Tatouage d'image semi-fragile pour appareil mobile intégré dans une chaîne de certification

Yves Stadler

Université de Lorraine

École doctorale IAEM

Informatique, Automatique, Électronique–Électrotechnique et Mathématiques

T H È S E

pour obtenir le titre de

Docteur en Informatique

de l'Université de Lorraine

Présentée et soutenue par

Yves Stadler

Tatouage d'image semi-fragile pour appareil mobile intégré dans une chaîne de certification

préparée à CodaSystem et Université de Lorraine, équipe LITA

soutenue le 29 novembre 2012

Jury :

Rapporteurs :	Lionel Fillatre	-	Université de Nice
	William Puech	-	Université de Montpellier 2
Examinateurs :	Imed Kacem	-	Université de Lorraine
	Djamel Khadraoui	-	CRP Henri Tudor
	Michaël Krajecki	-	Université de Reims Champagne-Ardennes
	Jean-Marie Moureaux	-	Université de Lorraine (CRAN)
Directeurs de thèse :	Yann Lanuel	-	Université de Lorraine
	Anass Nagih	-	Université de Lorraine

Abstract

In order to come to a judgement, the judge needs to assess the evidences submitted by the different parties. The trustiness of her decision is based on the trust she can have in the different pieces of information. But, with the paperless office, this notion is loosening. Despite the numeric-analog equivalents, keeping documents properties is not immediate. It is necessary to set new means, which guaranty authentication and integrity of a document.

This thesis has been made with an industrial-based funding (french CIFRE funding), in partnership with the firm CodaSystem. Its activity consist in developing pieces of software for mobiles to capture images and assure its probative value. Ensuring this involves the use of watermarking techniques to embed relevant pieces of information.

The "certification chain" is made of successive operations, which protect the image and its related information. To guaranty the quality of the security management, the "certification chain" has to be certified, following the ISO2700 norm. This certification is key element for a firm which will act act as a trusted third party. Within the certification process, risk analysis is the first step. This analysis describes the vulnerabilities, impacts and threats relevant to the information. It has pointed out critical risks in the different phases of the chain. Most important was the use of the image watermarking. Risks involved includes mark removal, chosen mark insertion, or mark duplication. Despite a good watermarking, the risk analysis revealed the necessity of assuring authenticity and fiability of the pieces of information that are watermarked. This includes the user authentication, timestamping authentication and geolocation authentication. This thesis subject is the study and conception of a smartphone image-watermarking application to embed pieces of information meeting the security needs.

Chapter 1 studies the probative image problematic. It also provides the "certification chain" risk analysis.

In the 2nd chapter, image watermarking properties are studied along with a classification of possible attacks on watermarking. This classification is use to identify the relevant attack scenarios, leading to the definition of the needed properties required for a "certification watermarking": imperceptibility, capacity, performance, semi-fragility and localisation. The proposed algorithm is a derivative of the F5 algorithm, tuned to meet the requirements. The image is watermarked blockwise, after the quantisation step and before the compression step in the JPEG scheme. To ensure a better invisibility, the encoding matrix method has been used and modifies only the *Least Significant Bit* - Bit de poids faible (LSB) of the *Discrete Cosine Transform* - Transformée en Cosinus Discrète (DCT) coefficients. The embedded message can be part of the metadata (pieces of information related to the evidence) or an integrity fingerprint. A solid cryptographic protocol has been developed to interleaves these marks to avoid a attacker being able to extract these marks without having to deal with

a computation problem. All blocks are not suitable for watermarking. If a block does not hold enough information, any modification can be used to get insight of the mark. Setting the threshold for a block watermarking is a non negligible problem. A set of tests validate the algorithm implementation and can be used to define experimentally the appropriate parameters.

A linked aspect of the watermarking is the gathering of the pieces of information which are embedded, more specifically the geolocation. These propositions have been mostly done for the ATLAS european project. The approach has been limited to the software part of the smartphones (exluding all intervention on the *Global Positioning System* (GPS) infrastructure, or embedded electronic). Four type of controls are proposed: *a posteriori* controls, *a priori* controls, corrective controls and external controls. An *a priori* control is realised by a quality of service. Pieces of information are gathered and compiled by a trusted third party to retro-actively determine the quality of a positionning solution computation. *A priori* control is based on the use of data gathered from terrestrial infrastructures. They are used to compute a quality indicator, which can be display on the mobile phone as an icon to inform the use in real-time. Moreover, this service can be used to correct the gelocation of a user. External controls are composed of two controls, providing *a priori* information. First control is solely based on the smartphone capabilities. It monitors the variation in GPS signal (strength, speed and thrust). If the phone is connected to other sources of geolocation information (*Global System for Mobile Communications* (GSM), *Wireless Fidelity* (Wi-Fi)), then it can use correlations between the signals to ensure the correctness of the calculations. The second control is based on the use of an external device, developped by our partner in the project. This dongle receives raw signal from the satellites, which can be transmitted to a server for treatment. The computed location is then send back to the smartphone and can be used for watermarking purpose. These propositions have in common the need for a trusted third party. This is a traditionnal approach in information security when authenticity is required. In the same way, authenticity for timestamping and user identification is need.

Remerciements

La réalisation d'une thèse est certes une réussite personnelle, mais elle est rarement possible sans l'implication de beaucoup de personnes. Cette section est dédiée à ces personnes, qui ont permis, par leur travail et leur soutien, l'aboutissement de ce travail.

J'aimerai tout d'abord remercier Francine Herrmann qui m'a mis en contact avec la société CodaSystem et a permis le démarrage de cette thèse CIFRE. Merci aussi pour sa participation dans les projets européens et son aide à la relecture des différents papiers publiés et le manuscrit de thèse.

Je tiens à remercier aussi, Anass Nagih et Yann Lanuel, qui ont accepté d'être mes co-directeurs de thèse. Merci pour leur soutien, pour leur travail et leur aide continue tout au long de ces quatre années.

Merci aux sociétés CodaSystem et STS-Group pour m'avoir accueilli et proposé un sujet de recherche riche et intéressant.

Ma gratitude va aussi à Ludovic Guillemot, qui m'a encadré au sein de CodaSystem pendant mon stage de master, ainsi qu'au début de ma thèse. Son aide précieuse dans le domaine du traitement de l'image et son implication dans ma formation et ma recherche ont su créer les conditions nécessaires à la réussite ce travail.

Merci aux personnes qui ont contribué aux travaux de recherches, plus spécialement Charles Gilbertz, mais aussi Florent Lesueur, Benoît Chenal, Christophe Clémence, Jean-Philippe Blaise, Philippe Falvo, M. Lauer et J. Lombardi.

J'adresse aussi mes remerciements aux personnes qui ont participé activement au projet ATLAS : Frederic Vanholder, Yann Lanuel, Francine Herrmann, Alexandra Mehat, William Roberts, Lee Banfield et la société NSL.

Je remercie les membres de mon laboratoire de rattachement et plus spécialement les personnels administratifs qui font un travail méticuleux pour faciliter le travail des chercheurs.

Un grand merci à Elisabeth Stadler, pour son soutien et sa participation plus qu'active à la correction de cette thèse.

Un merci particulier à Martial Lienert, qui a su souffrir de relire mon code tout en l'adaptant à la fois pour iOS et Android. *Fire and don't forget !*

Enfin, un clin d'œil à Olivier Commowick qui fournit gracieusement un patron de thèse pour la rédaction sur LaTeX[1].

Merci aussi à mes collègues successifs à CodaSystem, UFR MIM et à l'IUT.

Enfin, merci à mes amis de la 'Division Zz' qui m'ont soutenu plus qu'activement durant tout ce temps.

1. http ://olivier.commowick.org/

Table des matières

Introduction

Le marché des *smartphones* a connu un développement fulgurants ces dernières années. L'augmentation des ventes dans ce secteur est exponentielle depuis 2008. Initialement présents chez les catégories socio-professionnelles les plus élevées, la pénétration de ce type d'équipement concerne toutes les couches de la population et tous les âges. Plusieurs facteurs concourent au succès de ces équipements, malgré leur prix. Tout d'abord, leur « hyper-connectivité ». À la partie téléphonique (*Third Generation* (3G), *Fourth Generation* (4G), *etc.*), s'ajoutent naturellement les possibilités de connexion Wi-Fi et Bluetooth. Le *smartphone* devient un terminal informatique capable de naviguer sur Internet, gérer sa messagerie, ses comptes bancaires, son emploi du temps, etc. Ensuite, le *smartphone* est devenu un « hyper-capteur » intégrant non seulement le son, mais aussi l'image, la géolocalisation, le mouvement (grâce à l'accéléromètre), etc. Les écrans tactiles ont également révolutionné les interactions homme/machine. L'utilisation d'un *smartphone* est à la fois plus efficace et plus efficiente de par l'affordance des interactions tactiles. Enfin, les *smartphones* sont devenus des supports très ouverts sur lesquels les utilisateurs peuvent ajouter des applications couvrant tous les usages (productivité, jeux, tourisme, etc.). De fait, l'objet standard que l'utilisateur achète en rayon, devient rapidement un objet très personnel, tant par les aspects visibles (sonneries de téléphone, applications installées, etc.) que par les usages (stockage d'informations personnelles, voire confidentielles, accès à des services privés sensibles, etc.). La *dématérialisation des procédures* consiste à utiliser ces nouvelles technologies de l'information et de la communication pour réaliser les procédures qui, auparavant, utilisaient le support papier et la communication par voie postale. Les avantages de la dématérialisation sont nombreux. D'abord les nouvelles technologies offrent une palette d'outils extrêmement large pour créer le support avec des assistances très évoluées (par exemple, un traitement de texte avec correcteur orthographique et grammatical, un logiciel de capture et de retouche de photo, *etc.*). L'instantanéité des communications qui, grâce au maillage des réseaux de communication, sont accessibles partout et à chaque instant. La duplication, le stockage, la recherche sont également des avantages indéniables. Enfin, le coût est également un avantage décisif. Un email étant d'un coût infinitésimal comparé à une enveloppe timbrée.

Mais derrière cette révolution aux avantages très visibles, se cache une révolution « culturelle » qui l'est beaucoup moins. En général, un courrier que l'on reçoit par la poste a naturellement une valeur bien comprise par tous. Le courrier identifie clairement le destinataire. Son expéditeur est également identifié, lequel, de plus, appose en principe sa signature. Sauf si c'est une carte postale, le courrier est placé dans une enveloppe qui préserve la confidentialité. Le lieu et la date d'envoi sont précisés également grâce au cachet de la poste, laquelle

apporte sa caution légale. Si l'on considère qu'un email correspond à la dématérialisation d'un envoi postal, force est de constater que par commodité, on lui accorde la même valeur qu'un courrier postal, alors que la comparaison est largement à son désavantage. En effet, dans le cas général, un email est envoyé « en clair » (pas de confidentialité) à partir d'une adresse qui peut être très facilement usurpée et être envoyé depuis n'importe où. Dans la mesure, où l'on ne signe pas numériquement son message, n'importe qui peut finalement vous envoyer n'importe quoi. Donner, ou redonner, de la « valeur » aux informations numériques échangées ou diffusées est un défi qui touche tout le monde et dans toutes les situations.

Notre étude se place dans le domaine des images capturées depuis un *smartphone*. Comme on l'a évoqué précédemment, un *smartphone* est toujours équipé d'un appareil photographique. Toujours à portée de main, il devient le témoin des événements quotidiens, des plus futiles aux plus sensibles. Dans le cas d'un événement sensible (un scoop, un accident, etc.), la photo peut devenir un élément important pour attester, décrire ou comprendre l'événement. Elle devient d'autant plus importante que cet événement devient l'objet d'une procédure judiciaire. La société CodaSystem est spécialisée dans l'image probante capturée depuis un téléphone mobile. Dans le chapitre 1, nous décrirons précisément le contexte dans lequel a été réalisée cette étude. Nous commencerons par y définir la notion de preuve. La preuve peut être vue plus ou moins comme un procédé de signature cryptographique, largement utilisé de nos jours. La manière de lier cette signature à un document quelconque passe communément par le calcul puis le chiffrement d'une empreinte du document. Le résultat est ensuite concaténé au document lui-même. Ici, le type de document traité est une image. Et le défi consiste non pas à concaténer la preuve à l'image, mais à les *fusionner* par un procédé de tatouage. Le défi industriel consiste plus précisément à réaliser ce tatouage dans le contexte particulier où l'image est capturée depuis un smartphone. Pour bien cerner les enjeux en terme de sécurité informatique, nous présentons l'analyse de risque menée sur la chaine de certification, dont le tatouage n'est qu'un maillon. À partir des risques résiduels identifiés dans cette analyse, nous proposerons, dans le chapitre 2, un algorithme de tatouage semi-fragile qui satisfait les contraintes imposées par le contexte : imperceptibilité, capacité, intégrité et localisation. Les points clés de l'algorithme reposent premièrement sur l'utilisation de la technique des *matrix encoding*, laquelle permet de réaliser de manière satisfaisante à la fois la capacité et l'imperceptibilité. Deuxièmement, la sécurité de la marque et de l'intégrité de l'image s'appuie sur l'utilisation de clés de chiffrement destinées à calculer et à répartir les modifications que subit l'image. Cet algorithme sera validé par un ensemble d'expérimentations. La preuve tatouée est composée de l'identité de l'auteur, de l'horodatage et de la géolocalisation. Il va de soi, que si ces données sont erronées ou ne sont pas authentiques, la preuve n'a plus aucune valeur. Le chapitre 3 est consacré à l'authentification de la géolocalisation. Nous décrirons les infrastructures de géolocalisation et

leurs limites. Nous proposerons différents scénarios pour informer l'utilisateur des conditions d'acquisition, voire pour améliorer l'acquisition elle-même. Là encore, les contraintes industrielles nous imposent d'intervenir uniquement au niveau logiciel sur le smartphone (CodaSystem n'avait pas l'ambition de proposer les spécifications d'un nouveau récepteur GPS intégrable dans les smartphones, ni de faire évoluer les systèmes de navigation globaux !). Le fil rouge dans les divers scénarios proposés est de s'appuyer sur *l'hyper-connectivité* d'un smartphone. En effet, même si le récepteur GPS d'un smartphone est le capteur théoriquement le plus précis et le plus fiable pour délivrer l'information de géolocalisation, il est néanmoins possible d'utiliser les autres capteurs (Wi-Fi, GSM) pour collecter d'autres informations afin de confirmer ou mettre en doute cette géolocalisation. Nous verrons également qu'il est possible de faire appel à d'autres services liées aux systèmes de navigation pour affiner cette information. Enfin, dans le chapitre A, nous discuterons de la problématique de l'authentification en général, qu'elle porte sur l'utilisateur, l'horodatage ou la géolocalisation, en insistant notamment sur la notion de tiers de confiance. En général, dans le contexte de la dématérialisation, il n'est pas possible de *s'auto-authentifier*. C'est pourquoi, nous faisons généralement appel à un tiers de confiance, c'est-à-dire un individu (comme un notaire) ou une organisation (comme une autorité de certification) assermentés par les autorités gouvernementales, et qui possèdent les informations et les procédures qui introduisent la confiance recherchée.

Chapitre 1

Contexte

Sommaire

1.1 La sécurité

Lorsqu'Alice sortit de son garage sa voiture neuve, elle était loin d'imaginer se qui allait se passer ce jour-là. Avant de partir au travail, elle sortit son téléphone portable, histoire de prendre quelques photos de sa voiture, afin de rendre ses amies jalouses. Hélas pour Alice, sa joie va être de courte durée. En effet, au premier feu, un automobiliste brule le feu rouge et emboutit sa voiture toute neuve.

Après quelques instants de panique, c'est l'énervement. Qui est cet idiot qui vient de ruiner sa voiture toute neuve ? La personne en question, est Bob. Alice et lui vont devoir maintenant faire leur constat. Alice ressort à nouveau son téléphone pour prendre quelques photos du sinistre. Ce que fait Bob également avec son propre appareil. Mais leurs voitures immobilisées au milieu du carrefour gênent la circulation. Et le concert de klaxons et d'invectives des autres automobilistes rend l'atmosphère encore plus électrique. Bob et Alice sont contraints de déplacer leurs véhicules. C'est alors que les policiers arrivent sur les lieux, alertés par un témoin sans doute. Ils dressent leur propre procès verbal et l'accompagnent de leurs propres photos. Sauf que les véhicules ont changé de place. La version de l'accident que Bob, qui est responsable, décrit aux gendarmes est assez différente de celle d'Alice, qui est la victime...

Deux jours plus tard, l'assurance d'Alice reçoit tous les documents : les photos provenant du téléphone d'Alice, le procès-verbal et les photos des policiers, le constat, les photos prises par le garagiste pour l'expertise des dégâts.

Malheureusement les photos ne sont pas datées, et l'on ne sait pas qui les a prises. Et parfois le cliché ne permet pas de savoir si la photo a été prise chez le concessionnaire ou si elle a été prise sur les lieux de l'accident, si c'est bien la voiture d'Alice qui figure sur le cliché (la plaque d'immatriculation n'est pas toujours visible) et non pas une autre. De plus, comment s'assurer que ces photographies n'ont pas été trafiquées, ou simplement détériorées ?

L'arbitrage de ce litige entre Alice et Bob se heurte à plusieurs problèmes : la quantité d'informations, la qualité de ces informations, la forme de ces informations (photos numériques, témoignages oraux, traces, etc.) et le rôle des différents acteurs de l'accident. Finalement, le dénominateur commun à ces différents problèmes est la « confiance ». Comment faire confiance à ces images (leur provenance, leur contenu, leur intégrité), à ces témoignages ? Comment transférer la confiance qu'une entité a d'une personne (le client de l'assurance, les policiers) à une tierce-partie (le juge, l'assurance concurrente). La réponse nous est donnée par la loi, il faut des preuves. Autant le procès-verbal et les constatations des policiers seront admises comme preuves par un juge, de par leur assermentation, autant les mêmes informations recueillies par Alice et Bob vont être mises en doute. Que faire si les informations divergent ? Produire des éléments conformes à la vérité, éviter de mal les interpréter, de les déformer constituent des défis importants.

Ces considérations nous amènent à nous demander quels moyens nous devons employer pour garantir la force probante, notamment quand la forme est dématérialisée comme les photos numériques. Nous verrons tout au long de ce travail, que transformer une image en preuve requiert l'utilisation de plusieurs mécanismes ainsi que l'application d'un protocole strict. C'est l'ensemble de ces moyens qui créent la preuve. Le sentiment de confiance est souvent voisin du sentiment de sécurité. D'ailleurs, la définition du CNRTL (Centre National des Ressources Textuelles et Lexicales) est : « État d'esprit confiant et tranquille qui résulte du sentiment, bien ou mal fondé, que l'on est à l'abri de tout danger. »[1] Ainsi, un système de sécurité est un ensemble de mesures qui permettent de se sentir « à l'abri du danger ». Le « doudou » d'un enfant est un système de sécurité au même titre que l'alarme d'une voiture ou le coffre-fort d'une banque. Le besoin de sécurité n'est évidemment pas récent. Des anecdotes remontant à l'antiquité illustrent des besoins particuliers dans le domaine de la sécurité de l'information. Demaratus, roi de Sparte, s'est servi de la stéganographie pour transmettre un message à ses alliés en 484 avant Jésus Christ, dans un conflit contre les perses. Il gratte la cire d'une tablette de scribe, y grave son message secret, puis recouvre à nouveau la tablette avec de la cire. De même, il est connu que Jules César (100-44 av. J.-C.), chiffrait ses communications avec un simple décalage de l'alphabet. Cette problématique particulière de la sécurité de l'information, plutôt réservée à l'armée et à la diplomatie, concernent aujourd'hui tous les compartiments de la société, des

1. http ://www.cnrtl.fr/definition/sécurité

activités industrielles et commerciales, à la vie privée.

Le doudou de l'enfant, ou le coffre-fort du banquier sont des systèmes de sécurité très différents, qui répondent à des besoins très différents. Mais généralement, la sécurité se décrit à partir de plusieurs concepts communs : l'assurance, les contre-mesures, la défense en profondeur, les attaques, les risques, les menaces ainsi que les vulnérabilités. L'assurance est une promesse que le système de sécurité va garantir. La menace est un signe par lequel se manifeste l'imminence d'un danger[2]. La vulnérabilité est définie comme une faiblesse dans un système [informatique] permettant à un attaquant de porter atteinte à l'intégrité du système.[3]

Le vocabulaire peut éventuellement varier d'un contexte à l'autre mais l'idée reste généralement la même. Par exemple, en sécurité informatique, une attaque réussie est souvent désignée par le terme « exploit ». Plus précisément, l'exploit est un programme ou une technique démontrant et utilisant l'existence d'une vulnérabilité au sein d'un logiciel informatique. Celui-ci peut se limiter à une démonstration de faisabilité d'exploitation de cette faille (preuve par l'exemple) ou être utilisé pour mener une action malveillante.[4]

Le sentiment de sécurité d'un individu, qu'on désigne souvent par le terme de sécurité globale, est le résultat d'une conjonction de plusieurs catégories de sécurité. Elle commence intuitivement par la sécurité physique, mais elle se complète rapidement par la sécurité politique, financière, *etc.* La sécurité informatique est une catégorie comme une autre. Cependant, ce « sentiment de sécurité informatique » a fortement évolué ces dernières années. Il a tout d'abord été la préoccupation des organisations, des entreprises (banques, administrations, *etc.*), et a été peu ressenti par les individus. Mais rapidement, avec le développement des technologies numériques qui prolongent toutes les facettes de sa vie, les fraudes, les escroqueries, jusqu'aux atteintes à la réputation, ont exacerbé le sentiment d'insécurité de l'individu. Et la sécurité informatique devient aujourd'hui une préoccupation majeure pour lui aussi.

Il est difficile d'améliorer la sécurité dans une catégorie particulière, sans prendre en compte la sécurité dans les autres catégories. L'interdépendance est souvent plus forte qu'il n'y paraît. Par exemple, on ne fait jamais *qu'uniquement* de la sécurité informatique. La protection logique d'un serveur ne sert à rien si son accès physique n'est pas protégé, ainsi que sa source énergétique. Il n'est pas nécessaire de déployer les grands moyens informatiques pour provoquer un déni de service, si tous le monde a accès à la prise de courant qui alimente le serveur ! L'inter-dépendance entre les catégories de sécurité est forte. Et l'on ne peut pas satisfaire tous les besoins en même temps : c'est un travail trop important et trop coûteux.

2. http ://www.cnrtl.fr/definition/menace
3. http ://www.clusif.asso.fr/fr/production/glossaire
4. http ://www.clusif.asso.fr/fr/production/glossaire

> Que vos inquiétudes soient liées à votre sécurité personnelle face à l'augmentation des crimes, liées à la sécurité de l'information pour vos affaires d'entreprise, ou encore à la sécurité face au terrorisme, les problèmes de sécurité nous affectent de plus en plus dans notre vie de tous les jours et nous devons tous faire un effort pour mieux les comprendre. Nous devons arrêter d'accepter aveuglement ce que les politiciens et les grands pontes nous disent. Nous devons aller de l'avant et commencer à faire des compromis de sécurité raisonnables. [5]
>
> Beyond fear : thinking sensibly about security in an uncertain world
> [Schneier 2003]

Comme l'illustre la citation de Bruce Schneier, ce n'est pas raisonnable et un compromis s'impose. On peut également ajouter, qu'il n'existe pas un seul compromis, mais qu'il en existe autant que de besoins. Dans la mesure où notre étude est centrée sur la sécurité informatique, nous nous efforçons de préciser quel compromis est raisonnable dans ce domaine particulier, et pour les besoins spécifiques de l'image probante.

Revenons sur l'histoire d'Alice et Bob. Dans cette histoire, l'assureur requiert plusieurs garanties. Premièrement, la justification de l'origine des photos : est-ce bien Alice qui a pris les photos ? Est-ce que ce sont les policiers ? Ou bien est-ce Bob, ou encore un parfait inconnu ? Cette fonction est *l'authentification*. On veut pouvoir garantir l'origine des informations dont on dispose. On veut également être certain que ces photos n'ont pas fait l'objet de modifications, qu'elles soient intentionnelles ou non. On parle alors *d'intégrité* de la photo. L'assureur peut aussi vouloir conserver secrètes les informations qu'il reçoit, c'est-à-dire non accessibles à des personnes non autorisées. Pour cela, il lui faudra mettre en place des mesures de *confidentialité*. Cette fonction est une autre facette importante de la sécurité informatique.

Finalement, il est important que les informations soient accessibles aux personnes autorisées. Il s'agit du critère de *disponibilité*. En effet, une information non accessible n'est d'aucun usage. C'est d'ailleurs souvent un objectif visé par l'attaquant. L'attaque visant à empêcher l'accès à l'information s'appelle un *déni de service*.

Du point de vue de l'assurance, la confiance qu'elle peut avoir dans les photographies qu'on lui a confiées, relève typiquement de la sécurité informatique. Le contexte est conforme à la définition donnée par l'ISO :

5. Whether your concern is personal security in the face of increasing crime, computer security for yourself or your business, or security against terrorism, security issues affect us more and more in our daily lives, an we should all make an effort to understand them better. We need to stop accepting uncritically what politicians and pundits are telling us. We need to move beyond fear and start making sensible security trade-offs

> Protection de la confidentialité, de l'intégrité et de la disponibilité de l'information ; en outre, d'autres propriétés, telles que l'authenticité, l'imputabilité, la non-répudiation et la fiabilité peuvent également être concernées.
>
> *Normes ISO27001:2005 [ISO 2005]*

En bref, la sécurité en science de l'information est décrite par ses fonctions. Les trois principales, comme nous l'avons vu, sont : la confidentialité, l'intégrité et la disponibilité. Mais il est habituel d'inclure aussi l'authentification et la *non-répudiation* (le fait qu'une information ne puisse être reniée par son auteur). Dans la publication "Standards for Security Categorization of Federal Information and Information Systems", le NIST présente les objectifs de sécurité en terme d'impact potentiel sur ces trois fonctions.

