang and S. Wang. Efficient steganographic embedding by exploiting modification direction. *IEEE Communications Letters*, 10(11) :781–783, 2006.

[ZW09] W. Zhang and X. Wang. Generalization of the zzw embedding construction for steganography. *IEEE Transactions on Information Forensics and Security*, 4(3) :564–569, 2009.

[ZWWL03] H. Zhao, M. Wu, J. Wang, and K.J.R. Liu. Nonlinear collusion attacks on independent fingerprints for multimedia. In *IEEE ICASSP'03*, 2003. poster.

[ZWWL05] H.V. Zhao, M. Wu, Z.J. Wang, and K.J.R. Liu. Forensic analysis of nonlinear collusion attacks for multimedia fingerprinting. *IEEE Transactions on Image Processing*, 14(5) :646–661, 2005.

[ZZ04] F. Zhang and H. Zhang. Digital watermarking capacity research. In *International Conference on Communication, Circuits and Systems*, volume 2, pages 796–799, 2004.

[ZZ09] W. Zhang and X. Zhu. Improving the embedding efficiency of wet paper codes by paper folding. *IEEE Signal Processing Letters*, 16(9) :794–797, 2009.

[ZZW07] W. Zhang, X. Zhang, and S. Wang. A double layered "plus-minus one" data embedding scheme. *IEEE Signal Processing Letters*, 14(11) :848–851, 2007.

[ZZW08] W. Zhang, X. Zhang, and Shuozhong Wang. Maximizing Steganographic Embedding Efficiency by Combining Hamming Codes and Wet Paper Codes. In *Information Hiding, 10th International Workshop - IH 2008*, volume 5284 of *Lecture Notes in Computer Science*. Springer-Verlag, 2008.

[ZZW10] W. Zhang, X. Zhang, and S. Wang. Near-optimal codes for information embedding in gray-scale signals. *IEEE Transactions on Information Theory*, 56(3) :1262–1270, 2010.

Table des figures

1 Modélisation de la stéganographie, du tatouage et du *fingerprinting* comme des problématiques de transmission. Le vocabulaire propre à la stéganographie est mis entre parenthèses. La flèche en pointillés indique que dans certains cas particuliers de tatouage ou *fingerprinting* le document original peut être disponible lors de la détection/extraction. xv

2 Nombre annuel de publications IEEE portant sur le tatouage et la stéganographie [CMB+08]. xvi

3 Thématiques abordées. Les noms entre crochets permettent de faire le lien avec la liste des projets donnée dans la section C.1 de l'annexe C. . xvii

1.1 Stéganographie : principe général et notations. 3

1.2 Principe du codage par syndrome avec un code linéaire $\mathcal{C}$ de longueur n et dimension k. Les vecteurs $\mathbf{x}$ et $\mathbf{y}$ sont de longueur n, le message $\mathbf{m}$ est de longueur $m = n - k$. On autorise au plus T modifications du vecteur $\mathbf{x}$ pour réaliser l'insertion. 13

1.3 Insertion à deux niveaux de Zhang, Zhang et Wang (ZZW) 22

1.4 Quelques repères concernant l'efficacité du décodage en liste par rapport à l'interpolation de Lagrange. 27

1.5 Évolution du nombre $n - k - r$ de symboles de message que l'on peut insérer, en fonction du nombre de composantes verrouillées $\ell_{\mathcal{W}}$. 29

2.1 Ces deux schémas modélisent les opérations d'insertion et de détection/extraction. Ils posent par ailleurs les notations utilisées dans ce manuscrit. Les lignes en pointillés montrent les entrées optionnelles. Pour l'insertion, la ligne pointillée correspond à l'utilisation d'une information adjacente. Pour la détection/extraction, elle correspond à la notion de détection/extraction aveugle ou informée. En retrouve en vert les informations connues de l'attaquant, et en orange les informations tenues secrètes. Les données qui peuvent être connues de l'attaquant ou non selon les contextes sont laissées en noir. 35