Nous avons donné ici des définitions simples mais communément admises par la communauté. La norme ISO27001:2005 donne des définitions très voisines et d'autres définitions sont également disponibles dans le NIST Handbook [NIST 1995].

Les moyens mis en œuvre pour satisfaire ces différentes propriétés de sécurité sont des outils classiques de cryptographie (fonctions de chiffrement, de hashage) ou de théorie du codage (codes correcteurs), ainsi que des méthodes de contrôles d'accès au sens large (contrôle physique, filtrage, ...). Ces disciplines sont beaucoup plus vastes que ce dont on aura besoin dans le cadre de cette étude. De même que la sécurité informatique n'est qu'une partie de la sécurité globale, les techniques que nous utiliserons ne sont qu'une partie de la sécurité informatique. Il n'est pas nécessaire de prendre en compte tous les problèmes de sécurité. C'est une tâche vaine et coûteuse. C'est pour cela que dans le domaine de la sécurité, une entreprise n'a pas d'obligation de résultats, mais a une obligation de moyens. En d'autres termes, la fiabilité absolue n'existe pas. Il existe toujours un risque résiduel. C'est pourquoi, la réponse à la question « quels moyens mettre en place ? » est obtenue en réalisant l'analyse (identification, compréhension) des besoins de sécurité. Les sections 1.2 sur la dématérialisation et 1.3 sur la preuve vont nous permettre de bien comprendre les besoins et les objectifs généraux de la sécurité liés aux documents numériques. Nous verrons, dans la section 1.4, le contexte industriel dans lequel se place notre étude. Nous en déduirons les besoins et les risques particuliers du système étudié. Cela nous amènera finalement à déterminer les solutions à déployer pour répondre aux obligations de moyens.

1.2 La dématérialisation

Dans la section précédente, nous avons donné quelques définitions générales de la sécurité, notamment de la sécurité informatique. Pour augmenter ce sentiment de confiance qu'est la sécurité, les moyens de protection sont nombreux et variés, que ce soit du point de vue technologique, financier ou organisationnel. Une boutade souvent entendue chez les administrateurs réseau est que : « si vous ne voulez pas que votre ordinateur subisse une attaque, déconnectez-le du réseau ! ». Mais plutôt que d'en arriver à cette extrémité, il est préférable de reprendre la citation de Schneier : « Nous devons aller de l'avant et commencer à faire des compromis de sécurité raisonnables ». Et pour être raisonnable, il est indispensable de bien comprendre les besoins et estimer les risques. Nous évaluons tous les jours, et le plus souvent sans nous en rendre compte, la confiance des documents « papier » que nous manipulons. Le contrat de vente signé chez un notaire apparaît comme d'une sécurité parfaite. Personne ne met en doute l'authenticité de la feuille d'impôt ou du relevé bancaire qu'il reçoit dans sa boite aux lettres. Et cela alors même que les moyens d'impression dont tout un chacun peut disposer sont suffisamment performants pour reproduire fidèlement ces documents. Cette section propose de faire le parallèle entre les documents papiers et les documents numériques afin de mettre en évidence ce qui change, ou au contraire ce qui reste identique.

Et c'est un fait acquis que les usages actuels tendent à la conversion au tout numérique, à la dématérialisation. Cette constatation est illustrée par l'avènement des cartes à puces, cartes Radio-frequency identification (RFID), de la biométrie (passeport, future carte d'identité électronique, ...), qui deviennent des compagnons de tous les jours. De multiples usages quotidiens ont été importés sur l'Internet, comme par exemple la photographie. En effet, après une forte croissance chez les professionnels avec l'apparition du réflex numérique, elle connaît maintenant un véritable boom depuis l'arrivée sur le marché grand public de téléphones toujours plus performants et équipés d'appareils photographiques numériques de série. Pour appuyer cette remarque, citons le changement de stratégie de Kodak en 2004 qui arrêta de produire ses films argentiques, suivi par AGFA (renommée AGFAPhoto) quelques mois plus tard. L'un de leurs concurrents direct sur le plan argentique, ILFORD, fit faillite peu de temps après en dépit de son succès passé.

Le passage au numérique ne se fait pas sans changement de valeurs. Ce qui était avant original, de par la nature même de l'argentique des appareils photographiques, ne vaut plus pour le numérique. Ainsi pouvait-on considérer une photographie prise devant l'horloge de l'hôtel de ville de Paris comme datée et géo-localisée. Mais les traitements de l'image numérique évoluant toujours, cette donnée n'est plus fiable aujourd'hui. Les manipulations et ré-exploitations sont bien trop aisées.

Impossible de parler numérique sans penser Internet. Or depuis peu,

l'usage de l'Internet devient omniprésent. Cela implique une multiplication des utilisateurs et la massification des usages. Dans le même temps, l'utilisation du mobile comme outil de communication va croissant. Le mobile est devenu un vecteur multi-canaux : utilisation de la 3G, du Wi-Fi, du BlueTooth, mais il est devenu également multi-applications et trouve des utilisations dans un nombre croissant d'activités comme la cuisine, le sport, les loisirs. On citera à titre d'exemple le développement des *applications stores* de Apple, Research In Motion (RIM), Microsoft ou encore Android, qui proposent toujours plus de logiciels pour toujours plus de nouvelles utilisations. En cela le mobile devient le véritable prolongement de l'ordinateur personnel, toujours à portée de main. Du côté des entreprises, celles-ci trouvent pratique d'équiper leurs agents nomades avec des appareils polyvalents pouvant bénéficier d'applications de remontées de terrain, ou de géo-localisation. Le téléphone tire profit de ses capacités de communication à la fois pour envoyer et recevoir des informations. Cela permet ainsi d'effectuer ou de déléguer plus facilement certaines tâches.

Le potentiel des applications grand public, invite à profiter de l'équipement mobile de l'utilisateur pour qu'il produise et transmette lui-même des informations vers un organisme (entreprise ou administration). Il est important de préciser que cette démarche doit être vue comme un moyen d'améliorer la qualité de service globale en réduisant par exemple les temps de collecte de l'information et/ou les temps de traitement. Cela peut également contribuer au développement durable en évitant des déplacements inutiles et des impressions coûteuses. Actuellement, lorsqu'un usager confie son véhicule sinistré à un garage, dans la plupart des cas, ce n'est plus l'expert qui se déplace. Le garagiste prend des photos numériques qu'il transmet aussitôt à l'expert via le réseau informatique. Autrement dit, l'utilisation des équipements mobiles permet d'accroître l'efficacité générale des procédures administratives entre utilisateurs et organisations. Il faut cependant tenir compte d'un problème critique : ces informations sont-elles pertinentes et fiables ?

De par la nature du signal numérique, il est complexe de s'assurer qu'une photographie n'est pas une capture d'une autre photographie par exemple. La photographie d'un accident de face ou de profil n'apporte pas forcément les mêmes détails. L'un des deux angles peut être inutile. De même, les détails capturés volontairement ou non par l'utilisateur contribueront, ou empêcheront l'analyse. Cet exemple met en évidence un phénomène important : la mise à disposition de ces technologies induit un transfert de responsabilité. Dans l'exemple d'un sinistre, il s'agit de la responsabilité de *la collecte* des informations d'expertise. Mais dans d'autres cas, cela pourra être la responsabilité de *la diffusion* des informations (mises en lignes de photos ou de vidéos). Si la technologie connaît bien entendu des limites, l'élément humain reste un point crucial. Cette forme d'utilisation de la technologie a un impact qui n'est pas toujours bien mesuré. De plus, elle met en valeur l'importance primordiale de la formation de l'utilisateur.

.

La production et la diffusion de contenus numériques dans le cadre professionnel ou privé s'accompagne d'une demande toujours plus forte de protection. Mais on peut s'interroger sur ce qui doit être protégé. Les besoins actuels sont très variés : protection de la vie privée avec le développement des réseaux sociaux ; protection de la propriété intellectuelle pour les diffusions de films à la demande ; protection contre la fraude depuis le déploiement du paiement en ligne ; protection contre l'usurpation d'identité sur les services de l'administration en ligne ; *etc.* L'histoire d'Alice et Bob illustre de manière réaliste les usages et problèmes actuels. Pour répondre à ces problèmes, de multiples solutions de chiffrement, d'authentification ou de contrôle d'intégrité ont d'ores et déjà été proposées et sont en évolution constante. Les premiers standards de cryptographie comme le Cyclic Redundancy Check (CRC) et Data Encryption Standard (DES) ont progressé pour donner les outils bien connus aujourd'hui que sont Rivest Shamir Adleman algorithm (RSA), Secure Hash Algorithm (SHA) et Advanced Encryption Standard (AES). Essayons de faire un inventaire des usages courant des documents physiques, et voyons quelles méthodes numériques peuvent les remplacer. (Tables 1.1 et 1.2)

Table 1.1 – Monde physique

Fonction	**Usage physique**
Conserver, éviter la dégradation	Archivage, protection physique des supports
Restreindre l'accès	Gestion de l'accès physique
Accélérer l'accès	Tri, indexation, classement
Certifier la propriété	Enregistrement officiel, copie conforme
Acter l'engagement	Signature manuscrite
Empêcher/détecter la reproduction	Filigrane, impression taille-douce, fils de sécurité

Pour conserver et éviter de dégrader les supports, les usages physiques et numériques se ressemblent : conservation des supports dans des salles prévues à cet effet. L'avantage du numérique est que la duplication des informations est beaucoup plus aisée. Cependant, sur la durée, l'archivage d'un document numérique n'est pas aussi évident qu'il n'y paraît. En effet, si vous souhaitez relire un tableau MULTIPLAN (remplacé par EXCEL en 1984 !) que vous aviez stocké sur une disquette 5"1/4, votre seule solution est de vous rendre dans un musée de l'informatique car il est peu probable que vous ayez pris soin de conserver un lecteur de disquette 5"1/4 en état de marche avec un ordinateur capable d'exécuter MULTIPLAN, et dans la bonne version de votre document... L'archivage numérique, et c'est un des rares inconvénients de la dématérialisation,

Table 1.2 – Monde numérique

Fonction	Usage numérique
Conserver, éviter la dégradation	Protections pour support numériques. Backups, archivage numérique
Restreindre l'accès	Gestion des droits d'accès, chiffrement
Accélérer l'accès	Tri, indexation, classement
Certifier la propriété	Estampillage, authentification
Acter l'engagement	Signature numérique
Empêcher/détecter la reproduction	Contrôle d'intégrité, estampillage, tatouage

impose des procédures de gestion de document électronique très rigoureuses, beaucoup plus contraignantes que le stockage physique.

La restriction d'accès est gérée par un contrôle des personnes ayant accès à l'information. Dans le mode physique, il faut restreindre l'accès physique au lieu de conservation des informations : carte d'accès, coffre-forts. Pour le numérique, la même méthode peut être employée mais on peut aussi utiliser d'autre techniques comme le chiffrement des données.

Pour avoir une meilleure disponibilité de l'information, il est parfois nécessaire d'accélérer l'accès à ces données. Des techniques d'archivage et des moyens techniques existent mais restent limités : armoires, classeurs, ... Dans ce domaine, l'informatique apporte une réelle plus value grâce aux techniques de tri et d'indexation automatique, des moteurs de recherches, le tout avec des performances inégalables.

Pour certifier la propriété, différents mécanismes sont utilisés dans le monde physique : copie conforme, enveloppe Soleaux ainsi que le recours à des organismes tiers comme la SACEM, l'INPI. Dans le monde numérique, ces techniques restent valables exceptée la copie conforme qui est facilement contestable. Cependant l'horodatage certifié permet aux documents numériques d'être facilement marqués comme existants à une date t.

L'engagement (à un contrat par exemple) est matérialisé par la signature. Signer est une procédure qui apparaît assez simple intuitivement. Mais en réalité, ce qui en fait un engagement légal repose sur plusieurs propriétés que nous rappellerons dans la section suivante. Il reste que la signature manuscrite est un geste, parfois simple, parfois compliqué, mais dans tous les cas aisément falsifiable. La signature numérique, grâce aux techniques cryptographiques apparaît nettement plus sûre, et est légalement reconnue depuis 2000 en France. Par ailleurs, lorsque la signature manuscrite engage non pas l'in-

dividu qui signe, mais l'organisation qu'il représente (comme le commercial qui signe un contrat de vente au nom d'une multinationale), les habilitations du signataire ne sont pas toujours simples à gérer et à contrôler. Là encore, l'informatique apporte sa plus value en masquant cette complexité.

Pour empêcher la reproduction, ou au moins être en mesure de la détecter, de nombreuses techniques sont employées dans le monde physique : les filigranes, les impressions taille-douce et les fils de sécurité. Beaucoup de ces techniques sont utilisées pour imprimer les billets de banques ou les diplômes. Dans le domaine numérique, les techniques de contrôle d'intégrité, d'horodatage et de tatouage sont des méthodes qui permettent de répondre à ce problème.

Tous les documents sont soumis à ces fonctions. La dématérialisation ne fait apparaître finalement aucune fonction nouvelle, ni n'évite les fonctions habituelles des documents physiques. Au contraire, l'informatique apporte son efficacité. Mais tous ne nécessitent pas le même niveau de sécurité pour chacune de ces fonctions. Par exemple, un document public ne requiert aucune confidentialité, mais peut nécessité un stockage sûr et pérenne. Dans le contexte que nous développerons par la suite, les documents sont des images et l'on s'intéressera principalement à deux fonctions, l'authentification et la détection de modification (l'intégrité). L'objectif final est de construire une image probante, c'est-à-dire telle que sa valeur soit la plus indiscutable possible dans un tribunal. Mais, si l'utilisation d'une preuve par le juge est un processus intuitif, une preuve attestée permet au juge une décision juste. La manière ou les éléments que le juge apprécient pour accorder une valeur probante sont plus compliqués. C'est pourquoi, dans la section suivante, nous rappelons comment la loi française décrit cette notion fondamentale à l'application de la justice.

1.3 La preuve littérale

Si Alice et Bob n'arrivent à s'entendre à l'amiable sur les circonstances de l'accident, et donc sur les responsabilités de l'un et de l'autre, ils (ou leurs assureurs) devront faire appel au juge pour résoudre leur litige. Chacun va tenter d'apporter la *preuve* de sa bonne foi en produisant ses propres éléments d'information, parmi lesquels nous trouverons vraisemblablement les photos, mais aussi des témoignages, etc. Mais qu'est-ce qu'une preuve ? Comment le juge décide-t-il d'accepter un élément comme preuve ou de le rejeter. Nous résumons, dans cette section, les notions fondamentales que la loi dicte en matière de preuve.

Chaque jour, nos actes, parfois même les plus insignifiants, sont régis par un protocole. Par exemple, prendre le train pour se rendre à une certaine destination, nécessite la réalisation ordonnée d'une série d'actions : indiquer la

destination à l'agent, payer le billet de train, monter dans le bon train et finalement descendre à la destination souhaitée. Ce protocole commercial de base comme tous les protocoles qui régissent les interactions entre les participants (au moins deux !) répond à un certain nombre de caractéristiques :

- tous les participants doivent connaître le protocole ;
- tous les participants acceptent le protocole ;
- tous les participants comprennent les étapes du protocole de la même manière (non ambiguïté du protocole) ;
- toutes les situations possibles sont prévues par le protocole (complétude du protocole).

Les règles d'un jeu de société sont typiquement un protocole. On voit très aisément que si l'une de ces caractéristiques n'est pas satisfaite par ces règles, les conflits apparaissent très rapidement entre les joueurs. Il en va de même pour tous les protocoles que l'on utilise dans la vie courante, de l'achat du billet de train ou d'une maison, au protocole diplomatique pour résoudre un différend international.

Si l'on part du principe qu'un protocole est complet et non ambigu, les difficultés proviennent alors des participants eux-mêmes. Soit ils ne connaissent pas le protocole (mais « nul n'est censé ignorer la loi »), soit ils n'acceptent pas le protocole. Et la loi se charge alors d'arbitrer les conflits entre les participants.

> « Celui qui réclame l'exécution d'une obligation doit la prouver. Réciproquement, celui qui se prétend libéré doit justifier le paiement ou le fait qui a produit l'extinction de son obligation »
>
> Code civil, article 1315.

Cet article 1315 du code civil français, est le premier du chapitre VI : « De la preuve des obligations et de celle du paiement ». Cette partie du code civil énumère cinq moyens de preuve :

- **La preuve littérale** : c'est la preuve écrite. Depuis l'ordonnance de Moulins de 1566, le droit français repose essentiellement sur la forme écrite. La signature manuscrite devient le moyen le plus fiable d'authentifier les actes. La loi du 13 mars 2000 adapte le code civil à la forme électronique. La signature électronique a désormais la même valeur que la signature manuscrite. Ces actes écrits sont dits « authentiques » s'ils sont rédigés par un officier public (notaire, huissier, etc.), ou dits simplement « acte privé » (acte sous seing privé, les écrits spéciaux, les quittances, etc.) ;
- **La preuve testimoniale** : c'est la preuve qui s'appuie sur un témoignage. C'est alors le juge qui détermine la force probante du témoignage ;

- **La présomption** : la présomption simple est une preuve obtenue par déduction à partir d'indices multiples et concordants (article 1353). La encore, c'est au juge d'estimer la valeur probante : « Les présomptions qui ne sont point établies par la loi, sont abandonnées aux lumières et à la prudence du magistrat... » ;
- **L'aveu** : un aveu est fait de manière unilatérale. Intuitivement il apparaît comme une preuve convaincante. Toutefois, lorsqu'il est contraire à la réalité des faits, le juge peut ne lui accorder aucune force probante ;
- **Le serment** : « Levez la main droite et dites "je le jure" ». Ces paroles entendues dans tous les procès, représentent un exemple de serment, dont l'énoncé est définitif et implique celui qui le prononce. Un faux serment est réprimé par la loi.

Le droit français reconnaît globalement deux niveaux de fiabilité d'une preuve :

- la preuve parfaite (ou preuve absolue, incontestable)
- la preuve imparfaite (ou présomption simple)

Seule une preuve littérale est considérée comme parfaite. Sa force probante est constituée de deux aspects indissociables : la force probante quant à l'origine de l'acte et la force probante quant au contenu de l'acte. Ces deux objectifs sont généralement réalisés par la signature. La signature manuscrite est habituellement un graphisme, autrement dit un « geste très personnel », qui identifie la personne qui l'appose. C'est cette signature qui donne à l'acte sa force probante. Sans elle, un acte n'a juridiquement aucune valeur.

On peut rappeler ici les propriétés recherchées par une signature, qu'elle soit manuscrite ou électronique :

- **l'authenticité** : la signature, dans la mesure où c'est une marque personnelle, indique qui a signé le document. Dans le cas de la signature électronique, cette marque personnelle peut, par exemple, prendre la forme d'un certificat contenant une clé privée associée à l'identité de celui auquel elle appartient ;
- **l'infalsifiabilité** : la signature est un geste personnel, sous entendu qu'elle ne peut être reproduite que par son auteur. Dans le cas d'une signature électronique, et dans l'exemple où elle s'appuie sur un certificat c'est-à-dire un fichier informatique, cette propriété ne peut être le résultat que d'une gestion particulièrement rigoureuse de ce fichier (ne pas le prêter, ne pas le laisser accessible sur un ordinateur, etc.) ;
- **l'inaltérabilité** : la signature apposée sur un acte, ne peut pas être supprimée (on ne signe jamais au crayon de papier !). Dans le cas d'une signature électronique, aucun procédé informatique ne doit pouvoir supprimer la signature ;
- **la non-répudiation** : si la signature apparaît sur un acte, l'auteur de la signature ne peut pas renier avoir pris connaissance de l'acte et l'avoir approuvé ;

- **l'intégrité** : une fois l'acte signé, il ne peut plus être modifié ;
- **l'unicité** : cette propriété est propre à la signature numérique. Elle fait dépendre la signature du document lui-même. Ainsi, deux documents différents signés par le même auteur (avec son certificat par exemple), produira deux signatures différentes.

Nous avons rappelé ici quelques éléments juridiques fondamentaux destinés à mieux comprendre ce qu'est une preuve et quelles caractéristiques elle doit avoir pour obtenir au mieux la confiance du magistrat. Il reste que la force d'une preuve ne résulte pas du simple choix de sa forme, mais résulte de toute la démarche qui a entouré la production de cette preuve, et qui continuera à préserver sa valeur tout au long de son existence. Par exemple, un aveu obtenu sous la contrainte n'a pas de valeur probante. Une signature numérique réalisée avec un certificat auquel tout le monde peut accéder n'a pas de valeur. Dans le domaine de la preuve numérique, de nombreuses technologies, algorithmes et protocoles, permettent en théorie d'atteindre un niveau de confiance qui dépasse souvent celui obtenu dans un contexte écrit traditionnel. Les infrastructures à clefs publique (*Public Key Infrastructure* (PKI)) sont un exemple de protocole d'algorithme de signature particulièrement utilisé. Cet exemple sera détaillé au chapitre A. En médecine, l'efficacité d'un médicament n'est jamais garantie à 100%. Cette efficacité dépend de la précision du diagnostic du médecin (la pathologie est-elle précisément celle ciblée par le médicament ?). Elle dépend également de la posologie (le médecin a-t-il bien évalué le poids, l'âge, etc. pour définir le dosage optimal ?). Enfin, l'efficacité du médicament dépend aussi du patient lui-même (prend-il son médicament au bon moment, dans les bonnes conditions et sur la bonne durée ?). Dans une métaphore un peu risquée, la sécurité informatique par rapport au monde numérique correspond à ce que la médecine est à l'organisme humain. C'est une science inexacte, qui permet de prévenir ou de réagir à des défaillances de l'infrastructure. Le traitement à appliquer pour une sécurité optimale dépend de la finesse de l'analyse des risques du contexte, du niveau des mesures de sécurité choisies pour combler les problèmes identifiés, et leur respect par les acteurs du contexte.

La migration des usages vers le numérique a résolu certains problèmes : facilité de traitements, fiabilité de conservation, rapidité d'accès et de tri. Cependant, si la conservation est plus aisée, c'est parce la duplication à l'identique a été rendue plus facile par le numérique... Par exemple, dans le cas de la photocopie certifiée conforme, la modification du document était plus difficile et plus facilement détectable grâce à certains mécanismes simples comme l'apparition d'artéfacts visuels caractéristiques. Dans le cas du numérique, cela n'est plus le cas car un bit ne se distingue pas d'un autre. Dès lors, il n'est pas possible de constater une perte d'intégrité grâce au seul contenu du document. Il faut des techniques supplémentaires (Table 1.2). Ces variations majeures du scénario physique force à porter une attention très particulière vis-à-vis des moyens mis en œuvre dans le domaine numérique.

Dans la partie suivante, nous nous attacherons à présenter le contexte industriel dans lequel s'inscrit notre étude. Nous présenterons le parternaire industriel, la société CodaSystem. Nous décrirons la démarche suivie pour réaliser l'analyse de risques de la chaine de certification d'images de CodaSystem et les résultats obtenus, résultats qui ont conduit à une évolution majeure de leur logiciel de tatouage Shoot&Proof.

1.4 CodaSystem

Nous avons vu dans les sections précédentes que les nouveaux usages numériques réclament de nouveaux moyens de protection allant jusqu'à la nécessité d'avoir force probante. C'est dans cette optique que travaille CodaSystem. Plus précisément leur application Shoot&Proof vise à apporter à la photographie une valeur de preuve. Shoot&Proof s'appuie sur une infrastructure sécurisée qui permet de certifier les documents multimédia produits sur appareil mobile. S'agissant de la photo, l'objectif de CodaSystem peut se résumer par la formule suivante : « donner à la capture photo la notion d'original numérique ». Le dispositif mis en place pour atteindre cet objectif est basé sur l'utilisation d'une technique de tatouage permettant de certifier que la capture photo a été effectuée à tel endroit, telle heure, par telle personne, avec tel appareil, *etc.* A l'instar de l'usage du négatif dans le cas de l'argentique, la photo certifiée doit obligatoirement être dupliquée et c'est la copie qui pourra subir des traitements d'images, recadrage, *etc.* Une image ayant subi ce genre de traitements ne pourra être qualifiée de conforme. La distinction entre la copie et l'original ne peut pas seulement s'appuyer sur des critères visuels, mais doit également faire appel à des procédés numériques spécifiques, comme le tatouage d'image par exemple. C'est cette dernière idée qui a attiré l'attention de CodaSystem, qui a donc mis en place un projet de recherche composé d'un stage de Master 2 suivi d'une thèse industrielle.

Depuis le concept original, le processus de certification des photos numériques est composé d'une suite de modules complémentaires qui composent trois étapes principales de la chaîne de certification : la capture (I), la transmission (II) et le stockage (III). Ces étapes sont illustrées sur la figure 1.1. Notons que la notion de tatouage initialement présente dans cette chaîne est proche d'un HMAC lié à une méthode de stéganographie. Elle consiste a embarquer dans l'entête d'une image les informations codées.

L'étape de capture débute par l'identification de l'utilisateur (par mot-de-passe, code Personnal Identification Number (PIN), etc.) suivie de la capture proprement dite, au cours de laquelle est appliqué le tatouage. L'étape de transfert vers les serveurs de CodaSystem, effectuée *via* une connexion mobile dite *over the air* (General Packet Radio Service (GPRS) Enhanced Data Rates for GSM Evolution (EDGE) ou Universal Mobile Telecommunications System (UMTS)), est précédée par une phase de cryptage (cryptage temporaire si la

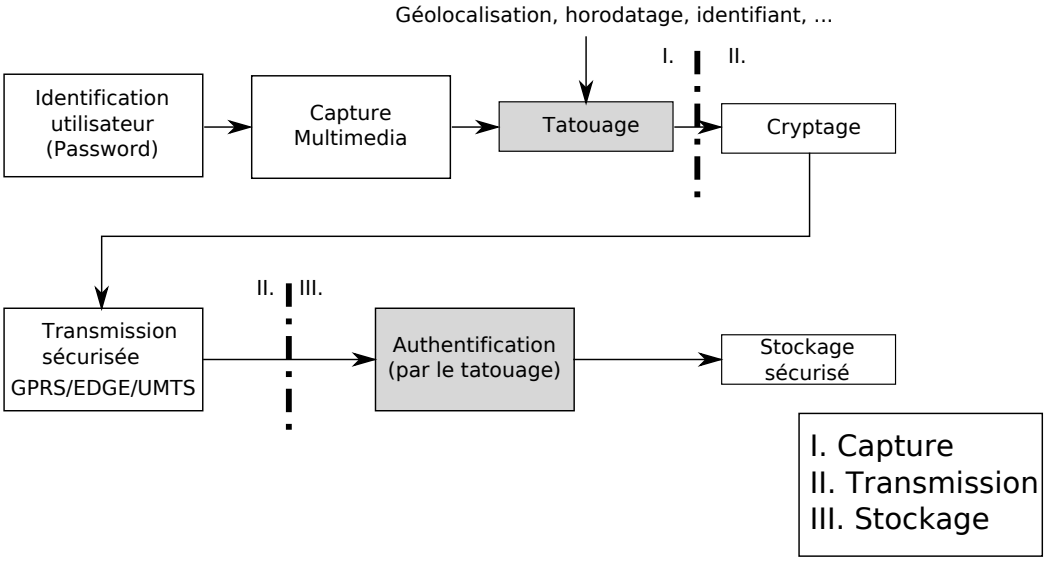


Figure 1.1 – La chaîne de certification de CodaSystem, de la prise de photo au stockage sécurisé

transmission ne peut avoir lieu, puis cryptage pour l'envoi du document). Une fois les données réceptionnées, elles sont authentifiées (on vérifie le tatouage) et stockées pour être disponibles pour l'utilisateur. Celui-ci peut alors consulter à la fois ses données en tant que telles et consulter les rapports *forensics* constitués des informations de certification et de la validation de chaque étape constituant la chaîne de certification.

Du côté client, le produit permet la production d'une image certifiée. On ne cherche pas à garantir la confidentialité de l'image après que l'original certifié est stocké de manière sûre. Entre la capture et le stockage, la confidentialité est évidemment critique pour minimiser la probabilité d'une attaque ou d'une altération de l'image.

L'objectif de recherche proposé par CodaSystem est d'obtenir une force probante pour des images capturées depuis des terminaux mobiles. En 2008, tout reposait sur la chaîne de certification et le HMAC dans l'entête JPEG. Un prototype, présenté comme un tatouage d'image, existait déjà (nous verrons plus tard que cette technique relevait plus de la stéganographie que du tatouage, bien que ces deux notions soient proches). Il apparaissait assez évident qu'il fallait renforcer ce module de tatouage.

Pendant le stage préalable à cette thèse, nous avons travaillé sur un prototype existant de tatouage. Nous avons analysé la technique stéganographique développée. Cette analyse a révélé plusieurs incohérences conduisant à une utilisation de temps exagérée lors de l'exécution sur mobile. Le stage s'est poursuivit par l'amélioration de ce prototype, par la simplification des étapes trop

coûteuses et la prises de choix pour combler les manques laissés par des étapes supprimées qui cherchaient la meilleure configuration. De plus, l'analyse a permis de changer le fonctionnement de la modification de l'image en supprimant certaines étapes de compression JPEG inutiles lorsque l'image d'entrée est déjà compressée.

Au début de la thèse, nous avons commencé par formaliser l'intérêt d'un tatouage d'image. L'utilisation du tatouage revêt plusieurs avantages par rapport aux techniques classiques basées sur un calcul d'empreinte. En premier lieu, les données permettant de vérifier l'intégrité, l'authenticité, la non-répudiation et la certification de l'image font partie intégrante de celle-ci. Par conséquent elles ne sont soumises à aucune contrainte liée à la mise en place d'un système d'information dédié à la gestion spécifique de ces données. Dès lors que l'on a en sa possession l'image, on peut à tout moment effectuer les opérations de contrôle liées au tatouage. En d'autre termes, ces mécanismes ne sont rattachés, par exemple, à l'utilisation d'aucune base de données et sont actifs tout au long de la vie de la photo numérique, cela même lorsque l'image est sortie de la chaîne de certification (*i.e* lorsque l'image a été restituée à son propriétaire). L'extraction du tatouage s'effectue alors à partir d'une plate-forme web dans laquelle l'image tatouée est chargée par le client afin d'y être analysée. L'accès relativement facile à l'image via Internet, pourrait faire craindre qu'elle puisse facilement être copiée. Nous insistons sur le fait que ce qui a de l'importance ici, c'est la protection de l'original. Par exemple, si un utilisateur est capable de réaliser ne serait-ce qu'une copie d'écran contenant l'image, puis de réencoder l'image dans le bon format et avec les bons paramètres, il copie certes les données tatouées mais n'obtient cependant qu'une copie conforme. L'original reste la référence en terme de preuve quelles que soient les utilisations réalisées avec les copies, conformes ou non. Une fois l'originale stockée de manière sûre, la confidentialité des copies relève de la responsabilité du client. En effet, la duplication d'une preuve n'est pas un problème en soi. Qui plus est, l'original de cette preuve reste conservé dans le coffre-fort numérique de CodaSystem et constitue toujours la référence. Des copies de cet original sont fournies « à la demande » au client, qui peut ainsi les utiliser pour ses besoins d'exploitations (en tant que preuve et/ou image). Il peut décider de les diffuser, sans pour autant perdre le statut de preuve.

Mais pour apporter une véritable sécurité, il était important de réaliser une étude plus approfondie de la chaîne de certification dans son ensemble. En effet, depuis 2001, la chaîne de certification n'a cessée d'être modifiée et d'évoluer. L'absence de sécurité absolue nous a obligé à réaliser une analyse des risques pesant sur la chaîne de certification. C'est ainsi que nous avons mis avantageusement à profit l'implémentation par CodaSystem de la norme ISO27001. Cette norme de sécurité informatique demande en effet de réaliser l'identification et l'analyse des risques pesant sur les actifs (anglais : asset - un actif est un composant du système d'information qui possède une valeur pour l'entreprise et qui peut être géré (qui possède un cycle de vie). [Définition Wi-

kipédia]) du système d'information de l'entreprise. Après un bref historique de la norme, nous détaillons l'analyse de risque de la chaîne de certification de CodaSystem et les remises en question que celle-ci apportent, notamment au niveau du tatouage.

1.4.1 ISO27001

La norme ISO/IEC 27001 [ISO 2005] décrit les spécifications pour établir, implémenter, surveiller, revoir, maintenir et améliorer un système de gestion pour les risques informatiques. C'est la première famille internationale de standards sur la sécurité informatique.

La norme provient d'un standard anglais nommé *ISMS*[6] *certification standard* BS7799-2:1999 puis BS7799-2:2002. En 2005 il devient ISO/IEC 27001:2005. Ce standard est aussi accompagné par le *Code of Practice standard* BS7799-1:1999 intégré par l'ISO en 2000 sous le nom ISO/IEC 17799:2000 révisé en 2005 (ISO/IEC 17799:2005) puis ISO/IEC27002:2005. Le standard est géré par l'ISO/IEC JTC 1/SC 27 (Joint Technical Comitee 1 / Sub-comitee 27).

Dans la révision de 2002 (BS7799-2) est introduit le modèle du cycle de Demming, aussi connu sous le nom de Plan-Do-Check-Act (PDCA). C'est un point important de la norme qui définit la gestion de la sécurité ISO27001 comme un cycle d'améliorations continu. En bref, le modèle PDCA fonctionne en quatre étapes :

- Plan : décrire ce que l'on va faire ;
- Do : faire ce que l'on a décrit ;
- Check : vérifier que l'on a fait ce que l'on a décrit ;
- Act : corriger lorsque la phase Check révèle des anomalies.

L'objectif d'ISO27001 est d'assurer la sécurité de l'information en termes de confidentialité, d'intégrité et de disponibilité. Cette sécurité doit s'aligner sur les objectifs stratégiques de l'entreprise pour préserver sa compétitivité, sa rentabilité ainsi que ses contraintes légales et règlementaires. Cela passe par la mise en place du Système de Management de la Sécurité de l'Information (SMSI) grâce auquel le Responsable des Systèmes de Sécurité de l'Information (RSSI) (Responsable des systèmes de sécurité de l'information) va pouvoir surveiller, contrôler et minimiser les risques, les contraintes ainsi que la bonne gestion de la sécurité. Le document ISO27005 explique en détail comment réaliser un traitement du risque dans la conformité avec ISO27001.

L'intérêt d'être certifié ISO27001 consiste, pour CodaSystem, à affirmer son rôle de Tiers de Confiance. Dans ce contexte, pouvoir se prévaloir d'une bonne gestion de la sécurité de l'information est un élément clef dans la construction d'un réseau de confiance. C'est pour cela que CodaSystem est devenue la pre-

6. ISMS ou SMSI : système de management de la sécurité de l'information

mière entreprise privée certifiée ISO27001 du Luxembourg. Les produits de CodaSystem fonctionnent en grande partie sur des téléphones mobiles. *C'est donc un défi supplémentaire pour l'application d'ISO27001 qui est, certes, adapté à tout type d'environnement, mais qui ne propose cependant que très peu d'outils pour la gestion de la* mobilité*. Le travail de cette thèse s'inscrit dans cette démarche. C'est l'analyse de risques qui a fourni les pistes pour l'amélioration de la chaîne de certification et des fonctionnalités nécessaires à la construction d'une preuve.*

1.4.2 Principe et méthode de l'analyse de risques

L'étape la plus importante de l'analyse dans ISO27001 est l'analyse de risques. C'est elle qui va permettre d'évaluer quels sont les risques qu'encourt actuellement l'entreprise, et va mettre en évidence les points à renforcer dans la sécurité du système. ISO27001 ne préconise aucune méthode particulière. Elle stipule cependant que :

> L'organisme doit effectuer les tâches suivantes : [...] définir et approuver l'approche d'appréciation du risque de l'organisme :
>
> 1. Identifier une méthodologie d'appréciation du risque adaptée au SMSI ainsi qu'à la sécurité de l'information identifiée de l'organisme et aux exigences légales réglementaires ;
> 2. Développer des critères d'appréciation des risques et identifier les niveaux de risques acceptables. La méthodologie d'appréciation du risque choisie doit assurer que les appréciations du risque produisent des résultats comparables et reproductibles.
>
> Extrait d'ISO 27001 – 4.2.1c [ISO 2005]

> La direction doit fournir la preuve de son implication dans l'établissement, la mise en œuvre, la surveillance et le réexamen, la mise à jour et l'amélioration du SMSI par : [...]
>
> f. La détermination des critères d'appréciation des risques et des niveaux de risques acceptables.
>
> Extrait d'ISO 27001 – 5.1.f [ISO 2005]

La méthode d'analyse de risques de CodaSystem repose sur les standards ISO27005 et la méthode EBIOS. Les critères de sécurité (confidentialité, intégrité et disponibilité) sont quantifiés selon une échelle à quatre niveaux. Ils déterminent la valeur de l'actif. Dans notre méthode, chaque fonction du SMSI

est classée sur une échelle de niveaux. Chaque niveau indique pour un critère quel serait l'impact si le critère n'était pas respecté. Les listes suivantes décrivent ces niveaux.

- Confidentialité :
  - Niveau 0 :Impact nul ;
  - Niveau 1 : Divulgation d'informations de CodaSystem ;
  - Niveau 2 : Divulgation d'informations de CodaSystem et de ses clients ;
  - Niveau 3 : Divulgation d'informations des utilisateurs (direct ou clients de clients).
- Intégrité
  - Niveau 0 : Impact nul ;
  - Niveau 1 : N/A ;
  - Niveau 2 : Modification d'informations non-critiques et/ou récupérables ;
  - Niveau 3 : Modification d'informations critiques et/ou irrécupérables.
- Disponibilité
  - Niveau 0 : Impact nul ;
  - Niveau 1 : Perturbation pour CodaSystem sans impact sur les clients ;
  - Niveau 2 : Perturbation mineure pour le client ;
  - Niveau 3 : Perturbation majeure pour le client.

Prenons quelques exemples concernant le niveau « Confidentiel ». Un communiqué de presse destiné à être diffusé à une date d ne doit pas être divulgué avant cette date pour ne pas trahir une opération marketing. Ce communiqué est donc classé « Confidentiel niveau 1 ». Si cette information venait à être publiée avant, cela n'impacterait que la stratégie de CodaSystem. Une fois l'annonce faite, le document passerait à Confidentiel niveau 0, puisque divulguer cette information n'aurait plus d'impact (l'information serait déjà publique). Les informations de la base de données utilisateurs sont, quant à elles, classées Confidentiel niveau 3, car il s'y trouve des informations de CodaSystem, des clients de CodaSystem ainsi que des clients de ses clients (données personnelles et nominatives entre autres).

Un risque est la combinaison d'une vulnérabilité et d'une menace. La menace n'est pas en notre contrôle et se distingue donc par sa probabilité (est-ce que l'attaque est facile ou non, ...). La vulnérabilité correspond à la possibilité de failles inconnues, mais dépend surtout des mesures que l'on a déjà mises en place (par exemple : réponse sur incident, isolation de composants, gestion des privilèges, ...).

Les critères de vraisemblance du risque sont qualifiés en tant que menace

et vulnérabilité. La quantification de ces critères suit la liste suivante :

- Menaces
  - Niveau 1 : Très improbable selon les statistiques, ou le coût du niveau d'expertise nécessaire
  - Niveau 2 : Peut arriver de temps en temps
  - Niveau 3 : Très probable, facile à mettre en œuvre, pas d'investissement ou d'expertise particulière
- Vulnérabilités
  - Niveau 0 : très faible : mesures en place efficaces contre la menace
  - Niveau 1 : moyen : mesures insuffisantes ou mal adaptées
  - Niveau 2 : élevée : pas de mesure de protection efficace en place
  - Niveau 3 : très élevée : manque de mesures, ou obsolètes ou inappliquées

La nomrme n'impose aucune formule pour l'évaluation du niveau de risque. Dans le cadre de l'implémentation du système de gestion de la sécurité de l'information de CodaSystem, l'équipe en charge de la sécurité, en partenariat avec le Centre de Recherche Public du Luxembourg Henri Tudor, a déterminé la formule de mesure du risque suivante :

$$\text{Niveau De Risque} = max(\text{Valeurs de l'actif}) \times (\text{Menace} + \text{Vulnérabilité} - 1)$$

Dans cette formule, les valeurs de l'actif correspondent aux trois valuations d'impact pour la confidentialité, l'intégrité et la disponibilité d'un actif. Cette formule donne un risque valué entre 0 et 15.

Il nous faut maintenant définir le niveau d'acceptabilité du risque. Nous avons jugé qu'un risque est inacceptable dans l'une des trois situations suivantes :

- Quand une attaque (ou une défaillance) peut arriver de temps en temps (menace >=2), et n'a pas de mesures de protection (vulnérabilité=3), et dont la valeur de l'actif est 3.
- Quand une attaque (ou une défaillance) est très probable et est facile à mettre en œuvre (menace = 3), ne fait pas l'objet de mesures de protection ou fait l'objet de mesures non efficaces (vulnérabilité>=2), et dont la valeur de l'actif est 3.
- Quand une attaque (une défaillance) est très probable et facile à mettre en œuvre (menace = 3) n'a pas de mesures de protection (vulnérabilité=3) dont la valeur de l'actif est 2.

En résumé, tout risque supérieur ou égal à 10 est considéré comme un risque inacceptable.

		Menace + vulnérabilité - 1					
		0	1	2	3	4	5
Valeur de l'actif	1	0	1	2	3	4	5
	2	0	2	4	6	8	10
	3	0	3	6	9	12	15

Table 1.3 – Mesures des risques

1.4.3 Bilan

Si le document d'analyse de risques n'est pas particulièrement pertinent dans cette thèse, il est cependant très utile pour mieux comprendre ce qu'est un risque critique. À partir de ces critères de criticité, la chaîne de certification a été analysée étape par étape.

Phase de capture : Les risques identifiés pendant la phase de capture sont : capture par un périphérique simulé (émulateurs de téléphone), empêcher l'accès au périphérique de capture, dériver le capteur sur un nouveau matériel. Tous ces risques ont été évalués comme non critiques, excepté pour la simulation de capture. La méthode choisie pour réduire ce risque est la vérification du module logiciel de capture photo. La réduction de ces risques ne fait pas l'objet de cette thèse.

Géolocalisation : Nous avons identifié, avec la société *Nottingham Scientific Limited* (NSL) dans le projet ATLAS (Chapitre 3), plusieurs risques. Ils sont regroupés en trois catégories : les perturbations naturelles, militaires et civils. Il n'existe actuellement pas de mesures pour réduire les risques naturels (impossibilité physique) ni militaires (impossibilité technologique ou coût trop excessif). Les risques principaux à traiter sont les risques de perturbation civils.

Identification de l'utilisateur : Les principaux risques sont le détournement du système de validation d'identité et l'usurpation d'identité. En effet, si quelqu'un vole un mobile, il serait en mesure de certifier des photographies en lieu et place de l'utilisateur légal.

Horodatage : le risque est l'utilisation d'une date différente de celle de la capture réelle. La fiabilité de cette information est indispensable à la construction d'une image probante.

Tatouage : les risques concernant le tatouage sont multiples : effacement de la marque, modification de la marque, insertion d'une marque choisie, copie

d'une marque existante sur un nouveau document, etc. Dans la mesure où c'est le procédé retenu pour construire physiquement la preuve dans l'image, il est nécessaire de veiller à ce que ce procédé soit le plus sûr possible.

Stockage sur équipement mobile : les risques pesant sur le stockage sont la perte d'intégrité des documents stockés, la diffusion non autorisée d'informations confidentielles, l'effacement de données ou la substitution de données. Des mesures de cryptage et d'identification du propriétaire sont déjà en place, limitant le risque à un niveau non critique. Cependant la mise en œuvre d'un mécanisme d'expiration des données mobiles a été retenu comme pertinent.

Transmission : Les principaux défauts qui peuvent impacter la transmission sont : le déni de service, l'interception de données ou la falsification de données transmises. Les mesures mises en place actuellement ont été estimées comme suffisantes.

1.5 Conclusion

Dans ce chapitre nous avons décrit ce qu'est la sécurité. Elle s'articule autour de quatre grands sous-domaines qui sont : la sécurité informatique, physique, politique et monétaire. Chacun de ces sous-domaines est fortement interdépendant, mais nous ne nous focalisons que sur le sous-domaine informatique. Même en réduisant le champ de la sécurité à l'informatique, nous avons vu qu'il n'est pas possible de mettre en place toutes les mesures de sécurité. La sécurité est une affaire de compromis. Il s'agit d'abord d'identifier les vulnérabilités et les menaces qui pèsent sur les actifs et l'impact que pourrait avoir une attaque sur ces actifs. Partant de là, on peut choisir de mettre en place des mesures pour protéger leur confidentialité, leur intégrité ou leur disponibilité. Mais les mesures à mettre en place ne se déduisent pas automatiquement. Elles dépendent entièrement du contexte et des besoins. De plus, le niveau d'effort à déployer n'est pas infini. Il doit avant tout correspondre à l'obligation de moyens nécessaire au niveau de confiance souhaité.

Pour bien cerner cet objectif, nous avons rappelé ce qu'est une preuve, identifié les fonctions de sécurité souhaitées par CodaSystem pour ses produits de certification et mené une analyse de risques pour évaluer leurs criticités. Cette dernière a été menée dans le cadre de la certification ISO27001 de l'entreprise. Elle a permis de révéler les points faibles de la chaîne de certification. Ces différents points nous donnent les différents axes sur lesquels est basée la recherche décrite dans cette étude.

Nous avons défini les objectifs de CodaSystem concernant la certification d'image. Cette certification passe par le développement d'un mécanisme de tatouage évolué permettant d'intégrer les informations nécessaires à la fabrication d'une preuve. L'analyse de risques a levé plusieurs problématiques concer-

nant le processus de création et de gestion de preuve. Le tatouage doit assurer un certain nombre de fonctionnalités : éviter la duplication, assurer l'intégrité et l'authentification du message. Tous ces points feront donc l'objet d'un traitement particulier durant les phases de conception et de réalisation détaillées au chapitre 2.

Chapitre 2

Tatouage de certification

Sommaire

Passer de formats analogiques à des formats numériques a non seulement modifié le format des informations, mais également les moyens de production et de diffusion. Par exemple, une photographie argentique est un objet physique dont la copie nécessite des équipements et des produits certes accessibles, mais qui restent au mieux du domaine de l'amateur éclairé. La diffusion des photographies subit ces contraintes de production et nécessite une organisation qui, de fait, est plus facile à contrôler. À l'inverse, les formats numériques ne nécessitent que des équipements et des logiciels informatiques tout à fait conventionnels, comme on en trouve dans chaque foyer. Les appareils photo numériques, caméscopes numériques, téléphones portables, « webcams », *etc.*

sont autant de moyens de productions simples, efficaces et peu onéreux. Et désormais, Internet offre un moyen de diffuser aussi facilement et massivement cette production. On peut également noter que cette diffusion peut se faire de manière complètement anonyme. Et sans être particulièrement formé, un individu peut facilement produire ou reproduire ce que contient son ordinateur. Dès lors, la question qui se pose est « quels sont les contrôles possibles dans ce contexte ? » La cryptographie tente d'apporter des solutions à ce problème de contrôle.

Dans sa thèse [Bas 2000], Patrick Bas décrit au chapître deux, les caractéristiques détaillées du tatouage, de la cryptographie et de la stéganographie. Nous n'en résumons ici que les différences majeures. La cryptographie consiste à chiffrer un message pour le rendre inintelligible. Les symboles qui composent le message clair sont recalculés. Ce qui produit en général un message de même longueur où la valeur des symboles et/ou leur position ont été modifiées. Le schéma cryptographique standard comprend deux fonctions, l'une de chiffrement pour l'émetteur du message, et l'autre de déchiffrement pour le récepteur. À la différence de la cryptographie, la stéganographie est une méthode de dissimulation. La valeur des symboles du message n'est pas nécessairement modifiée. Les symboles sont placés à l'intérieur d'un support hôte intelligible mais sans rapport avec le message, dont le rôle consiste à « leurrer » l'attaquant. Seule la personne avertie, celle qui sait « où » regarder, pourra retrouver le message. Enfin, le tatouage consiste quant à lui, à insérer une information en "superposition" d'un média existant. En cela, il se rapproche de la stéganographie, excepté que le tatouage n'est pas nécessairement dissimulé, et l'hôte n'est pas un leurre. L'hôte *est* l'information à transmettre. Selon l'objectif de sécurité visé sur un type d'information donné, il est possible d'associer les trois domaines, c'est-à-dire dissimuler dans l'hôte un message chiffré en rapport avec l'hôte lui-même.

Si le principe général du tatouage consiste à embarquer un message supplémentaire à l'intérieur même du support hôte, la nature de cette information et la manière d'embarquer cette information permettent de distinguer différents usages [Guillemot 2004]. L'une des premières applications importantes du tatouage est la protection de la propriété intellectuelle (ou droit d'auteur, qui se distingue sensiblement du *copyright* anglo-saxon). Dans un premier temps, le tatouage concentrait la majorité des efforts de recherche dans le domaine du traitement du signal, pour la protection des droits d'auteurs en particulier. Il s'agit d'identifier le propriétaire d'un média, en insérant un identifiant dans le média. Dans ce cas, la marque insérée doit pouvoir résister au mieux aux modifications qu'est susceptible de subir le média. Il s'agit de préserver au maximum ce qui va permettre aux ayants droits de revendiquer la possession du média. Par exemple, redimensionner une photographie tatouée ne doit pas faire disparaître son tatouage. L'auteur de la photo originale doit également être l'auteur de la photo redimensionnée.

La protection contre la copie (*copy control*) consiste en revanche à empê-

cher la copie du média. C'est notamment le cas avec les DVDs, pour lesquels on insère une marque destinée à reconnaitre la source des images si elles venaient à être copiées ([Petitcolas 1998]). De la même manière, un « beta-test » de jeu vidéo peut intégrer une marque tatouée sur l'écran. Cela permet d'identifier l'auteur d'une fuite d'information si des copies d'écrans sont diffusées.

Le tatouage sert également à protéger le contenu, c'est-à-dire à s'assurer que le média n'a pas été modifié. Il ne s'agit ni d'empêcher la copie, ni d'identifier l'auteur, mais de garantir par un contrôle d'intégrité la conformité du média à l'original. Différents travaux ont porté sur les niveaux de finesse de ce contrôle d'intégrité. Le niveau le plus bas consiste simplement à détecter la modification. À un degré plus élevé, il est possible de localiser ce qui a été modifié. Enfin, le niveau le plus fin consiste à déterminer la nature de la modification. Pour ce type de tatouage, Guillemot [Guillemot 2004] compare la marque incluse dans le média à un mouchard qui va témoigner des modifications effectuées.