2.2 Principe d'une insertion substitutive. 36

2.3 Principe d'une insertion additive avec étalement de spectre. 37

2.4 Le tatouage perçu comme une transmission avec information adjacente à l'émission. Deux bruits interviennent successivement pour perturber le signal transmis (le message) : tout d'abord le document original $\mathbf{x}$ (qui va forcer à modérer l'insertion par souci d'imperceptibilité), puis le bruit du canal de transmission $\mathbf{z}$ correspondant aux transformations et attaques que le document tatoué va subir. 37

2.5 Évolution de l'*equivocation*, qui mesure l'incertitude sur la clé, en fonction du nombre d'observations, pour des schémas de tatouages substitutifs. La courbe $H(K)$ correspond au cas WOA, la courbe $H(K|M)$ au cas KMA, et la courbe $H(K|X)$ au cas KOA. 44

2.6 Efficacité de l'estimation des porteuses pour les techniques d'insertion par étalement de spectre : corrélation entre les porteuses estimées et les vraies, en faisant abstraction de l'incertitude sur l'ordre des porteuses (on les met dans le bon ordre pour calculer les corrélations). On remarque que dans le cas WOA il ne sert à rien de procéder à un grand nombre d'itérations pour le calcul des porteuses. Avec 3 itérations et 500 images, on estime correctement 40% des porteuses. La figure 2.7 montre à quel point cette estimation est suffisante pour retirer la marque tout en conservant une excellente qualité à l'image. 45

2.7 Expérimentations effectuées sur des images de taille 512×512, tatouées avec une technique de tatouage par étalement de spectre avec information adjacente, réputée pour sa robustesse [PL03]. Les paramètres sont $n = 258058$ et $N = 8$. Comparaison de l'efficacité de deux stratégies d'attaque : (a) l'attaquant procède traditionnellement, à l'aveuglette ; (b) il estime la clé par notre méthode, en observant environ 1000 images tatouées avec la même clé. L'image « Lena » tatouée présentait un PSNR de 38 dB. (c) Nous avons mené ces tests sur 50 images de taille 512×512, et le PSNR de l'image obtenue avec l'estimation de la clé est en moyenne de 15 dB supérieure à celle de l'image obtenue avec une attaque à l'aveuglette. 46

2.8 Résistance à la première attaque de Westfeld [Wes08] : comparaison du `Broken Arrows` original (`BA`) et denotre variante `BA-AWC`. On trace le pourcentage d'images attaquées avec succès en fonction du PSNR des images attaquées. 50

2.9 Distances chordales normalisées SCD_{norm} comparées pour le `Broken Arrows` original (`BA`) avec différentes valeurs des paramètres N_v et N_c, et notre variante `BA-AWC+`. Rappelons que les paramètres originaux de `Broken Arrows` sont $N_v = 256$ et $N_c = 30$. 51

3.1 Principe d'une attaque par *collusion*. 56

3.2 Les modèles issus des travaux de Boneh et Shaw [BS95, BS98, GP00]. . . 58

3.3 Principe du découpage en blocs et aiguillage. Chaque document est découpé en une série de blocs. Dans chacun de ces blocs est inséré un symbole de l'identifiant. On pré-tatoue autant de copies maîtres qu'il y a de symboles dans l'alphabet, chaque copie maître contenant le même symbole dans tous ses blocs. Les copies personnalisées sont ensuite composées à la volée lors de la transaction (vidéo à la demande), ou du rendu utilisateur (lecture d'un disque blu-ray, ou décodeur TV par exemple). . . . 59

3.4 Comparaison de quelques schémas, proposée dans [HW07] : n est le nombre d'utilisateurs, ici pris égal à 10^7, N_{ech} le nombre d'échantillons de signal hôte, la probabilité d'accuser à tort de 10^{-3}. 62

3.5 Attaques de type « échange de blocs », « copier-coller-fg. On peut prédire que les symboles extraits de la copie pirate sont issus directement des identifiants des membres de la coalition. En jaune : la *marking assumption* (a) est respectée. On est dans le modèle (a)+(b). 65