Enfin, une autre fonction du tatouage est l'enrichissement de document, ou l'indexation. Dans ce cas, une information secondaire est ajoutée. Cette information peut être de nature diverse (lien, données personnelles, ou autre). [Bas 2000]

Dans le contexte de notre étude, la nature du média est l'image compressée, ici au format JPEG. Nous commencerons par rappeler les principales étapes de la compression JPEG. Dans les paragraphes précédents, nous avons utilisé implicitement un certain nombre de concepts comme *tatouage*, *hôte*, *marque*, *etc.*. Selon le point de vue où l'on se place, selon sa sensibilité scientifique (traitement du signal, sécurité informatique, *etc.*.), les termes qui correspondent à ces concepts sont parfois différents. Et des termes identiques recouvrent parfois des significations subtilement différentes. Dans la deuxième section, nous nous attacherons à donner une définition aussi neutre que possible du tatouage, même s'il est parfois difficile de se détacher complètement du contexte particulier de ce travail, où la sécurité est primordiale. Nous essaierons néanmoins de distinguer clairement les propriétés du tatouage et celles qui sont propres à la sécurité. Ensuite, nous définirons le concept de tatouage de certification en précisant ses caractéristiques requises et souhaitables. Dans une quatrième section, nous détaillerons notre proposition d'algorithme pour les images issues d'un équipement mobile. Enfin nous analyserons le comportement de cet algorithme, ses performances et résultats.

2.1 La compression JPEG

Dans cette section, nous allons nous attacher à développer le fonctionnement de la compression JPEG. Ce format est le format de prédilection pour des images dites « réelles », car il a été conçu pour tenir compte de caractéristiques qui se retrouvent dans une image naturelle. Le format JPEG a été normalisé par

l'ISO/IEC dans la norme [ISO 2004].

La compression JPEG peut se décomposer en quatre grandes étapes (cf. Fig 2.1) :

- Transformation de l'espace colorimétrique ;
- Transformation en cosinus discrets ;
- Quantification ;
- Codage run-length et codage entropique.

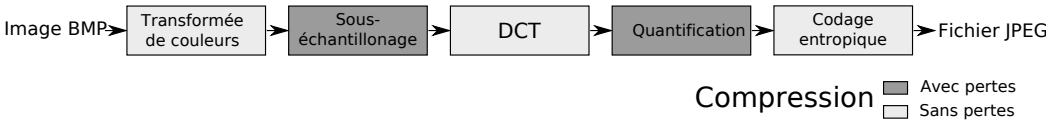


Figure 2.1 – Processus de la compression JPEG

Transformation de l'espace colorimétrique : Il s'agit d'une étape extrêmement coûteuse en calculs qui consiste à transformer de manière linéaire l'image de l'espace RGB[1] dans un espace YCbCr[2] afin de réordonner l'information contenue dans l'image. Chaque pixel peut s'exprimer selon une composante appelée luminance (intensité lumineuse du pixel) et deux composantes correspondant à l'information de couleurs (chrominances). Le contenu des chrominances est moins important du point de vue psycho-visuel. Ces composantes sont souvent sous-échantillonnées afin de diminuer la quantité d'informations à transmettre. La conversion de l'espace RGB à YCbCr est réalisable grâce aux équations 2.1, 2.2 et 2.3.

$$Y = 0,299.R + 0,587.G + 0,114.B \quad (2.1)$$

$$Cb = -0,1687.R - 0,3313.G + 0,5.B + 128 \quad (2.2)$$

$$Cr = 0,5.R - 0,4187.G - 0,0813.B + 128 \quad (2.3)$$

Transformation en cosinus discrets : La transformation DCT s'applique à des blocs de pixels de taille 8x8 qui deviennent, après transformation des blocs, des coefficients DCT (blocs DCT). Elle a pour but de réorganiser l'information contenue dans le bloc par fréquence spatiale. Elle a été présentée par Ahmed et *al.* dans [Ahmed 1974]. Intuitivement, les basses fréquences représentent les variations lentes au sein du bloc alors que les hautes fréquences correspondent aux détails. Cette réorganisation va permettre de faciliter le choix des données à « sacrifier » pour diminuer le poids de l'image tout en maintenant une qualité acceptable (phase de quantification). Le calcul du coefficient DCT (i, j) (i et j

1. *Red Green Blue* - RVB Rouge Vert Bleu
2. Luminance (Y), chrominance bleue (Cb), chrominance rouge (Cr)

étant les coordonnées du coefficient dans le bloc) d'un bloc $N \times N$ s'effectue ainsi :

$$\mathrm{DCT}(i,j) = \frac{2}{N} C(i)C(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \mathrm{pixel}(x,y) \cos\left[\frac{(2x+1)i\pi}{2N}\right] \cos\left[\frac{(2y+1)j\pi}{2N}\right] \tag{2.4}$$

La DCT est une transformation linéaire qui associe un vecteur n-dimensions à un ensemble de coefficients. La combinaison linéaire de n vecteurs de base connus pondérés par ces coefficients donnera le vecteur original. Les vecteurs de base sont des vecteurs sinusoïdal. Dans JPEG, on utilise un transformation à deux dimensions. Dans l'équation précédente, $C(i), C(j) = \sqrt{\frac{1}{N}}$ si $i, j = 0$ et $C(i), C(j) = \sqrt{\frac{2}{N}}$ si $i, j \neq 0$, u la fréquence spatiale horizontale, v la fréquence spatiale verticale et N vaut ici 8.

Le choix des blocs de données 8x8 (total de 64 coefficients) comme structure de base de la chaîne de compression JPEG s'explique par le fait que la décomposition fréquentielle DCT s'applique à des signaux stationnaires. Les images dites « naturelles » sont de manière notoire hautement non stationnaires. Il est, en revanche, plus probable de penser qu'un groupe formé de 64 pixels constitue une zone spatiale aux propriétés relativement similaires sur toute sa surface.

Ce traitement des données de manière séquentielle (par bloc de 64 coefficients) permet de minimiser la mémoire requise lors de l'exécution du programme de compression JPEG et permet de multiples optimisations du code, notamment dans le domaine de l'embarqué. C'est pour cette raison que l'algorithme de tatouage admet comme contrainte un traitement séquentiel des données, bloc par bloc.

On distingue deux types de blocs DCT :

- le coefficient DC correspond à la valeur moyenne de la sinusoïde (premier coefficient du bloc).
- les coefficients AC regroupent les variations dues aux détails (tous les autres). Plus on avance dans la liste des coefficients AC, plus on considère des valeurs qui ont une faible influence dans le codage de l'image.

On peut ainsi représenter cet ensemble de coefficients par une matrice, comme celle présentée sur l'image 2.2.

Quantification : La quantification est l'étape engendrant les pertes d'informations majeures au sein de la chaine de compression. Cela consiste à appliquer à un bloc DCT donné une matrice de quantification qui va diminuer la précision de la représentation de chaque coefficient DCT. L'équation 2.5 montre comment l'on calcule le coefficient DCT quantifié C_{QT} :

$$C_{QT}(i,j) = \lfloor DCT(i,j)/QT(i,j) \rfloor \tag{2.5}$$

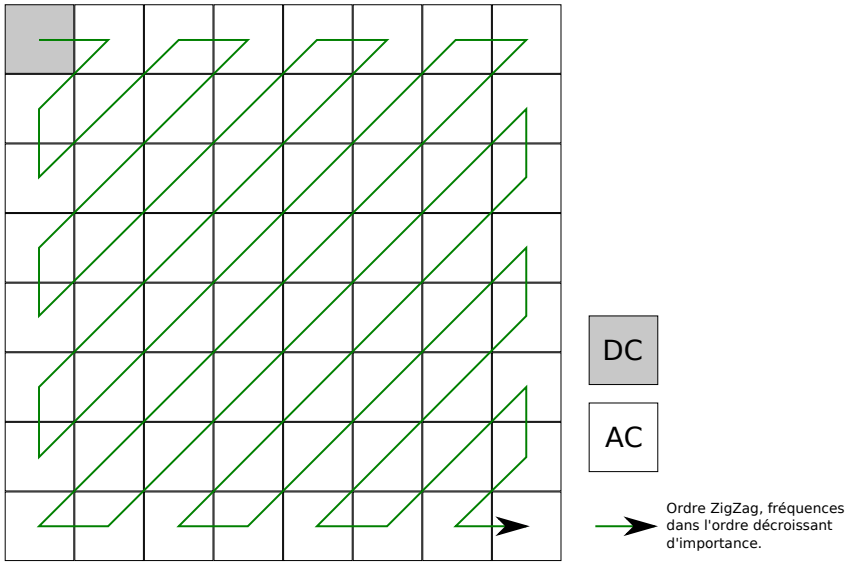


Figure 2.2 – Coefficients dans un blocs DCT et importance des fréquences.

```
DCT: matrice des coefficients DCT, QT: matrice de quantification
```

On obtient un bloc 8x8 de 64 coefficients de 8 bits. On remarquera la large quantité de coefficients réduits à zéro. C'est grâce à cela que l'image pourra être fortement compressée.

Un détail supplémentaire qui nous sera utile par la suite est l'interprétation binaire d'un coefficient. Un coefficient est un octet se composant d'un bit de poids faible nommé LSB (pour *Least Significant Bit*). Les autres sont des bits de poids fort, est sont appelés MSB (*Most Significant Bits*), comme le montre la figure 2.4.

LSB s'effectue ici sous certaines contraintes liées à un compromis

Codage *run-length* : Une fois les données quantifiées, elles sont compressées sans perte à l'aide d'un codeur entropique. Cette compression se passe en deux étapes, la première est un codage *run-length,* et le second est un codage de Huffman. Le codage *run-length* tire avantage des longues plages de zéros créées par la quantification. En particulier, la dernière plage de zéros est codée par le code « fin de bloc ».

Exemple :	4 4 4 4	5 5	3	0 0	1 1 1 1 1	2	0 0 0
	(4 ; 4)	(5 ; 2)	(3 ; 1)	(0 ; 2)	(1 ; 5)	(2 ; 1)	EOB

Codage entropique : Le codage de Huffman utilise des tables de compression pré-caculées spécialement pour JPEG. Elles permettent de réduire l'espace utilisé en réécrivant avec des codes binaires plus courts les symboles les plus

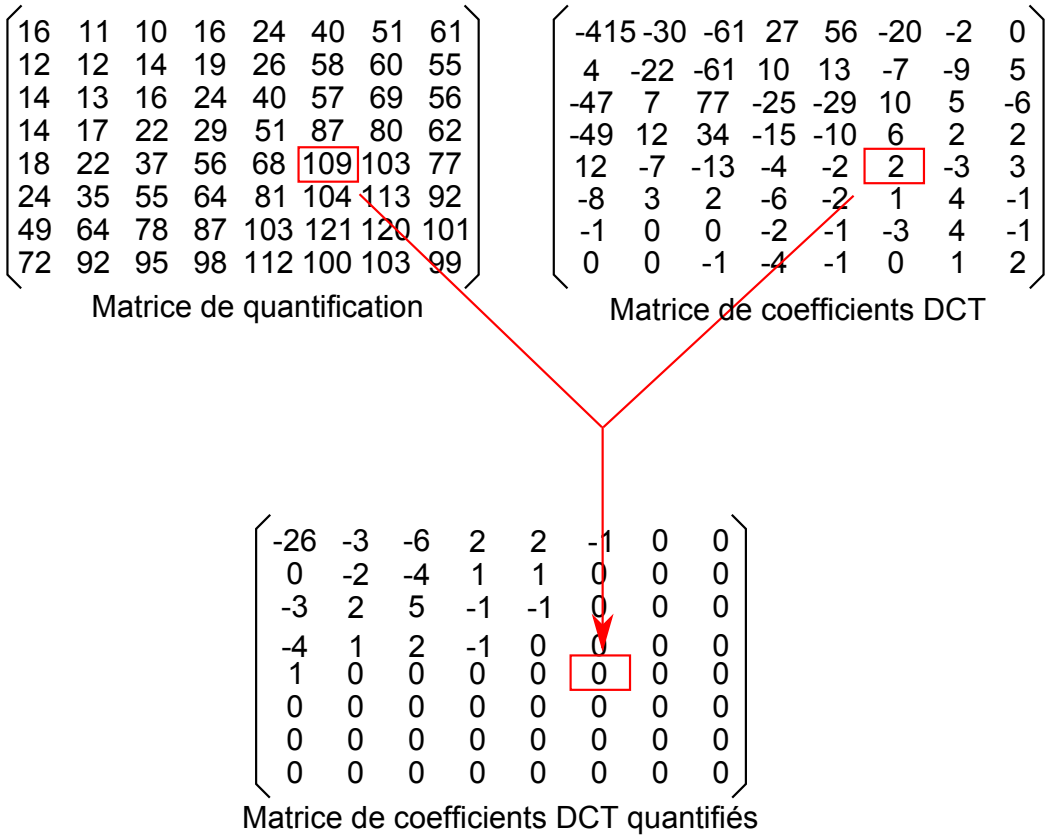


Figure 2.3 – Quantification des coefficients DCT par la matrice de quantification, où par exemple, 2 / 109 = 0

Figure 2.4 – Bits de poids faible/fort

fréquents. C'est un algorithme de compression classique en informatique. Il permet d'obtenir une compression sans perte. Un exemple de coadage par un code de Huffman est illustré sur l'arbre de la figure 2.5. Le code d'une lettre est donné par le parcours depuis la racine jusqu'à la feuille choisie en codant 0 pour chaque branche gauche parcourue et 1 pour chaque branche droite parcourue.

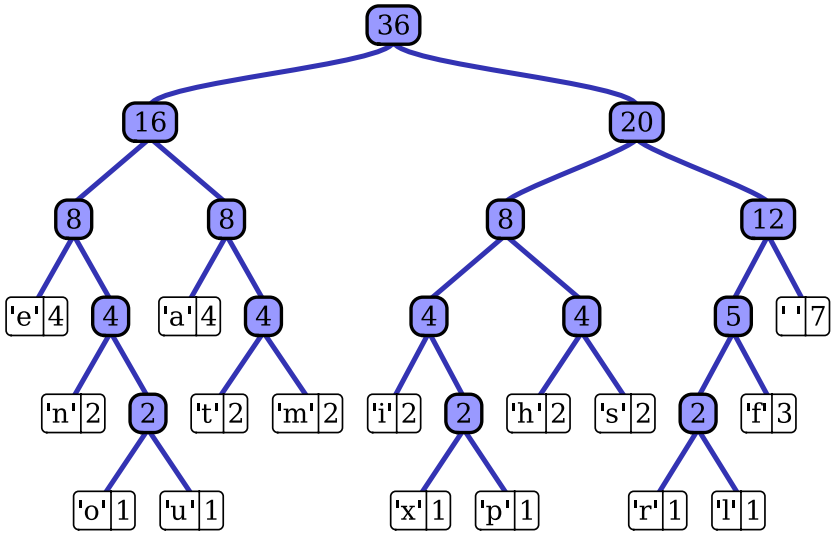


Figure 2.5 – Exemple d'arbre de Huffman

2.2 Le tatouage d'images

Le tatouage en général, et le tatouage d'images en particulier, est un domaine très actif depuis la fin des années 90. De nombreux travaux ont approfondi le tatouage dans des contextes et avec des objectifs très variés. De fait, toute une terminologie a été introduite, souvent claire mais parfois confuse dès lors que d'autres domaines sont concernés, par exemple la sécurité informa-

tique. Aussi, pour éviter ces confusions, nous proposons de clarifier ces termes en rappelant les définitions de la communauté lorsqu'elles sont admises, ou en choisissant celles qui conviennent le mieux à notre contexte en cas d'ambigüité.

2.2.1 Définitions

Comme nous l'avons souligné, le terme tatouage est souvent connoté suivant le contexte d'utilisation. Par exemple dans l'article de [Furon 2002] le tatouage est très fortement lié au traitement du signal. De plus, comme l'objectif visé par ces travaux est la protection des droits d'auteur, la propriété de robustesse est sous-entendue en permanence. A cette définition, nous préférons nous appuyer sur la définition plus générale donnée dans [Cox 2008].

_ Définition 1, Tatouage
Le tatouage est le processus qui consiste à modifier une œuvre en y incrustant une information relative à l'œuvre elle-même : la marque.

Le terme « œuvre » utilisé dans cette définition est bien le terme inscrit dans la loi française concernant les droits d'auteurs par exemple. Par ailleurs, cette définition est suffisamment générale pour couvrir tous les types de supports (multimédia, ou autres) et toutes les caractéristiques possibles du tatouage que nous introduirons par la suite. Mais dans le contexte de cette étude, les œuvres sont des œuvres numériques. Là encore, la terminologie varie selon les communautés. On utilisera dans la suite de multiples synonymes tels que média, support, signal, ou hôte.

Comme nous l'avons vu précédemment, le tatouage a plusieurs fonctions. Pour chacune de ces fonctions, la marque est composée de manière différente. Ce peut être une information sur le producteur du média, une image, des données de reconstruction, etc. Ce qui importe c'est d'insérer cette marque dans l'hôte.

La détection ou l'extraction de la marque varie selon les objectifs et les qualités visés. Suivant le cas, cette phase peut nécessiter des informations complémentaires, comme éventuellement l'hôte d'origine. On peut alors distinguer deux types de méthodes : le tatouage informé, et le tatouage aveugle.

_ Définition 2, Tatouage informé
Le tatouage informé (*informed watermarking*, ou *source extraction*) utilise le média original pour détecter ou extraire la marque dans le média tatoué.

_ Définition 3, Tatouage aveugle
Le tatouage aveugle (*blind watermarking*, ou *destination extraction*) consiste à extraire la marque du média tatoué sans utiliser l'original.

On trouve un exemple de comparaison entre tatouage informé et aveugle dans [Miller 2000]. Les deux types de tatouages impliquent un traitement différent de la source. Dans le premier cas, le tatouage informé, le média représente

un bruit. On lui ajoute la marque qui sera ensuite retirée par soustraction du bruit original. En revanche, dans le cas du tatouage aveugle, on ne peut considérer l'image comme du bruit puisque qu'on ne dispose pas de l'original comme référence. Nous devons utiliser un maximum d'informations concernant cette source pour embarquer la marque de la manière la plus fiable possible.

Eggers et Girod ont présenté dans [Eggers 2001] un tatouage aveugle d'authentification utilisant un schéma de tatouage appelé *Scalar Costa Scheme*. Ce schéma, développé dans un précédent papier, permet aux auteurs de récupérer l'information d'authentification sans l'image originale. C'est un concept qu'il nous faudra reprendre pour la validation d'une image en tant que preuve. En effet, l'image doit nous fournir à elle seule l'information qui nous est nécessaire pour l'authentifier. On peut remarquer que dans ces deux cas, informé et aveugle, la marque peut-être ou non connue à l'avance.

D'autres types de tatouage existent, mais sont plus marginaux. [Lu 2001] introduit le cocktail de tatouage défini ainsi :

_ **Définition 4, Cocktail de tatouage**
Le cocktail de tatouage est une méthode réunissant plusieurs tatouages possédant des caractéristiques opposées. Ces tatouages sont utilisés indépendamment dans la même image pour embarquer deux marques différentes.

C'est une technique qui nécessite de prendre certaines précautions, notamment pour éviter que le premier marquage ne soit effacé par le second.

[Kundur 1999] définit, quant à lui, le tatouage témoin :

_ **Définition 5, Tatouage témoin**
Le tatouage témoin (*Telltale tamper-proofing watermarking*), est un tatouage qui permet :

- d'indiquer avec une grande probabilité si une modification donnée s'est produite ou non ;
- de donner une mesure de la distorsion ;
- de caractériser le type de distorsion ;
- de valider le signal sans autre information secondaire.

Il est introduit par Kundur et Hatzinakos qui l'opposent au tatouage standard (dit *tamper-proofing watermarking*) qui ne permet que de déterminer avec une grande probabilité et sans l'usage d'autres sources, si une image a subi une modification. La différence majeure (comme expliqué dans [Hassan 2006]) consiste à déterminer le type de modification qu'un média a subi, et en déduire si elle est ou non légitime.

Tout tatouage est effectué en modifiant l'hôte. Cependant, ce dernier peut se présenter sous différentes formes, équivalentes en terme d'information, mais différentes en terme de présentation. Chacune de ces formes a des caractéristiques qui lui sont propres, et qui permettent de tatouer d'une certaine manière.

_ Définition 6, Domaine d'insertion
Le domaine d'insertion désigne la forme de l'hôte qui est utilisée pour insérer la marque.

Dans le cas d'une image, on distingue essentiellement deux formes. La forme « spatiale », comme un bitmap, est une représentation visuelle traditionnelle. L'image est un tableau tridimensionnel de pixels. Les deux premières dimensions définissent la position du pixel. La troisième dimension contient les informations colorimétriques (dépendantes du modèle utilisé). Un exemple de tatouage fonctionnant dans ce domaine est l'algorithme de Mintzer et Yeung [Mintzer 1999] (décrit plus loin dans l'état de l'art) dont Fridrich, Goljab et Memon ont fait l'analyse dans [Fridrich 2002b]. La seconde forme possible d'une image est la forme "fréquentielle". Différentes transformations numériques comme la transformée de Fourier, la DCT (Discrete Cosinus Transform), ou encore les ondelettes, permettent de transformer la représentation spatiale d'une image en une représentation fréquentielle. L'intérêt de la représentation fréquentielle est de mettre en évidence des composantes périodiques, plus favorables à la compression. Lin et Chang ont utilisé ce domaine pour insérer leur marque dans [Lin 2000a]. Une marque peut aussi être insérée dans le domaine des ondelettes comme dans [Kundur 1999]. C'est un domaine utilisé dans le format JPEG2000. Pour plus d'information sur les autres domaines, nous vous invitons à consulter les articles de traitement du signal qui y font référence ([Bas 2000]).

Tous les tatouages possèdent des fonctions, des qualités, et des faiblesses diverses. Voyons quelles sont les propriétés applicables aux tatouages.

2.2.2 Propriétés du tatouage

Pour la gestion des droits d'auteur, le tatouage appose dans le *media* un identifiant de l'auteur. Cet identifiant doit rester lisible au maximum, et ce même si l'hôte est modifié. En revanche, pour assurer l'authenticité du média, c'est-à-dire pour certifier qu'il est original ou une copie conforme de l'original, le tatouage est conçu pour se détruire à la moindre modification. Ces deux exemples extrêmes correspondent à la même caractéristique que nous désignerons par robustesse. La robustesse connait de multiples définitions ([Atupelage 2008, Cayre 2005, Furon 2005, Lin 2000b, Özgür Ekici 2004, Rey 2000]. Elles sont rarement identiques d'un article à l'autre. Nous proposons donc ici une synthèse de ces définitions.

_ Définition 7, Robustesse
La robustesse d'un tatouage caractérise sa capacité à résister aux modifications de l'hôte.

Ce caractère se décline généralement en trois degrés : robuste, semi-fragile et fragile. Un tatouage robuste correspond au degré le plus élevé. Par

exemple, Rey et Dugelay utilisent un algorithme de ce type dans [Rey 2000] pour distinguer les manipulations malicieuses d'une image. En revanche, un tatouage fragile correspond au degré de robustesse le plus faible. La moindre modification de l'hôte efface la marque. Par exemple, Wong [Wong 1998] se base sur une technique de tatouage fragile associé à une clé asymétrique pour vérifier l'authenticité et l'intégrité d'une image. Le tatouage semi-fragile est un degré intermédiaire entre ces deux extrêmes. Ce type de tatouage est robuste à un ensemble défini de manipulations et est fragile à d'autres. Par exemple, dans [Lin 2000a], les auteurs développent un tatouage embarquant une signature numérique pour authentifier une image. Dans leurs objectifs, l'image doit pouvoir subir un traitement de compression JPEG. En revanche, l'authentification nécessite d'avoir une détection fiable des modifications. C'est la perte de la signature qui révèle ces modifications. Ce tatouage est robuste à la compression JPEG, mais fragile pour toute autre modification.

On trouve, dans la littérature, des propriétés similaires. Elles reprennent en partie les concepts déjà évoqués ci-dessus. Par exemple dans [Atupelage 2008, Rey 2002] on introduit la sensibilité.

_ **Synonyme, Sensibilité**
La sensibilité d'un tatouage définit la façon dont il réagit face à une modification malicieuse.

Une autre formulation est donnée dans [Rey 2002].

_ **Synonyme, Tolérance**
La tolérance d'un tatouage définit sa conservation face à des modifications souhaitées comme la compression.

Ces termes sont parfois utilisés lorsque l'on traite de tatouage semi-fragile. Ils permettent de faire la différence entre les modifications qui participent au tatouage, et celles qui sont des attaques sur ce tatouage. On pourra se référer à [Atupelage 2008] qui traite ces deux points.

Par définition, le tatouage consiste à insérer une nouvelle information directement dans l'hôte, c'est-à-dire à modifier les informations elles-mêmes qui constituent l'hôte. Dans le domaine du tatouage de données multimédia (son, image, vidéo), la préservation de la qualité originale de l'hôte devient un aspect fondamental. Cette caractéristique est désignée par le terme d'imperceptibilité. Le terme d'imperceptibilité apparaît dans [Fei 2006, Kundur 1999, Lin 2000b], et des articles comme [Fridrich 1998a] en font un pré-requis.

_ **Définition 8, Imperceptibilité**
L'imperceptibilité du tatouage est le caractère qui permet d'affirmer qu'une image n'a pas été dégradée visuellement par l'insertion de la marque.

La valeur de cette caractéristique peut être très variable d'une application à l'autre. A une extrémité, l'illustration de la *perceptibilité* (sans intérêt dans

notre étude) est l'incrustation volontairement visuelle. Il s'agit, dans ce cas, de garantir la promotion d'une marque, ou d'un site par exemple. La sécurité réside alors dans l'impossibilité de reconstruire l'image originale lorsque la marque est supprimée. A l'autre extrémité, on peut citer les applications de type médical, dans lesquels on ne peut se permettre de fournir des médias dégradés. Un tatouage peut néanmoins y figurer s'il est imperceptible. Sachant qu'une imperceptibilité totale est impossible (ce qui reviendrait à ne pas modifier l'hôte, donc à ne rien insérer), il convient alors de définir un seuil à partir duquel on peut considérer la modification comme *imperceptible*. De plus, un tatouage peut être imperceptible de diverses manières. Soit parce que l'œil humain ne le voit pas (dans le cas d'une image), soit parce qu'une machine ne peut pas deviner sa présence. Une façon objective de mesurer l'imperceptibilité se fait par la détermination du *Peak Signal-to-Noise Ratio* (PSNR), [Petitcolas. 1999].

L'impact visuel est mesuré par le PSNR (Peak Signal to Noise Ratio). Cette mesure exprime le rapport entre la puissance maximale d'un signal et la puissance du bruit de corruption affectant la fidélité de la représentation.

Dans le domaine du traitement du signal, on estime qu'un bon PSNR se situe au-dessus de 38dB. Par exemple, le tatouage proposé par Kundur et Hatzinkos [Kundur 1999] permet d'obtenir un PSNR de 43dB.

La formule du PSNR est donnée ci-dessous (Équation 2.6)

$$PSNR = 10 \cdot \log_{10}\left(\frac{d^2}{EQM}\right) \tag{2.6}$$

$$EQM = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}||I_o(i,j) - I_r(i,j)||^2 \tag{2.7}$$

$$I_o \text{ image originale, } I_r \text{ image tatouée.} \tag{2.8}$$

Dans le cas de l'enrichissement de document, ou d'indexation, le plus important est d'avoir un tatouage qui permette d'insérer un grand nombre d'informations.

_ Définition 9, Taux d'insertion
Le taux d'insertion est la quantité d'informations que l'on peut tatouer dans une image sans dégrader sa qualité. On parle aussi de capacité.

[Petitcolas. 1999] indique qu'un taux d'insertion minimum est de 70 bits. Dans une méthode de stéganographie, Fridrich et Soukal [Fridrich 2006] ont mis en œuvre la technique de *matrix embedding*, pour augmenter la capacité d'une image à contenir un message. Cette méthode est aussi applicable au tatouage comme nous le montrerons dans nos travaux.

Pour des applications embarquées, mobiles, caméras de surveillance, les algorithmes doivent respecter des contraintes liées au matériel, et notamment pour des applications en temps-réel. Des cas précis sont abordés dans [Atupelage 2008, Fei 2006].

_ Définition 10, Complexité
La complexité d'un tatouage peut s'exprimer comme la complexité de l'algorithme qui sert à l'insertion et/ou à la détection de la marque.

Comme indiqué par [Hartung 1999], un tatouage en temps-réel nécessite une faible complexité, et donc devra (en général) travailler dans un domaine compressé. Fridrich a conçu un algorithme de tatouage à faible complexité dans [Fridrich 1998a]. Ce tatouage est dédié à l'utilisation sur caméra de surveillance.

Deux principes participent à améliorer les résultats de certains cas d'utilisation : la localisation, qui permet de savoir avec un certain grain quelle partie du média est modifiée, et la reconstruction qui, à partir des informations embarquées, est en mesure de restaurer une dégradation. Cette notion trouve une définition dans [Atupelage 2008, Lin 2000b].

_ Définition 11, Localisation
La localisation est la possibilité que peut offrir une marque extraite, de déterminer quelles sont les zones de l'images dont les modifications ont endommagé cette marque.

Cette propriété est en générale associée à un tatouage semi-fragile ou robuste. En effet, un tatouage fragile se devant d'être sensible à toutes les modifications imaginables, il n'est pas pertinent de chercher à localiser. La moindre modification cassant la marque, il ne sera pas possible d'avoir une plus grande granularité. En revanche, disposer d'une certaine latitude quant à la robustesse du tatouage, permet de lier la marque à des positions connues de l'image. Ainsi, lors de l'extraction de la marque, il est possible de déterminer les régions où se trouvent des erreurs. Fridrich analyse ces types de tatouages et propose un nouvel algorithme dans [Fridrich 2002a]

_ Définition 12, Reconstruction
La reconstruction est la possibilité offerte par une marque de régénérer une altération de celle-ci.

Ce type de tatouage marque l'image avec une version dégradée d'elle même. Ces informations mises en corrélation avec l'image permettent, en cas de dégradation, d'estimer une correction. Hassan et Gilani [Hassan 2006] proposent une technique d'authentification d'image avec reconstruction. Les informations sont stockées dans des macro-blocs de 2x2 pour l'authentification, alors que les blocs périphériques embarquent l'information de reconstruction.

2.2.3 Propriétés de sécurité du tatouage

Nous venons de voir que les algorithmes de tatouages peuvent avoir de multiples propriétés. On pourrait penser que la sécurité d'un tatouage se réfère à sa robustesse, mais ce n'est pas le cas. Suivant les définitions fournies

par Kalker dans [Kalker 2001], et reprises dans le projet européen Certimark [UCL 2002], la robustesse doit être considérée avec une importance moindre que la sécurité.

Kalker (et [Cayre 2005, Furon 2005] par la suite) propose une définition d'un tatouage robuste, en indiquant qu'il s'agit d'un mécanisme pour créer un canal de communication qui est multiplexé dans le contenu original. Pour justifier la terminologie de robuste, il ajoute que ce tatouage doit répondre à deux critères :

- La dégradation perceptible, induite par l'ajout de la marque, est minimale ;
- La capacité du canal de communication diminue selon une fonction « lisse » plus le contenu marqué est dégradé.

Pour expliciter ce dernier point, il faut comprendre que la quantité d'informations recueillies dans un document tatoué et altéré suit une fonction lisse décroissante qui est liée à la dégradation (fonction « lisse » : infiniment dérivables sur l'ensemble des réels - voir les classes de régularité des fonctions pour une définition plus complète). Cette définition correspond donc à un tatouage aveugle robuste selon nos définitions. On notera cependant les détails supplémentaires que donne Kalker concernant la manière dont la marque se doit de disparaître lors des modifications.

Concernant la sécurité du tatouage, Kalker nous donne la définition suivante :

_ **Définition 13, Sécurité**
La sécurité du tatouage est l'impossibilité pour un utilisateur non autorisé d'accéder au canal de communication.

Cela veut dire, qu'un tatouage sécurisé n'autorise pas une personne non initiée à accéder à la marque. Le projet Certimark précise ces définitions, en donnant le périmètre d'application. La sécurité traite à la fois de l'effacement de la marque (la robustesse), de l'insertion frauduleuse et de la détection non autorisée (présence, et lecture).

Une définition alternative est celle de l'*accès*.

_ **Synonyme, Accès**
L'accès à la marque constitue la possibilité d'insérer, de modifier, de copier, de lire et de détecter cette marque.

Une remarque importante est que la robustesse traite d'attaque pour lesquelles on ne sait pas si la modification est volontaire ou non. En général il s'agit de modification propre au domaine du traitement d'image, comme la découpe, la redimension ou le filtrage. Cela ne dépend pas du tatouage. En revanche la sécurité est testée par des attaques précises, et ciblée sur l'algorithme utilisé. À ce titre, il faut supposer que l'attaquant a une très bonne connaissance de l'algorithme.

Cela nous amène à dire qu'un tatouage sécurisé possède des mécanismes de protection contre :

- La lecture ;
- La modification ;
- La copie ;
- L'effacement
- La détection.

Ces points peuvent être résolus par différentes techniques, comme l'utilisation de clés. Voyons cela plus en détail. La confidentialité impose de faire un choix technique nous permettant de rendre le message inintelligible. Il ne faut pas oublier que l'attaquant connait notre algorithme. Les méthodes de cryptage actuelles sont valides. Ainsi, on peut mentionner [Atupelage 2008] où les auteurs font usage d'une PKI pour intégrer leur marque. Leur idée est d'utiliser le principe de signature pour résoudre des problèmes d'authentification, mais la littérature en sécurité nous indique aussi qu'une PKI permet de réaliser des chiffrements asymétriques. Ainsi on peut protéger une marque grâce à un cryptage fort. D'autres techniques comme celle présentée par [Eggers 2001] utilisent un cryptage par clé qui permet de faire varier l'insertion. Par la même méthode, on obtient une protection contre la lecture et la détection. On appellera cette sécurité la protection de la marque.

Empêcher la copie est généralement bien traité lorsque l'on ajoute à l'algorithme un système pour le faire dépendre du signal hôte. Toute copie invalide alors la relecture de la marque. On pourra regarder à titre d'exemple le schéma de Walton [Walton 1995].

Il reste un dernier élément à prendre en compte : les clés secrètes de l'algorithme. En effet, on sait depuis [Kerckhoffs 1883] que la sécurité d'un algorithme ne peut pas être basée sur le secret de l'algorithme lui-même (tout du moins sur le long terme), et qu'il est indispensable d'utiliser un mécanisme de clés qui ne peuvent être découvertes en un temps raisonnable. C'est un principe qui, comme il est rappelé dans [Cayre 2005], s'applique aussi au tatouage.

2.2.4 Classification des attaques du tatouage

Dans cette sous-section nous allons tenter de classifier les attaques qui peuvent être appliquées au tatouage. Nous commencerons par analyser des classifications existantes, puis nous verrons qu'il nous faut proposer une nouvelle classification pour tenir compte du contexte.

2.2.4.1 Classifications existantes

Une classification importante est donnée par Kutter et Voloshynovskiy dans [Kutter 2000]. Dans ce document les auteurs présentent quatre types d'attaques (voir la figure 2.6).

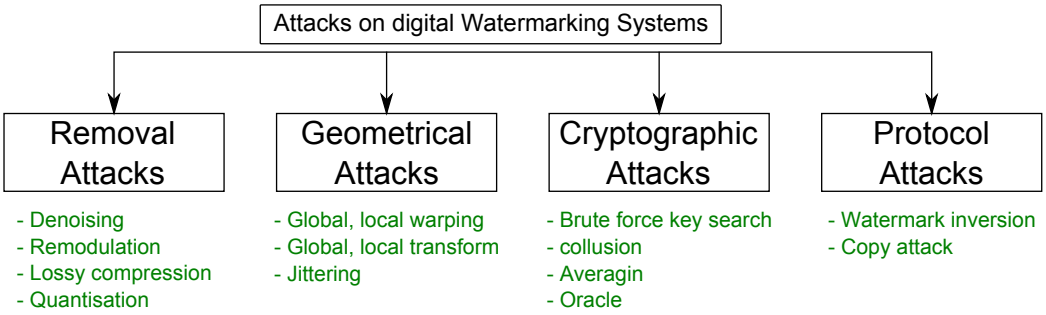


Figure 2.6 – Types of attacks on digital watermarking systems - Image de [Kutter 2000]

Le premier de ces types d'attaque est constitué par les attaques de suppression. Elles visent à enlever la marque contenue dans un hôte. Globalement pour parvenir à un effacement total de la marque, celle-ci est considérée comme un bruit avec ses statistiques. Les attaques tentent alors d'estimer l'original en retirant ce bruit.

Le second type d'attaques, plus faciles, correspond aux attaques géométriques. Leur objectif est la distorsion de la marque, ou la désynchronisation, avec comme résultat le fait que le lecteur n'est plus capable de relire la marque, même si celle-ci est toujours là. On pourra prendre comme exemple les rotations d'images.

Le troisième type d'attaque est la cryptographie, qui propose une vaste diversité de méthodes. L'objectif de ces attaques est de découvrir des informations sur les paramètres secrets comme les clefs.

Finalement, on trouve les attaques de protocole. Elles n'essayent ni de supprimer ni d'altérer la marque, mais cherchent à la manipuler par exemple en inversant la marque, en la copiant, ou en accédant au codeur.

Plus récemment dans [Fridrich 2002a], l'auteur présente une classification différente qui prend en compte les objectifs de l'attaquant. Fridrich avance ainsi trois types de scénarios.

Le premier type est : la modification non détectée : l'attaquant cherche à faire un changement de l'image avec une probabilité raisonnable pour que le détecteur ne voit pas le changement.

Le second type est : la collecte d'informations : l'objectif est d'obtenir des informations sur les clés secrètes, de placer des informations dans les blocs, ou encore de trouver des motifs de synchronisation, des *Look Up Table*...

Et le dernier type est la faille du protocole : lorsque l'algorithme n'est pas vulnérable en soi, la manière de l'utiliser peut le rendre attaquable. [Rey 2002] ajoute à ces trois intentions, celle de créer une nouvelle marque qui sera considérée comme authentique. Elle est à rapprocher de la modification non détectée, mais sans perdre de vue qu'ici on cherche à choisir l'hôte de la marque.

Dans la suite de la classification, Fridrich donne les conditions dans lesquelles peuvent se réaliser les attaques. [Fridrich 2002a] en dénombre cinq :

1. L'attaque s'effectue sur l'image tatouée (*stego-image attack*), l'attaquant possède une image authentifiée et veut faire un changement sans compromettre l'authenticité, ou récupérer une information pour être en mesure de reproduire le tatouage.
2. L'attaque se fait sur de multiples images tatouées (*multiple stego-image attack*), même principe avec plusieurs images.
3. L'attaque cible le vérificateur (*Verification device attack*, ou oracle), l'attaquant a accès au système de vérification et peut donc vérifier l'authenticité de n'importe quelle image. La sortie peut être un oui/non, ou la liste des blocs compromis.
4. L'attaque peut se faire sur l'hôte (*Cover-image attack*), l'attaquant dispose alors de paires d'images originales et authentifiées.
5. L'attaque peut être menée à partir d'hôtes choisis (*Chosen cover-image attack*), l'attaquant peut authentifier l'image qu'il souhaite.

[Rey 2002] donne des exemples d'attaques :

- L'attaque par copie, on copie simplement la zone de tatouage sur une autre image.
- L'attaque par collage [Fridrich 2002a] où l'on créé une image en collant des blocs provenant d'une base d'images authentifiées.
- Et l'attaque force-brute, qui est la recherche exhaustive des clés.

[Furon 2005] classifie les attaques à la manière de la classification cryptographique :

- Attaques à marque choisie : on fournit au codeur la marque qu'il va embarquer dans un hôte donné.
- Attaque à original choisi : ici c'est l'hôte qui est fournit au codeur.
- Attaque « original unique, marques multiples », le même original est tatoué plusieurs fois.
- Attaque « originaux multiples, marques multiples », qui s'inspire de la même idée mais pour laquelle on a accès à plusieurs originaux.

Ces classifications sont intéressantes pour avoir une vision globale des attaques possibles. Mais on remarque qu'elles ne tiennent pas compte du contexte. Or, selon le contexte, l'information à incruster, l'information à protéger ou les usages sont très variables. Ainsi, connaître l'impact de l'attaque et le gain potentiel de l'attaquant est important pour orienter l'effort à consacrer à la protection. Pour cela nous proposons une classification contextuelle, qui fait apparaître les attaques par famille de risques.

2.2.4.2 Classification contextuelle des attaques sur le tatouage

Nous rappelons ici que le contexte est celui de la construction de la preuve photographique. Une photographie capturée depuis un appareil mobile doit intégrer suffisamment d'informations pour que l'on puisse attribuer à tout moment la propriété de la photographie à son auteur, et vérifier que cette photographie n'a fait l'objet d'aucune modification suspecte.

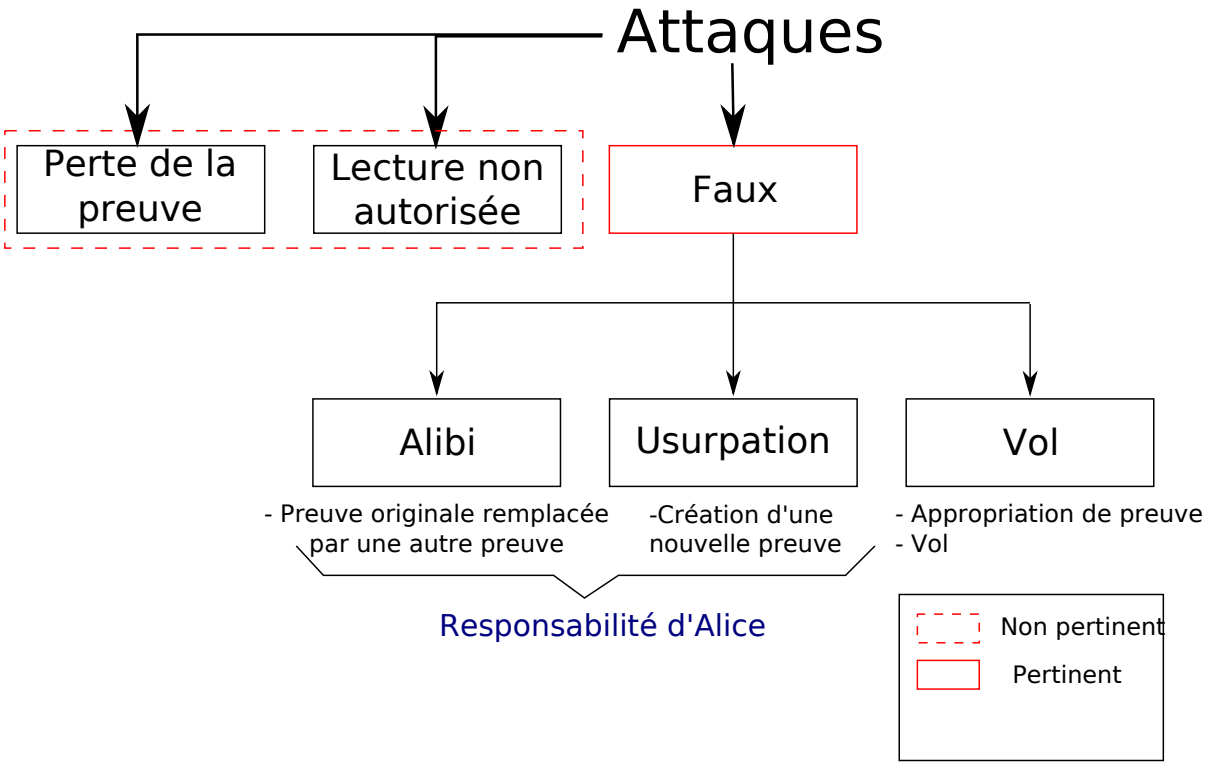


Figure 2.7 – Classification des attaques

Le premier niveau présente trois familles de risques.

1. Tout d'abord, l'attaquant peut chercher à détruire la preuve que le tatouage embarque. L'attaquant dispose alors de l'image sans les éléments qui en font un original probant. Cela n'est pas pertinent du point de vue du tatouage. En effet, si l'attaquant dispose d'une copie conforme (tatouée) de l'original, et qu'il fait disparaître le tatouage sur cette copie, il n'en reste pas moins que l'auteur continue à posséder l'original qui, lui, reste probant. Si l'attaquant peut supprimer la marque de l'original lui-même, alors le problème n'est pas uniquement celui de la qualité du tatouage, mais aussi celui de la conservation sécurisée de l'original : le coffre-fort numérique. Ce point est évidemment crucial dans la chaine de certification, mais n'est pas l'objet de ce chapitre. Les attaques par suppression,

par débruitage ou encore les attaques géométriques ne sont pas une cible pour la conception d'un tatouage de certification.

2. La deuxième famille de risques est l'accès non autorisé aux informations tatouées. Il s'agit pour l'attaquant de lire, par exemple, les coordonnées géographiques de l'image ou son horodatage. Là encore, ces attaques ne relèvent pas de la qualité du tatouage proprement dite car, même si la connaissance précise de ces informations n'est pas souhaitée, la preuve reste intacte, et donc ne perd pas sa valeur. On peut noter que le chiffrement de ces informations résout ce problème.
3. En revanche, la dernière famille, la création de faux, nous intéresse particulièrement. C'est même un point critique, car si l'on peut réaliser de telles attaques, la validité du système tout entier est menacée. Cette famille peut elle-même se décomposer en trois objectifs, le faux alibi, l'imitation de signature ou le vol. Ces termes, que nous allons définir, peuvent apparaître à certains endroits restrictifs sur les perspectives de l'attaquant. Leur signification usuelle illustrent cependant bien les intentions de l'attaquant.

Pour cerner précisément les différentes sortes de faux et les litiges qu'ils peuvent produire, nous définissons plusieurs fonctions reflétant le contenu sémantique de la marque.

- TS est la fonction d'horodatage. Cette fonction possède une relation d'ordre sur ses valeurs. Pour deux images x et y, si x est antérieure à y, alors $TS(x) \leq TS(y)$ est vrai.
- Geo est la fonction de géolocalisation. Cette fonction ne possède qu'une relation d'équivalence. Si deux images x et y ont été capturées au même lieu géographique, alors $Geo(x) = Geo(y)$ est vrai.
- $Auth$ est la fonction d'authentification de l'utilisateur. Si l'image x appartient à l'utilisateur u, alors $Auth(x, u)$ est vrai.

A ces trois fonctions s'ajoute une quatrième fonction implicite dans tous les scénarios. C'est la fonction d'intégrité qui doit systématiquement être satisfaite. Si l'intégrité n'est pas vérifiée, l'attaque n'en est pas une car la valeur de l'image obtenue par l'attaquant est nulle.

Le *vol* d'une preuve se définit par la production d'un document certifié pour s'approprier la propriété de la preuve d'autrui. Le vol est à distinguer de l'atteinte au droit d'auteur, dans le sens où l'attaquant s'approprie ici une preuve et non une œuvre. Autrement dit, le vol consiste pour l'attaquant à apposer sa propre signature sur une preuve dont il n'est pas le propriétaire. Supposons qu'Alice prenne une photographie d'un scoop, et la certifie dans le but d'en faire valoir ses droits. Oscar, qui n'était pas sur place, copie cette photographie et la signe avec sa propre signature. Il est sous-entendu ici, que cette photo

signée par Oscar ne contient plus la signature d'Alice. Chacun possède alors un document probant. Pour réussir cette attaque il faut :

- Créer une marque à son nom sur un « original » qui ne nous appartient pas, et qu'elle soit valide ;
- Que la marque de temps associée au document soit antérieure à celle du propriétaire légitime.

Pour réaliser cette attaque Oscar doit :

- Copier l'original d'Alice ;
- Faire disparaître la signature d'Alice (formule 2.9) ;
- Apposer sa propre signature (formules 2.10) ;
- L'horodatage associé au document d'Oscar doit être antérieur à celui d'Alice (formules 2.11) ;

On peut le formaliser ainsi :

$$Auth(Image_{Oscar}, "Alice") = Faux \tag{2.9}$$
$$Auth(Image_{Oscar}, "Oscar") = Vrai \tag{2.10}$$
$$TS(Image_{Alice}) > TS(Image_{Oscar}) \tag{2.11}$$

Résoudre le conflit nécessite de s'appuyer sur la marque de temps associée aux signatures respectives.

Éviter ce type d'attaque nécessite de protéger les deux aspects, l'authentification de l'auteur de l'image et l'instant de la capture de l'image. C'est essentiellement ce dernier point, l'antériorité, qui pose réellement un problème et qui nécessite, dans le processus de certification, un traitement approprié. Mais l'horodatage sûr a déjà fait l'objet de travaux approfondis, et des protocoles fiables existent ([Adams 2001, Bonnecaze 2006, ETSI 2006, Haber 1991, Pinkas 2003]). L'authentification de l'horodatage sera décrit plus en détails dans le chapitre A. Dans le contexte de la certification, nous posons comme pré-requis qu'un tiers de confiance possédera toujours l'image originale dûment authentifiée. Il n'est donc pas nécessaire de garantir que l'authentification des copies soit particulièrement robuste. En cas de litige, l'original, celui qui est conservé par le tiers de confiance, fera toujours foi.

La création d'un *alibi* est la création d'un faux qui appartient à l'auteur légitime, mais translaté, soit dans le temps (pour prouver autre chose en un lieu donné), soit dans l'espace (pour prouver autre chose à un instant donné).

Pour réaliser un alibi, Alice doit dans un premier temps, créer une marque à son nom sur un original choisi, éventuellement certifié mais avec d'autres informations (formule 2.12). Cette condition est commune aux différents types d'alibi.

$$Auth(Image_{Alice}, "Alice") = True \tag{2.12}$$

Dans un deuxième temps, Alice doit pouvoir modifier soit : l'horodatage, soit la géolocalisation (formules 2.13 ou 2.14).

$$TS(Image_{Alice}) \neq TS(Alibi) \tag{2.13}$$
$$Geo(Image_{Alice}) \neq Geo(Alibi) \tag{2.14}$$

Comme nous l'avons déjà évoqué précédemment, des solutions techniques fiables existent pour s'assurer de la validité et de l'authenticité de l'horodatage. En revanche, la géolocalisation reste aujourd'hui une information peu sûre. Les moyens de brouillage ou d'usurpation sont relativement faciles à mettre en œuvre. Forger ce type d'alibi (faire croire qu'à un instant donné, l'image a été prise ailleurs) est possible. Le chapitre 3 est entièrement consacré à ce problème.