3.6 Attaques de type « fusion de blocs ». On ne peut pas prédire quels symboles seront extraits de la copie pirate, cela dépendra de la technique de tatouage utilisée. En jaune : la *marking assumption* (a) est cependant toujours respectée. On est dans le modèle (a)+(b′), qui autorise des erreurs et des effacements. 65

3.7 Attaques de type « lessivage ». On ne peut pas prédire quels symboles seront extraits de la copie pirate, cela dépendra de la technique de tatouage utilisée. On est donc dans le modèle (b′). 66

3.8 Résumé du fonctionnement des codes de Tardos binaires tels que présentés dans [ŠKC08]. 71

3.9 Distribution des scores des utilisateurs (coupables en rouge, innocents en vert) pour un code de Tardos binaire tel que présenté dans [ŠKC08] avec les paramètres $m = 2\pi^2 c^2 \lceil \ln(1/\varepsilon_1) \rceil$ et $Z = 2\pi c \lceil \ln(1/\varepsilon_1) \rceil$ [ŠVCT08]. Les expériences sont menées avec $\varepsilon_1 = 10^{-3}$, et des attaques de type « Echange de blocs ». Les scores sont distribués suivant des gaussiennes. 73

3.10 Distance de Kullback-Leibler entre les distributions des scores des innocents et des coupables, en fonction du paramètre de forme de la distribution de Dirichlet κ pour : l'échange de blocs avec le calcul de scores classique de Škorić *et al.* [ŠKC08], la fusion par moyennage avec le calcul de scores de l'équation (3.1), et la fusion par moyennage avec le calcul de scores de l'équation (3.2). Les paramètres du système sont $m = 300$, $q = 4$ et $c = 20$. 78

3.11 Plate-forme `FANTOMAS` : attaque par moyennage (de type fusion). On voit en haut les symboles détectés pour chaque bloc (on détecte parfois plusieurs symboles simultanément). On affiche les scores triés par ordre décroissant. Les cinq coupables arrivent bien en tête, avec des scores nettement supérieurs à ceux des innocents. On trouve, dans la fenêtre en bas à gauche les probabilités précises d'accuser chaque utilisateur à tort [CFG08]. On voit que le risque que l'on prend en accusant le plus grand score est ici de l'ordre de 10^{-25}. Pour le deuxième plus grand score, elle est de 10^{-16}, pour le troisième de 10^{-15}, pour le quatrième de 10^{-14}, et pour le cinquième de 10^{-8}. L'innocent ayant le plus grand score serait, lui, accusé à tort avec une probabililté de 10^{-3}. 80

3.12 Evaluation des performances de nos deux fonctions de calcul de scores (3.1) et (3.2), avec les différentes variantes de `Broken Arrows` présentées à la section 2.4. Nous avons choisi pour cette comparaison de prendre $\kappa = 0.23$, car comme le montre la figure 3.10 c'est pour cette valeur que les scores classiques de [ŠKC08] sont les plus performants. Les paramètres du système sont les mêmes que précédemment, soient $m = 300$, $q = 4$ et $c = 20$. On retrouve dans la première colonne les résultats de la figure 3.10, s'appuyant sur le masque originel de `Broken Arrows`, et les paramètres $N_v = 256$ et $N_c = 30$, avec une attaque par échange de blocs pour le calcul classique des scores [ŠKC08], et une fusion par moyennage pour les calculs des scores avec les équations (3.1) et (3.2). Pour les trois variantes de `Broken Arrows`, nous avons testé quatre attaques non-linéaires de type fusion : moyennage des pixels, entrelacement des pixels, maximum des pixels, et enfin l'attaque MMX (*Moderated Minority eXtreme*) due à Schaathun [Sch08a]. 81

3.13 Estimation de la taille de la coalition ainsi que de la stratégie qu'elle a adoptée pour produire la copie pirate, et optimisation dynamique des fonctions d'accusation. 83