L'*imitation* consiste à produire un document certifié qui ne nous appartient pas. Pour réaliser cette attaque, Oscar doit créer une marque au nom d'Alice à partir d'un document qu'elle n'a pas certifié, et que cette marque soit valide.

$$Auth(Image_{Oscar}, "Alice") = True \tag{2.15}$$

Dans cette attaque, le système d'authentification est le point essentiel à protéger. Le protocole de certification ne doit divulguer aucune information sur les éléments qui permettent à Alice d'authentifier ses images. Les systèmes de signature actuels offrent des qualités suffisantes pour limiter les risques liés à ce type d'attaque.

Pour réaliser ces attaques, nous avons identifié quatre méthodes :

- Attaque par force brute ;
- Attaque par écrasement ;
- Attaque statistique ;
- Duplication.

Attaque par force brute : Il s'agit s'assurer que le cardinal de l'ensemble des paramètres secrets est assez élevé pour rendre extrêmement difficile une recherche séquentielle de ces paramètres. Typiquement, disposer d'une clé de 128 bits impose à l'attaquant de tester 2^{128} possibilités.

Attaque par écrasement : Le principe est d'utiliser la méthode de tatouage sur une image déjà tatouée. On ne prend aucune précaution particulière pour sauvegarder la marque précédente, ou pour la faire disparaître, on espère que recopier une nouvelle marque éliminera l'ancienne.

Attaque statistique : Il s'agit de considérer les éventuelles failles pouvant être exploitées lorsque l'attaquant dispose d'un ensemble assez grand d'images tatouées. Nous montrerons, par la suite, tous les dispositifs mis en place pour injecter de la variété dans les règles d'insertion afin de minimiser les corrélations que l'on appelle "transversales". (Figure 2.8)

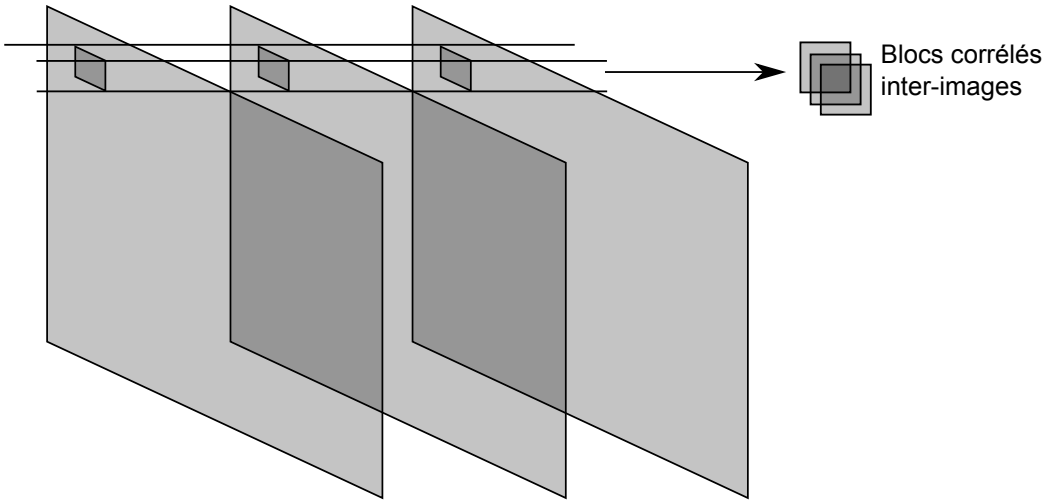


Figure 2.8 – Corrélations transversales entre mêmes blocs d'images distinctes

Duplication : Cette attaque concerne la possibilité de dupliquer la marque dans une autre image sans nécessairement être capable de décoder le message (sans connaissance des clefs). Il s'agit ici d'une attaque avec une connaissance partielle, voire faible, de l'algorithme. Il s'agit d'un problème crucial dans le domaine du tatouage de certification : on ne doit pas être capable de certifier une image par simple recopie du message provenant d'une autre image déjà certifiée.

Détaillons ce scénario. Le principe employé pour effectuer l'insertion se base sur la modification des LSB[3] LSB) des données suivant une règle bien déterminée. La modification de ces LSB engendre une dégradation limitée de l'image, par conséquent, dès lors que l'attaquant connait la nature des composants de l'image qui ont été tatoués, il peut aisément décalquer les LSB sur une autre image (Figure 2.9). Pour éviter ce problème de duplication il est indispensable de faire dépendre la règle d'insertion du contenu de l'image. Ainsi un même message sera inséré de deux manières différentes dans deux images distinctes.

Par la suite nous nommerons « caractéristique » cette dépendance au contenu. Finalement, afin d'être en adéquation avec les principes de Kerckhoffs [Kerckhoffs 1883], cette caractéristique devra dépendre d'un paramètre secret

3. Détaillé dans la section 2.4.2.1

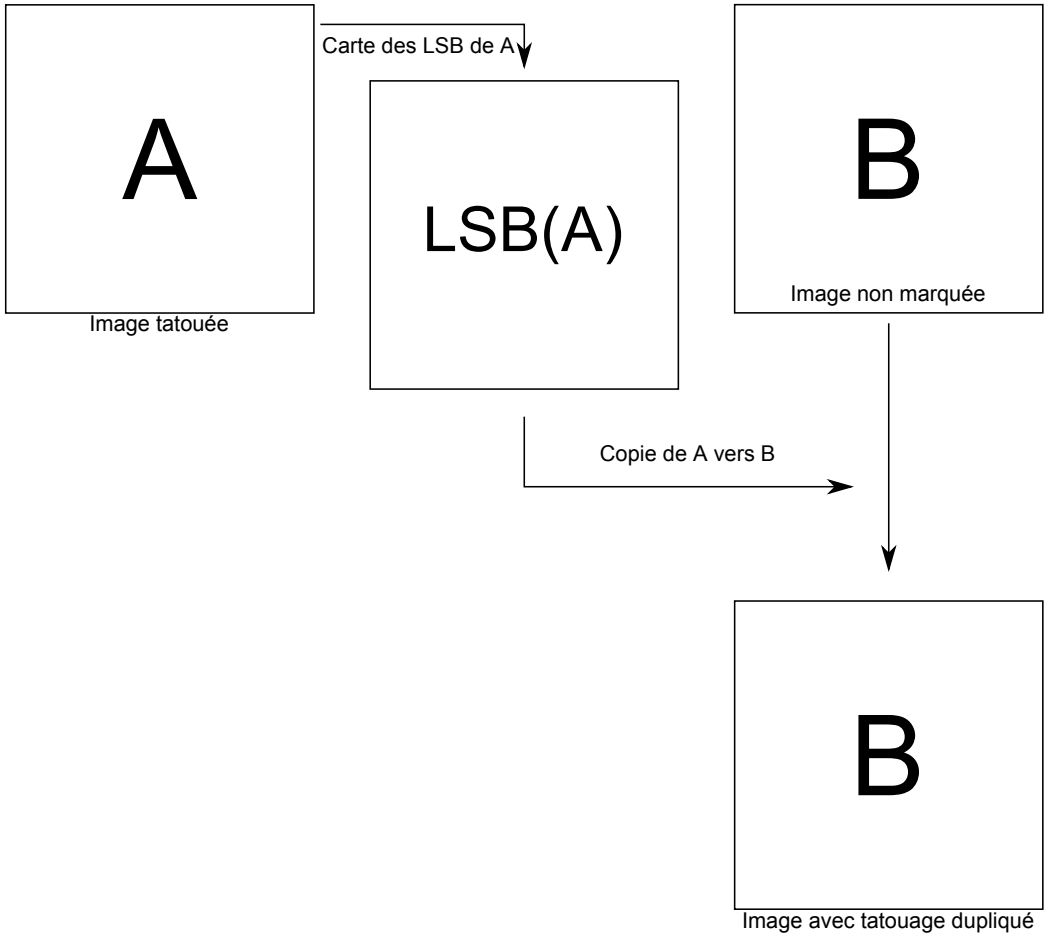


Figure 2.9 – Décalque d'une carte de LSB d'une image certifiée vers une image non certifiée.

afin de rendre quasiment impossible la duplication du message sans connaissance des paramètres secrets. (Figure 2.10)

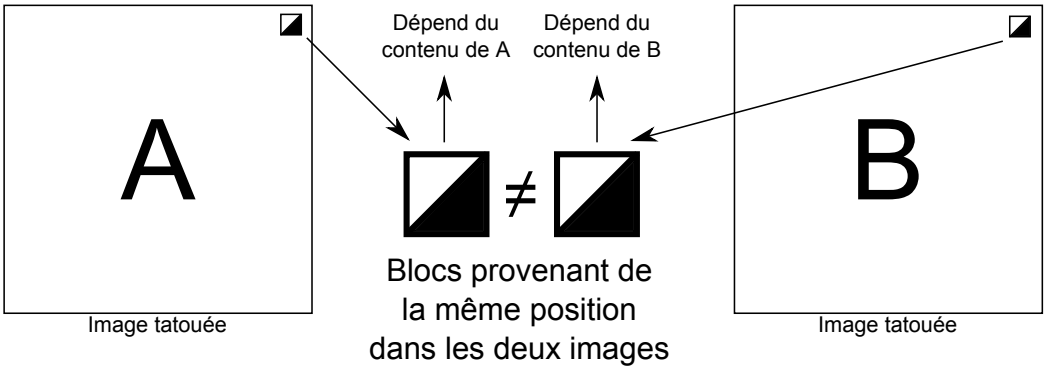


Figure 2.10 – Dépendance du tatouage avec le contenu de l'image

2.3 Tatouage de certification

Nous avons vu les propriétés que peut avoir un tatouage. Nous avons aussi vu quelles sont les attaques qu'il peut subir. Maintenant nous allons déterminer quelles sont les propriétés qu'un tatouage doit avoir pour resister à ces attaques.

2.3.1 Définition du tatouage de certification

Le tatouage de certification a pour objectif d'apporter la preuve qu'une image n'a pas été modifiée depuis l'insertion de la marque, et doit de plus intégrer les notions introduites dans la section 1.3 concernant la notion de preuve. On souhaite un algorithme de tatouage qui permette :

- d'authentifier l'utilisateur
- de contrôler l'intégrité du document
- d'embarquer les informations nécessaires à la création d'une preuve.
- de localiser les endroits qui ont perdu leur intégrité
- d'être exploitable par un mobile.
- de réaliser certaines manipulations, notamment la compression JPEG

Les quatre premiers points constituent le tatouage de certification, le cinquième est une contrainte technique. Enfin, le dernier point est une contrainte imposée par le contexte industriel de CodaSystem.

Les propriétés que requiert le tatouage de certification sont donc les suivantes :

- la fragilité, pour garantir l'effacement de la marque lorsque l'on essaie de modifier l'image.

– la robustesse, pour résister à la compression qui sera nécessaire pour les utilisateurs.

Par conséquent, ce tatouage devra être **semi-fragile**. Pour répondre aux autres demandes, nous proposons de définir le tatouage de certification par les points suivants.

La **sensibilité aux modifications** (contrôle d'intégrité) est nécessaire dans toute application d'authentification. Du fait de la nécessité de la robustesse à la compression, et de la propriété de localisation recherchée, elle ne peut pas être aussi sensible que les méthodes cryptographiques comme les fonctions de hashage. En effet, avec une fonction de hashage, la modification d'un seul bit de l'image compromet l'intégrité de toute l'image. Cette relaxation de l'intégrité permet de retrouver l'approche traditionnelle d'une photocopie conforme. En effet, le procédé de photocopie introduit de petites dégradations (saleté, variation des couleurs, déformations, etc.) qui ne remettent pas en cause la qualité de copie conforme. Au niveau du tatouage, cela est mis en évidence par la localisation des défauts. La détection de ces zones non intègres offre la possibilité de faire la différence entre une éventuelle falsification de l'image et une simple erreur de ré-encodage par exemple. Cependant, elle doit être suffisante pour ne pas autoriser de modifications majeures de l'image.

Une **dépendance à l'image** sera intégrée à ce tatouage, il ne faut pas en effet que l'on puisse dupliquer la marque d'une image certifiée à une autre image quelconque pour rendre cette dernière probante.

Une **protection de la marque** doit aussi être mise en œuvre : d'une part pour conserver la confidentialité et d'autre part pour éviter de potentielles attaques à texte connu.

L'imperceptibilité n'est paradoxalement pas une caractéristique fondamentale du tatouage même si celle-ci reste importante. En effet, si elle est appréciable pour l'utilisateur qui visualise l'image, une dégradation contrôlée est possible, voire souhaitable si elle apporte de la valeur probante. En effet, pour insérer des informations relatives à la photographie, il nous faudra nécessairement modifier l'image.

Enfin, la **capacité** du tatouage doit être suffisante pour embarquer l'ensemble des informations de la preuve.

En revanche, d'autres propriétés ne sont pas pertinentes pour la certification. Par exemple, la reconstruction des zones altérées ne constitue pas un point crucial, requise par [Lin 2000a] elle reste optionnelle pour [Rey 2002]. Cette dernière opinion peut être soutenue par le fait que, si l'on suppose la reconstruction possible, on ne peut y accorder de crédit dans le cadre de l'authentification et de la preuve ; en effet on ne peut pas garantir que la reconstruction restaure une version digne de confiance. Lin et Chang dans [Lin 2000a], définissent la notion de portabilité. Cette propriété indique que l'authentification peut être faite à partir du contenu reçu. Il n'est pas évident de lire sous cette appellation l'idée de tatouage aveugle. Pour dissiper l'ambiguïté, le tatouage de certification aura besoin d'être un tatouage aveugle pouvant nécessiter l'utili-

sation de clés (qui sont externes au contenu reçu).

En résumé, voici les critères que nous proposons pour un algorithme de tatouage de certification :

- robustesse aux distorsions inhérentes à l'application, (Un tatouage JPEG doit résister à la compression JPEG)
- sensibilité (validation de l'intégrité et localisation),
- protection de la marque,
- imperceptibilité,
- capacité à embarquer les informations nécessaires à la qualification de preuve,
- complexité limitée (contexte mobile).

Partant de ces propriétés nous allons maintenant étudier les différentes méthodes disponibles dans la littérature.

2.3.2 Revue littéraire des tatouages semi-fragiles

Notre recherche bibliographique s'est focalisée sur les algorithmes de tatouage semi-fragiles, car ce sont ceux qui sont les plus appropriés. Nous cherchons maintenant à comprendre les différents algorithmes et à en extraire les propriétés nécessaires ou intéressantes pour le tatouage de certification. Nous allons détailler les méthodes en fonctions des propriétés qu'elles fournissent : contrôles d'intégrité, techniques de localisation, éléments de sécurité, robustesse et capacité.

2.3.2.1 Gestion de l'intégrité

[Hassan 2006] L'idée est de s'inspirer des algorithmes de *hashage* utilisés en cryptographie. L'auteur propose donc d'utiliser SHA1 pour réaliser une empreinte d'un groupe de 4x4 blocs. Cependant ces empreintes ont une sortie dont la taille (160 bits) dépasse généralement la capacité de la zone que l'on *hashe*. Pour remédier à ce problème, les auteurs projettent ce *hash* sur un nombre inférieur de bits. Concernant notre tatouage, il est évident qu'un *hash* permettrait une vérification quasi parfaite de l'intégrité, avec peu de chances de réussir une attaque valable. Cependant, dans le cadre du tatouage de certification, on souhaite pouvoir localiser plus finement les zones erronées. Par conséquent, il nous faut un contrôle d'intégrité au niveau d'un seul bloc. Dans ce contexte, en plus de l'enbvironnement mobile, le calcul d'une empreinte par ce type d'algorithme est trop complexe.

[Lin 2000b] La marque se compose d'un bruit pseudo-aléatoire disposant de certaines caractéristiques (c'est une distribution Gaussienne normale : variance unitaire et moyenne nulle). Cette marque est insérée dans le domaine fréquentiel, plus précisément, les basses fréquences. De fait, la détection se

fait par mesure de corrélations. La méthode qui permet de vérifier la marque ne donne pas un résultat binaire ce qui rend l'analyse de l'intégrité moins précise. En revanche, on remarque que les auteurs utilisent les basses fréquences (coefficients DCT). En effet, les plus hautes fréquences sont plus susceptibles d'être détruites par la compression JPEG. L'utilisation de la Gaussienne normale ne nous sera pas utile, car son objectif est de camoufler la marque statistiquement, ce qui serait plus utile dans le cadre de la stéganographie.

[Fridrich 1998a] Le contrôle d'intégrité de cette méthode se fait par bloc de 64x64 pixels. L'auteur utilise une technique à spectre large pour créer un bruit à partir de M bits venant du bloc considéré. Pour des blocs similaires, ce bruit doit être approchant. On construit ensuite une séquence pseudo-aléatoire à partir de l'identifiant de l'appareil servant à la capture. Les blocs subissent un adoucissement et sont traités par un *low-pass filter* puis sont rendus indépendants du DC. On extrait ensuite une séquence (à partir des informations du bloc et de la séquence). Un paramètre de seuil permet de calculer un résultat 0 ou 1. Il est choisi de manière à ce que le nombre de 0 et de 1 soit à peu près équivalent. Notre intérêt dans cette algorithme concerne l'utilisation d'une chaîne pseudo-aléatoire initialisée à partir de l'identifiant.

[Atupelage 2008] Dans cet algorithme, les auteurs n'utilisent que le coefficient DC et deux ACs comme informations provenant d'un bloc. Cela représente selon eux 70% de l'information visuelle de l'image. Cette information est aussi confirmée dans [Weng 2007]. La vérification de l'intégrité se fait en combinaison avec une signature numérique. Cette signature se base sur la méthode des courbes elliptiques. L'inclusion de cette signature dans l'image se fait dans les coefficients DCT de toute l'image. Ce que l'on retient avant tout ici, est que les premiers coefficients réunissent la plus grande partie de l'information visuelle. C'est une information que l'on devrait prendre en compte lors du choix de la caractérisation d'un bloc et lors de l'insertion de la marque. Mais l'insertion d'une signature sur l'ensemble de l'image empêche une localisation précise d'éventuels problèmes d'intégrité.

[Fridrich 2002a] La méthode utilisée pour déterminer l'intégrité d'un bloc est l'utilisation d'une chaîne binaire, constituée d'informations concernant l'image (l'index de l'image, sa dimension, etc.), mais aussi du bloc (position dans l'image). La chaîne binaire ainsi constituée est *XORée* avec le hash des 128 pixels du macro-bloc `8x16`. Une clé particulière sert aussi à chiffrer la marque lors de l'insertion dans les LSB. La vérification se fait en vérifiant la redondance insérée dans la chaîne après déchiffrement et extraction. Composer une chaine binaire avec des informations spécifiques au bloc dans lequel on la tatoue constitue un moyen efficace pour localiser l'intégrité. Cependant, la composition de la chaine binaire et le hashage proposé par les auteurs sont

discutables du point de vue de la complexité au moins dans le contexte de la certification.

[Kundur 1999] La marque est insérée dans le domaine multi-résolutions (ondelettes). Elle se compose d'une clé de validation, d'une clé de sélection de coefficients, d'un paramètre de quantification et de la fonction porteuse mère pour le traitement par ondelettes. L'extraction se fait grâce à ces mêmes informations. Pour déterminer si l'image a été modifiée, une fonction nommée TAF (*Tampering Assessement Function*) permet, grâce à un seuil, de savoir si une modification est effectuée. Là encore, la détection d'une modification se fait par une estimation d'une part, et d'autre part, elle est globale à l'image.

2.3.2.2 Mécanismes de localisation

[Lin 2000a] L'algorithme proposé ici est semi-fragile et embarque deux marques. Une marque sert à l'authentification, la seconde à la récupération. C'est la première marque qui permet de détecter et localiser les erreurs. Pour cela, les auteurs effectuent une approximation de chaque bloc original. Il est possible alors d'intégrer une marque en modifiant des paramètres restant invariants lors de la quantification. Cette technique inclut de l'information pour reconstruire l'image en plus de la détection d'erreur. Ce n'est pas une technique très adaptée pour nous, car la capacité consommée par les informations de récupération est trop importante et se fait au détriment des autres propriétés du tatouage (complexité, invisibilité). Par ailleurs, la récupération d'une image plus ou moins originale à partir d'une image modifiée, n'est pas du tout une fonction pertinente dans le contexte.

[Hassan 2006] Cette méthode utilise l'espace colorimétrique YST de Francescso *et al.* (voir [Benedetto 2005]) pour embarquer la marque. Pour chaque groupe de 4x4 macro-blocs de l'image, un *hash* SHA1 est calculé avec les 12 blocs du bord, ainsi qu'avec des informations sur le propriétaire et l'horodatage. On y concatène un calcul de l'intensité du bloc pour la reconstruction. Les 160 bits sont projetés pour obtenir un nombre de bits correspondant à la capacité d'un sous-groupe de *2x2* blocs, le centre du groupe. L'avantage de la méthode est qu'elle travaille par macro-bloc. Une perte d'intégrité dans l'espace T permet de localiser une altération. Les informations d'intensité dans YS et T permettent de reconstruire l'image. La double conversion RGB, YST, puis RGB non adaptée au cadre mobile. En effet, cela constitue une complexité de calculs trop importante.

[Fridrich 2002a] Le principe d'insertion (décrit précédemment) d'une chaine binaire composée d'informations, entre autres, sur le macro-bloc, est un moyen pertinent pour obtenir une bonne localisation des problèmes d'intégrité. Pour qu'elle soit intéressante dans notre cas, il serait préférable d'avoir

des blocs de 8x8 pixels conformément au processus de compression JPEG standard.

[Lin 2000b] Un test statistique est effectué sur chaque bloc, avec un paramètre de seuil qui détermine si le bloc est altéré. La comparaison se fait par corrélation de pixels adjacents. On se base sur le fait que, sauf si une bordure est présente, la différence entre deux pixels correspond au tatouage. On peut donc estimer si cette variation représente une rupture de l'intégrité ou non. La granularité offerte par la méthode correspond aux objectifs fixés par le tatouage de certification (le macro-bloc 8x8 pixels). Chaque bloc peut être considéré intègre ou non, quelle que soit l'interprétation qui a été faite du bloc précédent.

2.3.2.3 Éléments de sécurité

[Lin 2000a] Chaque bloc est associé à un second bloc grâce à une fonction secrète. Pour chaque paire, on prend β_a coefficients parmi les 64 coefficients possibles. La marque est insérée grâce à une autre fonction de projection T_a. Cette méthode base essentiellement la sécurité sur la sélection des blocs et celle des coefficients où est insérée la marque. La fonction d'appariement des blocs est une fonction pseudo-aléatoire. Les auteurs proposent donc d'utiliser une graine pour initialiser cette fonction qui sera insérée dans l'image avec une méthode à large spectre. Ce processus se déroule avant l'étape de quantification. A partir d'une image compressée, comme c'est le cas dans notre étude, il est donc nécessaire de réaliser les deux premières étapes de la décompression (décodage entropique puis déquantification) pour réaliser ce type de tatouage. Ce qui n'est pas réaliste du point de vue de la complexité.

[Eggers 2001] L'algorithme utilise le SCS (*Scala Costa Scheme*) pour tatouer des informations dans une image en générant de la distorsion. La distorsion est dépendante d'une clé k. C'est donc une insertion paramétrée. L'algorithme base l'accès à la marque sur cette clé, qui sert à générer l'information tatouée. Générer une marque à partir d'une clé utilisateur permet de diversifier les marques possibles et rendre les attaques plus difficiles. C'est une pratique intéressante dans le cadre de la certification.

[Kundur 1999] Dans la méthode de Kundur et Hatzinakos, les auteurs utilisent plusieurs clés. La première appelée *validation key* est la marque de l'utilisateur. La seconde est la *coefficient key (ckey)* et indique quels coefficients dans l'image vont porter la marque. La dernière, la *quantisation key (qkey)* sert à rendre l'insertion irréalisable pour un attaquant qui ne la connaîtrait pas. Dans cette méthode, bien que l'utilisation faite des clés ne nous intéresse pas spécialement, l'utilisation de multiples clés en revanche nous semble un bon moyen pour contrôler les diverses opérations du processus.

[Fridrich 1998a] Cet algorithme robuste utilise des blocs de 64x64 pixels et la marque est insérée grâce à la technique de large spectre (*spread spectrum*, modulation des fréquences DCT). Fridrich utilise un générateur de nombres pseudo-aléatoires pour créer une chaîne binaire dépendante de l'image. L'utilisation d'un générateur de nombres pseudo-aléatoire attire notre attention, car il est moins coûteux de garder uniquement la graine que l'ensemble des motifs. De plus, cela permet de générer une séquence binaire difficile ment reproductible sans connaissance de la graine et qui varie fortement lorsque cette dernière change. Cela peut-être une méthode pour distinguer les marques d'un utilisateur particulier (la graine est liée à la clé de l'utilisateur).

[Fridrich 2002a] L'empreinte du bloc calculée par cette méthode (voir précédemment) est insérée dans les LSB. Ensuite, ces LSB sont chiffrés grâce à une clé. Par ce procédé, l'intégrité d'un bloc est protégée pour un utilisateur donné. Toute modification des MSB sera détectée par l'utilisateur en vérifiant les LSB. Cependant, n'importe quel autre utilisateur, muni de sa propre clé, peut appliquer le même procédé et remplacer la marque de l'utilisateur original par sa propre marque. La sécurité de l'image n'est donc pas complètement garantie dans la mesure où un attaquant peut se l'approprier. Dans le cadre de la certification, si l'idée de protéger l'intégrité est intéressante, elle n'est pas suffisante. D'autres informations doivent être tatouées pour compléter le dispositif et garantir qu'une image appartienne toujours à son auteur.

Fei *et al.* proposent un *framework* pour les systèmes d'authentification basés sur le tatouage. Concernant l'accès à la marque, ils recommande l'utilisation d'une fonction de hashage couplée à une clé k. Ce principe, similaire aux HMAC, utilise les *Most Significant Bit* - Bit de poids fort (MSB) comme entrée et le résultat est inséré dans les LSB. Cette technique nous intéresse, car elle permet de faire dépendre le tatouage de l'image elle-même.

[Atupelage 2008] L'algorithme insère un *hash* du coefficient DC et des deux premiers coefficients AC signés par la méthode des courbes elliptiques dans les LSB de l'image. La gestion des signatures repose, c'est une particularité dans cet article, sur l'utilisation d'une PKI. C'est une méthode intéressante, basée sur des outils connus et maîtrisés. Cependant notre cas ne prévoit pas de pouvoir utiliser des signatures de cette sorte. En effet, elles nécessitent au moins deux passe sur l'image, ce qui ne permet plus de réaliser un traitement au fil de l'eau. Cela ne remplit pas notre contrainte de complexité.

2.3.2.4 Propriété de robustesse

[Fei 2006] Les auteurs expliquent que l'avantage à utiliser un tatouage semi-fragile est d'avoir un plus grand potentiel pour caractériser les distorsions faites sur l'image, et aussi à concevoir une méthode robuste à certains types

d'opérations. Les algorithmes semi-fragiles doivent être traités comme un problème d'encodage/vérification. C'est tout à fait le type de tatouage que nous recherchons, car il doit être robuste à la compression JPEG mais fragile à toutes les autres modifications.

[Fridrich 2002a] L'algorithme utilise des macro-blocs de 8x16 pixels. De ces blocs on calcule un *hash* à partir des MSB, puis on insère dans la LSB une chaîne binaire XOR avec ce *hash*. Ces LSB sont ensuite chiffrés. Cet algorithme est conçu pour résister aux attaques de Holliman-Memon, aux *verification center attack*, aux attaques sur l'image originale et aux attaques à image originale choisie. Cette robustesse est obtenue par modification du schéma de tatouage de Wong [Wong 2000], en utilisant une chaîne différente contenant des informations sur les blocs et l'image. La résistance à ce type d'attaque nous intéresse. Cependant, la réalisation implique de pouvoir insérer une information relative à l'image dans son ensemble dans chaque bloc. Cela requiert une forte capacité pour un seul bloc, qui serait obtenue au détriment de l'imperceptibilité.

[Fridrich 1998b] L'algorithme utilisé par l'auteur utilise la technique à large spectre (*spread spectrum*, modulation des fréquences DCT) qui rend le tatouage robuste aux modifications suivantes : ajustement du contraste/luminosité, ajout de bruit, manipulation d'histogramme et découpage. De plus, il est robuste à la compression JPEG avec un facteur moyen de qualité. La technique utilise un spectre large. Cela n'est pas adapté à nos contraintes, car cette méthode s'applique avant la quantification.

[Eggers 2001] Cet algorithme est basé sur le schéma scalaire de Costa (*Scalar Costa Scheme*, SCS). Ce tatouage est aveugle. Il s'effectue par distorsion de l'image source et utilise une clef lors de l'insertion. Il s'effectue indépendamment de la compression JPEG et la détection s'effectue par un test probabiliste. Cette technique ne nous intéresse pas, à la fois pour des raisons de performance (faible intégration à JPEG et mauvais robustesse à JPEG) que pour la méthode de détection qui n'est pas immédiate.

[Atupelage 2008] L'algorithme insère un *hash* des DC et deux premiers AC signés par la méthode des courbes elliptiques dans les LSB de l'image. Les informations sont insérées avant la quantification. Avec cette technique, la marque n'est robuste qu'à la compression JPEG. En effet, la signature sera rendu invalide par n'importe quelle modification d'un DC ou d'un des deux premiers coefficients AC de n'importe quel bloc. Il n'y aura aucun moyen de savoir quel bloc a rendu la signature invalide. Cette technique nous intéresse, car il s'agit d'un moyen assez aisé d'obtenir de la robustesse à JPEG. Cependant, nous privilégierons la réalisation de toutes les étapes avec pertes avant le tatouage, car l'on souhaite limiter la complexité de notre algorithme.

[Lin 2000b] Cet algorithme utilise une technique simple d'étalement de spectre (*spread-spectrum*) avec un détecteur modifié. Le récepteur effectue un test statistique sur chaque bloc, avec un paramètre de seuil qui détermine si le bloc est altéré. La comparaison se fait par corrélation de pixels adjacents. Elle se base sur le fait que, sauf si une bordure est présente, la différence entre deux pixels correspond au tatouage. On peut donc estimer si cette variation représente une rupture de l'intégrité ou non. Dans cet article, l'algorithme est robuste à la compression JPEG. Pour obtenir cette robustesse, l'insertion est faite après réalisation de la DCT, dans les basses fréquences car celles-ci sont moins sensible à la quantification. La tatouage des basses fréquences DCT est un principe que l'on conservera pour garantir une robustesse suffisante lors des re-compression JPEG.

2.3.2.5 Gestion de la capacité

[Fridrich 2002a] L'algorithme utilise des macro-blocs de 8x16 pixels. De ces blocs on calcul un *hash* à partir des MSB, puis on insère dans la LSB une chaîne binaire XOR avec ce *hash*. Ces LSB sont ensuite chiffrés. L'insertion de l'information supplémentaire se fait dans la chaîne qui permet la vérification. La marque contient à la fois un moyen de vérifier et de transmettre une information. La vérification de l'information se base sur l'étude de la redondance contenue dans les données insérée. Or cela requiert une trop grande capacité par bloc.

[Hassan 2006] La méthode utilise l'intérieur des macro-blocs *4x4* (soit un macro-bloc de*2x2* blocs) dans lequel deux LSB (soit 8 bits) servent à mettre des informations de reconstruction. Dans cet article, les bits servent à la reconstruction, on peut imaginer y embarquer des informations sur l'auteur. Cette méthode peut modifier jusqu'à trois LSB d'un coefficient, ce qui confère une capacité élevée, mais l'impact sur l'imperceptibilité n'est pas neutre.

[Lu 2001] Dans cet article, les auteurs cherchent à insérer deux types d'information, un *copyright* et une information d'intégrité. La méthode qu'ils proposent repose sur un cocktail de tatouages. Le premier tatouage est robuste et garantit la conservation du *copyright*. Le second tatouage est fragile pour assurer l'intégrité de l'image. Si nous avons également deux types d'informations a tatouer (les métadonnées et l'intégrité), les deux types d'information doivent être tatouer avec la même fragilité. Par ailleurs, un seul algorithme aurait l'avantage d'éviter de traiter deux fois l'image, évitant ainsi des problèmes de performance.

Table 2.1 – Tableau des fonctionnalités des méthodes de tatouage

	Robustesse JPEG	Protection marque	Dépendance image	Imperceptibilité	Validation intégrité	Localisation
Tatouage de certification	oui	oui	oui	oui	oui	oui
Lin Chang [Lin 2000a]	oui	oui	oui	oui	oui	oui
Lin Podilchuck Delp [Lin 2000b]	oui	non	non	oui	oui	oui
Eggers Girod [Eggers 2001]	oui	oui	non	oui	oui	oui
Fridrich [Fridrich 1998a]	oui	oui	oui	oui	oui	oui
Kundur Hatzinkos [Kundur 1999]	oui	oui	non	oui	oui	oui
Atupelage Harada [Atupelage 2008]	oui	oui	oui	oui	oui	oui (4x4)
Hassan Gilani [Hassan 2006]	non	non	non	oui	oui	oui (4x4)
Fridrich [Fridrich 2002a]	oui	oui	oui	non	oui	oui (8x16)
Fei,Kundur [Fei 2006]	oui	oui	non	oui	oui	oui

2.3.3 Synthèse

Le tableau 2.1 récapitule les propriétés recherchées et la manière dont elles sont prises en compte dans ces articles. On remarque que ces algorithmes ne nous fournissent pas de solutions parfaitement adaptées aux contraintes de notre problème. Pour obtenir un algorithme de certification, il va nous falloir, à partir de ces différents algorithmes, trouver un compromis entre la complexité, la capacité, la localisation, l'imperceptibilité et la sécurité. Sur la base de cette analyse critique, nous allons, dans la section suivante, concevoir un nouvel algorithme de tatouage, permettant en même temps l'insertion des méta-données d'une part et d'autre part les marques d'intégrité, et assurant de plus la protection des éléments insérés.

2.4 Tatouage de certification CodaSystem

Abordons maintenant la partie plus technique de ce chapitre. Dans cette section nous allons montrer comment nous avons construit l'algorithme de tatouage de certification, tout en justifiant les choix effectués. La première sous-section montrera comment chacune des propriétés retenues est prise en compte dans la conception de l'algorithme. La seconde donnera des détails sur les choix liées à l'implémentation.

2.4.1 Spécification

Nous avons vu précédemment qu'un algorithme de tatouage est caractérisé par plusieurs aspects comme la robustesse ou l'imperceptibilité. L'objectif de

sécurité et le contexte d'utilisation imposent de réaliser un compromis sur la manière dont chacune de ces caractéristiques est satisfaite. Dans cette section, nous définirons précisément le compromis que nous avons choisi pour obtenir un algorithme de tatouage de certification. Nous commencerons par décrire la manière d'insérer les informations (contrôle d'intégrité et message) ; comment les disséminer dans l'image et comment les rendre imperceptibles. Pour terminer, nous verrons les techniques qui nous permettent de protéger ces informations.

2.4.1.1 Robustesse

L'un des usages que doit satisfaire le tatouage de certification envisagé est la recompression. En effet, une image compressée doit évidemment pouvoir être décompressée afin d'être visualisée. L'image vue est bien sûr tatouée. Mais cette image doit pouvoir ensuite être recompressée (dans les mêmes conditions) et conserver son tatouage.

Le tatouage de certification fait partie de la catégorie des algorithmes dits semi-fragiles : il doit en effet être robuste à certains traitements effectués sur l'image mais, en revanche, il ne doit pas résister à des manipulations dites "structurelles" de l'image s'il veut remplir efficacement sa fonction de contrôle d'intégrité. Dans un premier temps, le tatouage de certification doit principalement résister à une compression de type JPEG. En effet, il s'agit du mode de stockage le plus couramment utilisé au sein des appareils mobiles. Mais ce format de compression engendre des pertes. Plusieurs stratégies existent pour obtenir un algorithme résistant à la compression JPEG. La plus simple, et celle offrant un taux d'erreur nul, consiste à effectuer le tatouage à partir des données quantifiées, c'est-à-dire après la dernière opération destructive de la compression JPEG.

Nous verrons dans le chapitre consacré à la description de l'algorithme les pistes concernant la robustesse du tatouage face à des erreurs dues à un ré-encodage JPEG (avec paramètres identiques).

2.4.1.2 Invisibilité

Notre tatouage n'étant pas de la stéganographie, il n'a pas pour objectif d'être camouflé dans l'image. Cependant, il est bon de tenir compte des cas d'usages. Certifier une image trop dégradée par rapport à la réalité n'est pas très convaincant. De plus les personnes réalisant les photos doivent pouvoir les exploiter au delà de la simple preuve. L'invisibilité est donc souhaitable. Dans le domaine fréquentiel, les tests montrent que modifier trop de coefficients nuit à la qualité de l'image. Il faut absolument éviter de changer le coefficient DC qui décrit énormément d'informations. La robustesse nous imposant de rester dans la diagonale supérieure du bloc, nous commencerons à tatouer à partir du second coefficient. Il existe des méthodes issues de la stéganographie qui per-

mettent d'insérer une information de k bits en ne modifiant qu'un bit parmi 2^k. Cette technique, appelée *enconding matrix*, est basée sur les codes correcteurs de Hamming sera décrite en section 2.4.2.2.

2.4.1.3 Imperceptibilité

Pour pouvoir qualifier une image d'original numérique, l'imperceptibilité du tatouage doit être maximale. En particulier, aucun artefact dû au tatouage ne doit apparaître dans l'image. Afin de minimiser la quantité de données de l'image modifiées par le tatouage, un mode de codage particulier est utilisé : le *matrix encoding*. Le point clef de cette technique est la possibilité de tatouer le message en ne changeant qu'au plus un bit d'un même bloc. Par ailleurs, le processus de tatouage doit éviter au maximum de tatouer les zones dites homogènes (mur blanc, ciel bleu, etc.). En effet, ces zones sont visuellement sensibles à toute modification. De fait, elles pourraient facilement trahir la présence d'un bit de la marque et faciliter le travail de l'attaquant potentiel.

Puisque l'algorithme proposé est intégré dans le processus de compression JPEG le choix du taux de compression JPEG ne doit pas introduire d'artefacts.

2.4.1.4 Localisation

La méthode de localisation est principalement identique pour tous les algorithmes que nous avons étudiés. Nous effectuons un contrôle d'intégrité d'une zone, et s'il échoue, on marque la zone comme modifiée. La véritable différence est la granularité. Par exemple, le contrôle peut porter sur tous les coefficients de l'image simultanément (c'est le cas de nombreux algorithmes). Dans ce cas, si l'intégrité n'est pas vérifiée, c'est toute l'image qui est rejetée et ce, quelle que soit la modification, même mineure. En revanche, la granularité souhaitée dans notre contexte doit être beaucoup plus fine. En effet, nous autorisons la recompression de l'image tatouée. Et cette recompression peut introduire de temps à autre des modifications marginales d'arrondi. De ce fait, si une image subit ces altérations dans des zones informationnelles non pertinentes, les blocs intègres restants peuvent malgré tout constituer une image suffisamment probante.

Dans la mesure où la compression JPEG est au coeur de notre proposition, et que cette compression procède bloc par bloc, nous nous efforcerons de situer le contrôle d'intégrité à ce niveau. En d'autres termes, nous souhaitons pouvoir dire si un bloc 8x8 a été modifié ou non.

2.4.1.5 Intégrité

La méthode que nous choisissons d'appliquer se base sur l'utilisation d'un logo binaire. Pour ne pas inclure toujours le même logo et pour conserver facilement les diverses versions employées par les utilisateurs, nous nous servi-

rons d'un générateur pseudo-aléatoire paramétré par une graine dépendante de l'utilisateur.

2.4.1.6 Sécurité

La sécurité du tatouage s'appuie sur trois facteurs essentiels : l'indétectabilité, la confidentialité et la confiance. Le premier facteur, l'indétectabilité, consiste à faire dépendre la règle d'insertion de paramètres secrets, déduits d'une clé cryptographique liée à l'utilisateur. En effet, compte tenu des principes de Kerckhoffs, il est prudent de partir de l'idée que l'algorithme d'insertion est connu de l'attaquant. Par conséquent, définir une stratégie d'insertion basée sur des critères seulement dépendants de l'image réduit l'entropie du principe de tatouage. Entropie d'autant plus réduite que parmi les objectifs, les images doivent rester "visibles", ce qui en termes cryptographique est équivalent à "publiques". Ce qui permet naturellement à l'attaquant, d'exploiter les stratégies d'attaque par corrélation. En introduisant en plus de ces critères liés à l'image, une clé liée à l'utilisateur, donc "privée", l'entropie globale est accrue donc la marque est moins détectable. Le second facteur, la confidentialité, consiste à rendre les informations tatouées illisibles sans la connaissance de la clé. En effet, dans l'éventualité où la marque a été frauduleusement détectée et extraite, le chiffrement des informations rend leur exploitation impossible. Le principe cryptographique retenu se base sur la méthode classique appelée *Output FeedBack* (OFB) et consiste à combiner par « ou exclusif », les bits de la marque avec les bits d'une séquence générée pseudo-aléatoirement. Là encore, la clé de l'utilisateur aura un rôle important à jouer, en garantissant la synchronisation du générateur pseudo-aléatoire. Ces éléments cryptographiques (clé et générateur pseudo-aléatoire) doivent être bien choisis de manière à éliminer naturellement toute attaque par force brute. Enfin, le dernier facteur intervenant dans la sécurité de la méthode de tatouage est la confiance. Comme on l'a vu, la clé est un élément central dans la sécurité. Mais la sécurité contenue dans la clé ne vaut que si sa gestion est digne de confiance, c'est-à-dire que le protocole de distribution, de stockage et de répudiation est bien construit et respecté. Nous aborderons dans un paragraphe spécifique ces questions concernant le tatouage de certification. Typiquement, il semble difficile de mettre en place un système de gestion de type PKI dont les clés auraient une durée relativement courte. En effet, il semble contradictoire par exemple d'avoir besoin de la date d'utilisation du système pour décoder le message afin d'extraire la date à laquelle a été tatouée l'image. Des pistes pour contourner ce paradoxe seront évoquées dans le chapitre A.

2.4.1.7 Authentification

L'authentification se fait par la connaissance des clefs. En effet, ce sont celles-ci qui vont permettre de calculer les motifs insérés dans l'image lors du

codage. La vérification de l'intégrité grâce à ce motif, propre à chaque utilisateur, permet de constituer le caractère authentique et certifié à la photo.

2.4.1.8 Complexité

Le tatouage de certification proposé a la particularité d'être destiné à des usages nomades. Il s'agit de certifier où, quand et par qui a été prise la photo. Par conséquent, une contrainte fondamentale pour cet algorithme est d'être implémenté sur un équipement mobile. Il ne peut pas tolérer un temps d'exécution trop long pour des raisons à la fois d'usage et de consommation. Une autre contrainte réside dans le fait que le tatouage est intimement lié à la compression JPEG. La stratégie la plus avantageuse en terme de coût de calcul consiste à intégrer le tatouage dans ce processus JPEG au niveau de l'étape de quantification ou en aval (*i.e* dans le codeur JPEG lui-même) de celle-ci. (Voir détails des étapes de compression en 2.1)

2.4.1.9 Intégration

L'utilisation du code sur mobile nous oblige à faire attention à la complexité en temps et en espace. Pour résoudre ce problème, on se forcera à laisser un maximum d'opérations aux blocs dédiés du mobile, notamment pour la compression JPEG (calcul DCT, transformation colorimétrique, ...). Pour ne pas encombrer le processus, nous ne voulons exécuter notre code qu'une fois par image, donc nous ne ferons pas de tests multiples pour trouver une solution optimale, ni de méthode intégrant deux tatouages distincts. De même les calculs cryptographiques doivent être contrôlés, car nous ne pouvons pas nous permettre d'utiliser des solutions trop gourmandes en ressources. Pour tirer parti de la compression JPEG et ne pas consommer trop de mémoire, toutes les opérations devront être réalisées bloc par bloc.

La plupart du temps, les applications multimédia utilisées dans le domaine de la téléphonie sont optimisées pour les processeurs embarqués (de type ARM) et parfois même bénéficient de blocs hardware dédiés pour optimiser la rapidité et la consommation. Il est par conséquent très difficile d'avoir accès à la chaîne de compression JPEG et d'y intégrer directement un module de tatouage sans le concours du constructeur ou de l'intégrateur de l'appareil. Pour contourner cette difficulté, la solution consiste à concevoir un programme s'appliquant aux images JPEG déjà compressées. Cette stratégie présente deux avantages :

- Les images JPEG produites dans un téléphone mobile sont disponibles dans l'appareil et ne nécessitent pas de faire appel à un constructeur.
- Même si elle n'offre pas le même potentiel d'optimisation que l'intégration directe dans le codeur, cette stratégie permet de faire l'économie des traitements les plus coûteux (ce point sera détaillé dans la suite).

Toutefois, l'intégration directe ou partielle au processus JPEG implique un certain nombre de contraintes communes : place mémoire limitée et proces-

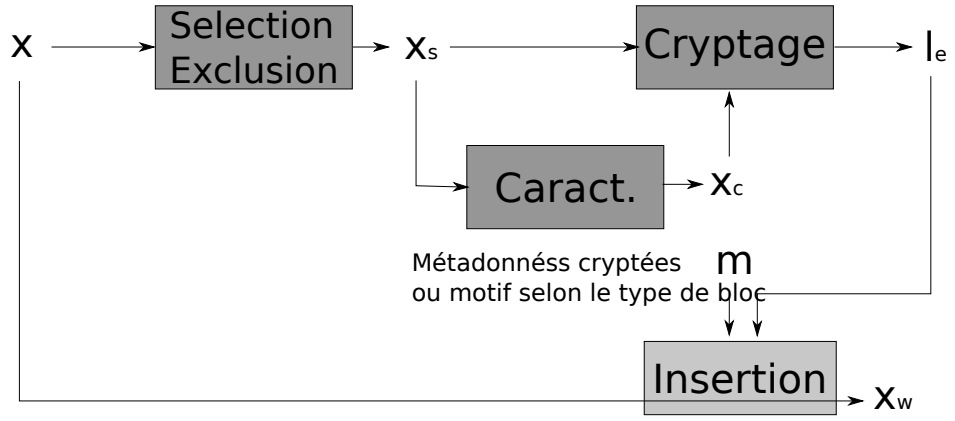


Figure 2.11 – Fonctionnement général du tatouage de certification.
X : bloc source ; X_s : Bloc sélectionné ; X_c : Caractéristique du bloc ; l_e : LSB extraits et cryptés ; X_w : bloc tatoué.

seurs moins puissants que dans le monde des ordinateurs. Pour assurer la prise en compte de ces deux points, la conception de l'algorithme de tatouage de certification s'appuie sur deux principes de base :

- Traitement des données séquentiellement par blocs de 8x8 pixels (nous verrons dans la suite qu'il s'agit là de la structure des données qu'utilise un encodeur JPEG)
- Insertion constituée d'opérations peu coûteuses (ex : opérateurs logiques bit à bit).

2.4.2 Implémentation

Nous allons maintenant présenter l'implémentation de l'algorithme. Nous commencerons par une présentation générale avant d'exposer la méthode d'insertion, basée sur les *encoding matrix*, et enfin les modifications que nous avons apporté à la technique initiale pour protéger la marque.

2.4.2.1 Présentation générale de l'algorithme de tatouage

La figure 2.11 représente le fonctionnement de l'algorithme dans un bloc DCT quantifié. L'algorithme s'applique à une séquence binaire Correspondant aux LSB des coefficients DCT choisis pour le tatouage (les coefficients DC et les plus hautes fréquences ne sont pas prises en compte).

Étape 1 : sélection et exclusion

Il s'agit d'une étape de décision relative au type d'informations que va contenir le bloc. Il y a trois cas de figure : soit ce bloc contient des métadonnées, soit des informations dédiées au contrôle d'intégrité (ils constituent une chaîne binaire que l'on appelle ici motif). Le troisième cas de figure est l'exclusion du bloc DCT si celui-ci ne respecte pas certains critères assurant l'invisibilité et la sécurité du tatouage. Le choix de ce critère découle de la constatation suivante : lorsque le bloc est massivement composé de coefficients nuls, celui-ci peut être considéré comme homogène, c'est-à-dire ayant peu de variation fréquentielle. Les zones homogènes ne sont pas adaptées au tatouage, celui-ci étant rapidement visible. Par ailleurs, tatouer les blocs DCT contenant un grand nombre de 0 peut faciliter les attaques statistiques. On nommera ce critère "seuil", car il indique le nombre maximum de zéros que l'on tolère dans un bloc DCT avant de l'exclure.

Étape 2 : calcul de la caractéristique

La caractéristique est une donnée calculée afin de faire dépendre la règle d'insertion du contenu de l'image pour parer aux attaques de type duplication. Cette caractéristique est une séquence binaire calculée à partir des MSB des coefficients DCT quantifiés. Le choix des MSB utilisés dans le calcul est un paramètre secret permettant d'assurer la robustesse du système face à une attaque de type duplication.

Étape 3 et 4 : cryptage et insertion

L'insertion est basée sur une technique de codage appelée *matrix encoding* extrêmement peu destructive pour le bloc DCT. Ce mode de codage est modifié de sorte à faire dépendre la règle d'insertion à la fois de la caractéristique et de paramètres secrets. L'idée clef consiste à perturber l'*encoding matrix* en effectuant une opération OU logique ($\oplus$) entre les LSB et la séquence binaire constituée de la caractéristique et de paramètres secrets. La méthode sera expliquée en détail dans le paragraphe 2.4.2.2. Il faut souligner toutefois que la perturbation de l'*encoding matrix* n'est pas robuste à une attaque sur les clés dans le cas d'un message connu. Ce point sera contourné par l'intermédiaire d'un cryptage du message.

2.4.2.2 Manière d'insérer

Intégration dans le processus JPEG : Le tatouage de certification est dédié aux processeurs embarqués sur équipement mobile à puissance et mémoire limitées. Il doit donc en premier lieu tirer profit des algorithmes optimisés présents en natif sur l'appareil. Comme cela a été précisé dans la section précédente, l'approche retenue afin d'être robuste à la compression JPEG et de minimiser le temps d'exécution de l'application est l'intégration du tatouage au sein du processus JPEG.

Cette approche entraîne des coûts de calcul supplémentaires : décodage et réencodage entropique. Nous verrons cependant que les temps d'exécution obtenus sont très acceptables pour un usage nomade.