3.14 Fonctions d'accusation qui maximisent l'espérance du score des coupables. 84

3.15 Probabilité d'accuser correctement le k-ième plus grand score, pour $k \in \{1, \ldots, 8\}$. $m = 1000$, $c = 8$, $n = 5000$. 400 expériences réalisées. On compare l'efficacité d'une accusation classique (a) ici mise en défaut par une longueur trop faible par rapport à celle préconisée, avec une accusation optimisée (b)(c)(d). 85

3.16 Impact de la modification du vecteur $\mathbf{p}$ sur le calcul des scores. Expériences menées pour $m = 1000$ avec $c = 3$. L'axe des abscisses indique le nombre de composantes modifiées dans le vecteur $\mathbf{p}$. Les lignes colorées en plein montrent comment les scores de dix utilisateurs pris au hasard augmentent. Les lignes en pointillés montrent les scores des coupables qui seraient calculés avec le bon vecteur $\mathbf{p}$. 87

Liste des tableaux

3.1 Valeurs de $cm\tilde{\mu}_C/\tilde{\sigma}_I$ obtenues après optimisation des fonctions d'accusation pour $m = 100$, $c = 3, 4, 5$. Entre parenthèses sont données celles obtenues par [FGC08]. Rappelons qu'avec les fonctions de [ŠKC08] on a 64 dans tous les cas. 84

Table des matières

Remerciements **iii**

Introduction générale **xi**

Overview **xix**

1 Assurer la furtivité des communications grâce à la stéganographie **1**
1.1 Introduction à la stéganographie 1
1.2 Une meilleure gestion des composantes verrouillées grâce aux codes de Reed-Solomon : séjour post-doctoral de F. Galand (2006-2007) . . . 25
1.3 Assurer l'insertion, coûte que coûte : thèse de M. Barbier (2008-2011) 28
1.4 Conclusion et perspectives 30

2 Protéger le droit d'auteur grâce à un tatouage robuste et sûr **31**
2.1 Introduction au tatouage robuste 32
2.2 Cryptanalyse des schémas de tatouage (2002-2005) 42
2.3 Utilisation de tatouage audio robuste pour sécuriser la diffusion de musique sur les téléphones mobiles 3G (2003-2006) 45
2.4 Amélioration de la robustesse et de la sécurité de `Broken Arrows` : thèse de F. Xie (2007-2010) 47
2.5 Conclusion et perspectives 52

3 Identifier la provenance de fraudes grâce à la personnalisation de copies (fingerprinting) **55**
3.1 Introduction à la personnalisation de copies 56
3.2 Association (et amélioration) des codes de Tardos avec le schéma de tatouage `Broken Arrows` pour contrer les attaques de type fusion : thèse de F. Xie (2007-2010) 76
3.3 Optimisation dynamique de l'accusation des codes de Tardos : thèse de A. Charpentier (2008-2011) 79
3.4 Intégration des codes de Tardos dans un protocole de traçage dit asymétrique, thèse de A. Charpentier (2008-2011) 86
3.5 Conclusion et perspectives 89

4 Conclusion et perspectives générales **91**

A Autres contributions **93**
A.1 Tatouage 93
A.2 Conception et attaque de systèmes de chiffrement 95

B Soutenance **101**

C Curriculum Vitæ, et liste de publications **121**
C.1 Curriculum Vitæ 121
C.2 Liste de publications 130

D Publications choisies **137**
[FG09] How Reed-Solomon Codes Can Improve Steganographic Schemes, *EURASIP Journal on Information Security*, 2009 138
[ABF11] Ensuring message embedding in wet paper steganography, *IMA 13th Conference on Cryptography and Coding*, 2011 148
[CFF05e] Watermarking Security : Theory and Practice, *IEEE Transactions on Signal Processing*, 2005 166
[CFFC11] An Asymmetric Fingerprinting Scheme based on Tardos Codes, *Information Hiding*, 2011 178
[CCCF01] On Cryptographic Properties of the Cosets of $R(1, m)$, *IEEE Transactions on Information Theory*, 2001 194

Bibliographie **215**

Table des figures **250**

Liste des tableaux **254**

Table des matières **255**