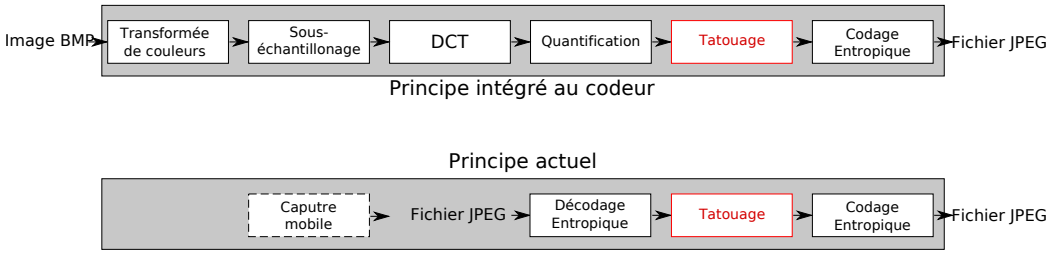


Figure 2.12 – Intégration du tatouage.
Dans l'encodeur JPEG en haut. Après l'encodage sur mobile en bas.

Comme nous l'avons précisé en introduction, le tatouage de certification a une double fonction d'insertion d'informations et de contrôle d'intégrité. Ces deux fonctionnalités sont essentiellement remplies par le contenu du message composé d'une part des métadonnées (géolocalisation, identification, horodatage - ~500 caractères) et d'une séquence binaire dédiée au contrôle d'intégrité. Ces informations sont regroupées en blocs de 5 bits[4] correspondant à un bloc 8x8 de coefficients DCT bien précis. Il y a donc une correspondance directe entre les emplacements des blocs message et des blocs DCT. C'est cette relation qui constitue le point clef permettant la localisation des modifications pour le contrôle d'intégrité de l'image.

Motif : Un bloc motif est une séquence binaire pseudo-aléatoire qui permet de contrôler l'intégrité du bloc DCT correspondant : en comparant le motif décodé au motif qui a été inséré, on est capable de déterminer si le bloc de l'image a été modifié ou non.

Nous verrons, dans le paragraphe décrivant les mécanismes mis en place pour empêcher la duplication du message (voir section 2.4.2.3), que l'insertion du bloc motif dépend de certains critères. Premièrement, les blocs homogènes ne seront pas tatoués, afin de satisfaire certaines contraintes d'invisibilité et de sécurité. Deuxièmement, dans certains cas le motif pourra être modulé avec d'autres informations permettant de décrire certaines propriétés de l'image utiles lors du contrôle d'intégrité. Enfin, le motif est une séquence binaire qui constitue un des paramètres secrets du système. La question du mode de génération de ce paramètre secret sera abordée dans le paragraphe 2.4.2.4.

Métadonnées : Les métadonnées représentent, pour une image de taille 640x480, environ 4% de la longueur totale du message. Afin d'assurer la confi-

4. La taille de ces blocs est imposé par la règle d'insertion. L'explication concernant le choix sera détaillée dans le paragraphe suivant 2.4.2.2

dentialité de ces données, elles occupent des blocs choisis aléatoirement au sein de l'image. La répartition aléatoire du message est une méthode classique utilisée dans le domaine du tatouage pour répartir uniformément le message sur toute l'image et assurer la confidentialité. Une approche très répandue dans le domaine du tatouage [Westfeld 2001] consiste à effectuer une répartition à l'aide d'une permutation des coefficients. Malheureusement, même si cette méthode offre l'intérêt d'un mélange encore plus difficile à casser par force brute, elle nécessite une grande quantité de mémoire et ne permet pas le traitement séquentiel par bloc. L'approche par bloc nécessite seulement de générer une liste d'entiers correspondant aux numéros des blocs contenant les métadonnées.

Le contenu des métadonnées est normalisé. Il contient les informations de géolocalisation, d'identification de l'utilisateur et de l'horodatage.

Ces informations ne peuvent être extraites que par une personne ayant à sa disposition les paramètres secrets pour décoder le message. Il peut s'avérer intéressant d'y ajouter des informations qui, elles, seront insérées de manière indépendante des autres données afin de délivrer des éléments facilitant le décodage du tatouage :

- Insertion d'un numéro de version de l'algorithme de tatouage pour permettre l'évolution du tatouage.
- Insertion d'un paramètre secret renseignant sur le numéro de clé utilisé lors du tatouage.

La correspondance entre ce numéro et la clé étant connu au décodage, Cette approche permet d'apporter de la variété dans les clefs utilisées et ainsi de diminuer le risque d'attaque statistique sur les clefs. Nous reviendrons sur ce point dans le paragraphe sur la gestion des clefs cryptographiques.

La figure 2.13 présente une image du message illustrant une alternance possible l'utilisation des blocs messages des deux types décrits ci-dessus.

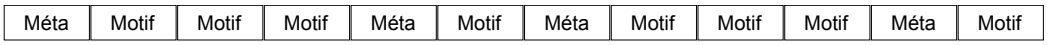


Figure 2.13 – Représentation du message, alternant les blocs messages de type métadonnées et motif

Les *encoding matrix* : Nous allons présenter ici le principe de l'insertion du message m dans un bloc donné x et ses LSB l. Le bloc tatoué, quant à lui, sera noté x_w et ses LSB l_w. On notera H l'application linéaire (*encoding matrix*) de $\mathbb{F}_2^n$ dans $\mathbb{F}_2^n$ utilisée comme codage de la séquence binaire l extraite de la source x. $\mathbb{F}_2$ étant le corps fini à deux éléments.

L'insertion I consiste à modifier la séquence binaire l de la source x de manière à transmettre k bits du message m (Un bloc message). x_w est donc l'image de x par I. Le principe des *encoding matrix*, introduit dans les travaux

de Ron Crandall dans [Crandall 1998] et implémenté par Andreas Westfeld dans [Westfeld 2001], dérive des codes de Hamming [Hamming 1950]. L'idée clé motivant l'utilisation des *encoding matrix* dans la mise au point de la fonction d'insertion I réside dans le fait que ces codes permettent de minimiser le nombre de bits modifiés lors de l'insertion. Plus précisément, les *encoding matrix* permettent d'obtenir une distance de Hamming au plus égale à 1 entre l et un l_w qui code le message m (Quels que soient l et m). Cette propriété rend les *encoding matrix* extrêmement intéressantes pour obtenir une fonction d'insertion I réalisant un bon compromis entre imperceptibilité et capcaité. En effet, pour tout l, on peut trouver un l_w dont un bit au plus diffère de l, et qui code le message m que l'on souhaite.

La fonction H est aussi le code (n, k) qui associe à k bits du message m le mot binaire l_w de longueur n embarquant de message. La relation entre k et n est la suivante :

$$n = 2^k - 1$$

L'encodage à l'aide de H s'effectue avec les calculs suivants :

$$s = H.l \tag{2.16}$$
$$h = m \oplus s \tag{2.17}$$
$$p_i = 1 \text{ pour } i = \sum_{r=1}^{n} h_r.2^{r-1} \text{ et } p_j = 0 \text{ pour } j \neq i \tag{2.18}$$
$$l_w = I(l, m) = l \oplus p \tag{2.19}$$

Ces équations définissent les quatre étapes de codage suivantes :

1. On calcul une séquence binaire s de taille k, appelée syndrome, correspondant à l'application de H à l (équation 2.16).
2. Le contrôle de parité h vaut 0 si le message m vaut déjà s. Si ce n'est pas le cas, on passe à l'étape suivante (équation 2.17).
3. D'après les travaux de Hamming sur les codes d'erreur, on peut calculer la position i du bit à modifier pour transformer l en m. On peut donc construire un vecteur unitaire p, de taille n, utilisé pour modifier l en l_w (équation 2.18)
4. L'insertion I du message m dans les LSB l (équation 2.19)

Le décodage est donné par la relation suivante :

$$E(l_w) = H.l_w = m$$

on notera cette fonction E. On remarquera que cette opération est la même que celle de l'étape 1 (équation 2.16) du codage. En effet, lors du codage, on évalue, en calculant s, la valeur qui sera décodée par l'extraction du message. Si cette

valeur n'est pas la valeur m que l'on veut transmettre, alors les opérations 2, 3 et 4 permettent de faire en sorte que

$$E(l_w) = H.l_w = m$$

Prenons un exemple simple avec $k = 2$, nous avons donc $n = 3$. La matrice H pour ces valeurs de k, n est $H = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$

Supposons que le message à insérer est l composé des bits l_2, l_1, l_0 suivants $\begin{pmatrix} 1 & 0 & 1 \end{pmatrix}$. Calculons le syndrome $s = H.l$.

$$s = H.l = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} . \begin{pmatrix} 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} l_2 \oplus l_0 \\ l_2 \oplus l_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Nous avons réalisé ici l'opération que le décodeur va effectuer. Si $s = m$ il suffit de s'arrêter là. Sinon, supposons que $m = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, il faut faire une modification de l. Calculons $h = m \oplus s = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \oplus \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Selon le principe des codes de Hamming, il faut transformer le troisième bit, soit l_2 (étape 3). On obtiens donc $l_w = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$. On vérifie que

$$H.l_w = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} . \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \overline{l_2} \oplus l_0 \\ l_2 \oplus l_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = m$$

La **propriété fondamentale** du principe d'*encoding matrix* est la suivante :

$$E(I(l, m)) = m$$

Comme nous pouvons le voir, l'insertion est composée d'opérations simples de type "XOR" ($\oplus$) et "ET" ($\wedge$) logiques, ce qui rend cette étape du tatouage relativement peu coûteuse. En informatique le calcul de 2.16 se fait avec la formule suivante :

$$s = \bigoplus_{i=1}^{n} l_i \wedge i \tag{2.20}$$

et 2.18 puis 2.19 par la conversion décimale de h qui nous permet de changer le bit i.

Cette technique nous permet d'insérer k bits de message en modifiant uniquement 1 bit parmi n coefficients ($n = 2^k - 1$). L'*encoding matrix* de taille maximale permettant de traiter des coefficients d'un même bloc est obtenue pour $k = 5$ et $n = 31$. Plus l'*encoding matrix* est grande (*i.e.* k grand) plus la proportion de coefficients modifiés diminue (quel que soit k (n), le nombre de

bits modifiés par ce principe est au plus un). Par conséquent, ce seront les paramètres retenus dans l'algorithme de tatouage de certification. Le taux d'insertion sera alors de 5 bits/bloc soit 0,078 bits/pixel. Le taux de modification sera au maximum d'un coefficient par bloc.

2.4.2.3 Protection des données

Dans la section précédente, nous avons décrit la technique permettant de tatouer un message m dans un bloc DCT x. Comme nous l'avons vu également, le message peut correspondre soit à une séquence de cinq bits de métadonnées ou cinq bits générés pseudo-aléatoirement pour assurer l'intégrité. Nous allons maintenant montrer comment nous renforçons la sécurité de la marque en perturbant le message grâce au contenu du bloc x. La fonction qui va nous donner le vecteur perturbateur est appelé caractéristique.

Caractéristique : Faire dépendre la perturbation du contenu du bloc permet de répondre efficacement au risque de duplication. Cela permet en outre de sécuriser l'insertion face à une attaque lorsque le message est connu.

La dépendance au contenu, appelée caractéristique, est définie par la fonction $\mathcal{C}_k$, équation 2.21. Pour la clé nous utiliserons $\mathcal{K}_c$.

Afin d'assurer l'extraction du message, cette caractéristique doit respecter la contrainte suivante :

$$\mathcal{C}_{\mathcal{K}_c}(x) = \mathcal{C}_{\mathcal{K}_c}(x_w) \tag{2.21}$$

En outre, cette caractéristique doit limiter les possibilités de collision, c'est-à-dire les possibilités que deux blocs différents admettent une même caractéristique. Cependant cette contrainte n'est pas aussi stricte que dans le cas d'un calcul d'empreinte par une fonction de hashage, puisqu'ici la collision n'impacterait qu'un seul bloc et non l'intégralité des données.

La méthode que nous proposons consiste à calculer la somme des y MSB des coefficients utilisés pour l'insertion. Nous obtenons donc un entier de 31 bits dont le bit de poids fort est la somme des MSB, et le bit de poids faible celui du MSB d'ordre $8 - y$. La figure 2.14 illustre le principe du calcul de la caractéristique.

Soit $\mathcal{C}_{k_i}$ le i$^{\text{ème}}$ bit de la caractéristique, $x_i(j)$ le i$^{\text{ème}}$ bit du j$^{\text{ème}}$ coefficient DCT de x, et $\mathcal{K}_i$ celui de la clé, on a :

$$\mathcal{C}_{k_i} = \bigoplus_{j=1}^{n} \mathcal{K}_i \wedge x_i(j) \tag{2.22}$$

On comprend aisément qu'utiliser cette caractéristique comme perturbation empêche la recopie simple de la marque : remplacer un bloc de l'image par celui d'une autre image engendrera une caractéristique différente de l'initiale.

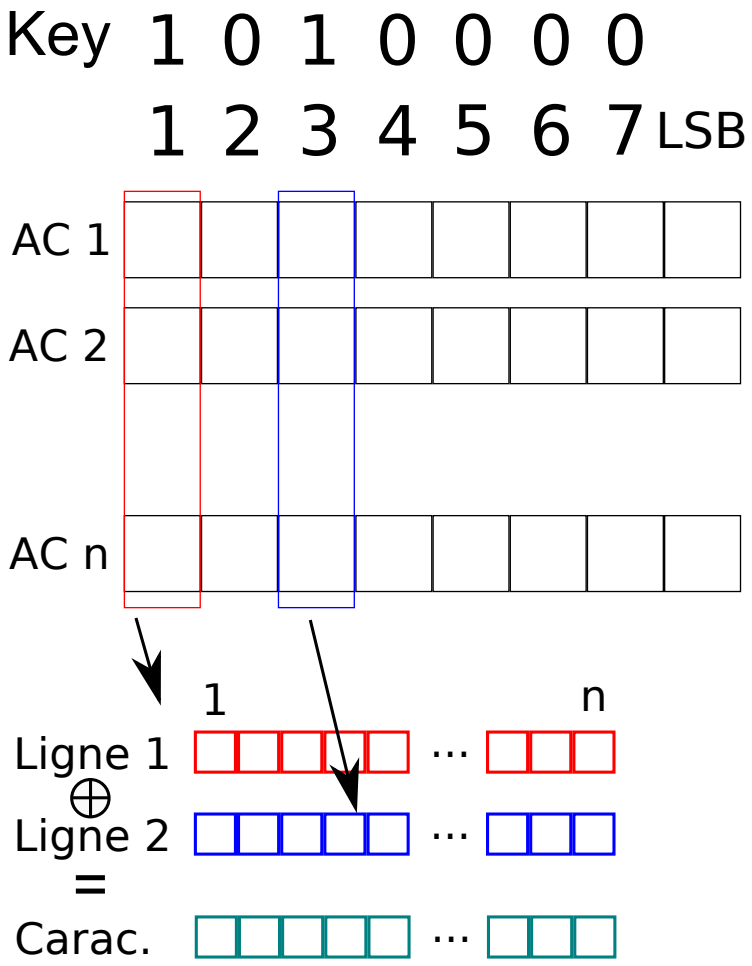


Figure 2.14 – Construction de la caractéristique

Toutefois pour contrecarrer une attaque de type duplication, cette caractéristique doit dépendre d'un paramètre secret.

Cryptage de la caractéristique : Le cryptage du calcul de la caractéristique s'effectue dans le choix des couches de MSB utilisées. 5 couches de MSB peuvent être potentiellement utilisées, par conséquent il y a 2^5 possibilités différentes ce qui constitue une clef secrète de 5 bits par bloc.

Exclusion des blocs homogènes : Les blocs contenus dans des zones homogènes de l'image admettent des coefficients DCT massivement nuls, ce qui pose plusieurs problèmes du point de vue du tatouage de certification. Premièrement, la modification d'un coefficient dans une telle zone va induire des changements plus visibles que dans les zones contenant des détails. Deuxièmement, la caractéristique utilise les MSB des coefficients DCT quantifiés. Dans un bloc homogène, la caractéristique sera donc nulle. Ce qui, comme nous l'avons vu, peut s'avérer critique en terme de sécurité.

Pour contourner ce problème, nous proposons d'exclure les blocs homogènes. Le critère d'exclusion est le suivant :

$$nombreDeCoefficientsNuls >= 27 \tag{2.23}$$

La justification de ce choix sera présenté dans une des sections des résultats expérimentaux 2.5.3. Ce choix tient compte de la satisfaction d'objectifs de sécurité et de qualité.

L'exclusion des blocs homogènes du processus de tatouage présente cependant un inconvénient : la modification d'une zone non homogène en zone homogène n'est pas détectée. Concrètement, l'effacement d'un objet contenu dans une zone homogène est susceptible de ne pas être détecté lors de l'extraction. Une solution à ce problème consiste à utiliser certains blocs motifs du message comme des compteurs des blocs exclus. De cette manière, il est possible de vérifier lors de l'extraction, si le nombre de blocs homogènes correspond bien au nombre tatoué.

2.4.2.4 Génération des paramètres secrets et gestion des clefs

Dans le but d'augmenter la difficulté de trouver nos clefs par une approche force brute, nous avons décidé de faire varier ces clefs pour chaque bloc. La technique sera expliquée dans le paragraphe suivant, et les conséquences de l'utilisation des différentes clefs seront présentées par la suite.

Génération des paramètres secrets pseudo-aléatoires Chaque paramètre secret utilisé dans l'algorithme est issu d'un générateur pseudo-aléatoire, initialisé par une clef de trente deux bits (un entier). Les paramètres secrets

peuvent utiliser directement le résultat de la pseudo-génération, ou seulement une partie. Cependant, que tout le résultat ou seulement une partie, soit consommé, lorsque le bloc change, on effectue une itération supplémentaire du générateur pseudo-aléatoire pour chaque clef (équation 2.24).

$$z_i = PRND(z_{i-1}) \tag{2.24}$$

Ce principe s'inspire de l'approche *Output FeedBack* (OFB).

Dans notre cas, ce paramètre secret a été remplacé par l'utilisation d'une fonction pseudo-aléatoire qui possède la particularité d'avoir un comportement difficilement prédictible sans la connaissance de sa clef d'initialisation.

Voici un résumé des différentes clefs utilisées dans l'algorithme :

- Clef d'insertion / paramètre e réduit par l'*encoding matrix* (32 bits / 5 bits) ;
- Clef de génération de motif (32 bits) ;
- Clef de cryptage de la caractéristique / paramètre secret (32 bits / 31 bits) ;
- Clef de sélection pseudo-aléatoire du message (32 bits)

Au total, l'algorithme de certification utilise une clef "cumulée" de 128 bits.

Gestion des clefs cryptographiques : La première justification de l'utilisation d'une clé dans la règle d'insertion était d'empêcher la recopie des méta-données d'une image dans une autre. Nous nous sommes vite rendu compte de la fragilité du système, puisqu'avec cette méthode, il est possible de recalculer la clé d'insertion si les méta-données sont connues. C'est aussi le cas lorsque le contrôle d'intégrité n'est codé qu'avec la caractéristique cryptée, car on peut effectuer une attaque statistique sachant que le motif est identique.

Pour résoudre ce problème, il faut ajouter un paramètre secret supplémentaire. Le cryptage de l'insertion vient donc compléter le cryptage de la caractéristique et le cryptage du message. Nous ajoutons donc une clé de cryptage $\mathcal{K}_I$ pour coder l'insertion. Combinée avec l'utilisation de la caractéristique on obtient la fonction $E_{\mathcal{K}_I,x_c}$. Pour ajouter à la difficulté de cryptanalyse, cette clé est dérivée pseudo-aléatoirement.

La question de la gestion des clés cryptographiques est un point crucial en sécurité informatique et l'est également en ce qui concerne le tatouage de certification. Nous avons présenté jusqu'à présent les différents mécanismes mis en œuvre pour se prémunir contre les attaques sur une image au niveau de la fonction d'insertion. Nous avons également présenté la méthode de génération des paramètres secrets destinée à injecter de la variété et diminuer au maximum le niveau de corrélation entre blocs d'une même image.

L'algorithme de tatouage de certification apporte *a priori* toutes les assurances nécessaires pour éviter une attaque sur une image de manière isolée. Nous avons vu également qu'un certain nombre de points techniques ont été

mis en place afin de se prémunir contre un attaquant qui aurait à sa disposition un groupe d'images certifiées pour effectuer une cryptanalyse d'ordre statistique, par exemple. Toutefois, quelque soit la qualité des composants développés contre ce type d'attaque, un des rouages essentiels de la sécurité réside dans la capacité à utiliser des clefs secrètes différentes selon l'utilisateur. Ce problème et celui de l'authentification de l'auteur et sera abordé dans le chapitre A.

2.5 Résultats expérimentaux

Dans la section précédente, nous avons spécifié et implémenté un algorithme de tatouage de certification. La qualité globale de cet algorithme dépend de la satisfaction de plusieurs objectifs : la robustesse, l'invisibilité, la localisation, la complexité et la sécurité. Nous avons vu dans la section 2.4.2.1 que le seuil conditionnant la sélection ou l'exclusion d'un bloc pour le tatouage est un paramètre important de l'algorithme. Il doit être choisi judicieusement pour optimiser l'ensemble des objectifs. Or les images peuvent être visuellement très variables, allant des images très homogènes (avec peu de détails, donc fortement compressées) à des images très chaotiques (avec beaucoup de détails, donc faiblement compressées).

Dans un premier temps, nous allons analyser l'influence du seuil sur la sélection des blocs en fonctions des images du jeu d'essai. Nous déduirons de ce analyse un intervalle de seuil acceptable pour être adapté à la diversité des images. Dans un second temps, nous analyserons la propriété d'imperceptibilité, qui aura son influence sur le choix d'un seuil définitif. Enfin, en utilisant cette valeur de seuil, dite idéale, nous continuerons à étudier chacun des objectifs du tatouage de certification, pour mesurer comment ils sont satisfait.

2.5.1 Présentation du jeu d'essai

Pour analyser les résultats de l'algorithme de certification, nous avons réalisé 22 photos pour composer le jeu d'essai (figure 2.15). Cet échantillon comprend des photos homogènes, détaillées et quelques cas extrêmes. La photo 1 est presque uniformément blanche, donc très favorable à la compression. La photo 22, bien que visuellement régulière, est très chaotique si l'on compare les blocs entre eux. Ce qui explique un taux de compression faible. Le taux de compression est une mesure indépendante du tatouage. Une image avant ou après le tatouage présente quasiment le même taux de compression. Les taux de compression sont présentés dans le tableau 2.2.

2.5.2 Seuil

Le seuil est un élément important de notre algorithme. Choisir le seuil, revient à effectuer le meilleur compromis pour préserver en priorité la sécurité.

Figure 2.15 – Jeu d'essai

Le seuil permet de déterminer quels sont les blocs qui ne sont pas adaptés au tatouage. En effet, tatouer un bloc homogène revient à laisser fuir de l'information sur les éléments ajoutés par le tatouage. Le seuil est une valeur délicate à choisir. Plus le seuil est élevé, plus on va tatouer de bloc, au détriment de l'invisibilité et de la fuite d'information (la marque ajoutée à un bloc homogène est aisément reconnaissable et extractible). Un seuil trop élevé a donc des conséquences sur la sécurité. En revanche, un seuil trop faible n'implique pas assez de blocs tatoués. Ainsi, on n'est plus en mesure de garantir l'intégrité de cette image, puisque trop de blocs sont laissés sans surveillance. Là encore, cela impacte la sécurité.

La figure 2.16 fait ressortir le pourcentage de bloc seuillés (non tatoués) en fonction du seuil (nombre de zéros à partir duquel un bloc est considéré comme homogène).

Nous considérerons que 50% des blocs doivent être tatoués pour garantir l'intégrité d'une image. Cette valeur est liée à la contrainte de sécurité qui sera discutée plus loin. Cette figure montre clairement que pour un seuil de 20, 19 images sur 22 ($4/5$) ont moins de 50% de blocs tatoués. Seules les images 20, 21 et 22 (les plus détaillées) sont propres au tatouage. On constate également que la plupart des images sont très sensibles au seuil et présentent de fortes variations. Seules les images 1, 21 et 22 ont des courbes plutôt stables. Ce sont logiquement les extrêmes. L'image 1 est la plus homogène et quel que soit le seuil, elle est impropre au tatouage. En revanche, les 21 et 22 sont toujours en deçà des 50% quel que soit le seuil et pourront donc toujours être tatouées. En dehors de ces cas particuliers, il reste à déterminer la valeur optimale du seuil pour garantir l'intégrité au plus grand nombre. A l'examen des courbes, il apparait clairement qu'un seuil de 27 garantie un pourcentage de blocs tatoués supérieur à 50%. Seules trois images (1, 2 et 3) sont exclues par ce seuil. Ce sont les images très homogènes dont le taux de compression est très significativement supérieur aux autres images. On peut en déduire que la mesure du taux de compression permet une première estimation pour une détection simple et efficace des images impropres.

Si on scinde le tableau 2.2 en deux parties, alors on peut tracer les courbes 2.17 et 2.18 qui séparent ces deux types d'images. La figure 2.17 montre que les images à fort taux de compression se retrouvent dans le haut du graphique. Cela indique, comme on pouvait s'y attendre, que beaucoup de blocs sont seuillés. Les trois images dites "irréfutablement" homogènes (1, 2, 3) se retrouvent en haut du graphique et ont les plus forts taux de compression aussi. On note que le pourcentage de blocs homogènes chute lorsque le seuil excède 26. Les sept images restantes ne présentent pas visuellement un caractère flagrant d'homogénéité. Certes, certaines ont plus de zones homogènes que d'autres, mais dans l'ensemble, on s'attend à ce que ces photos soient des cas fréquents. En observant le seuil 23, on se rend compte que toutes ces images dépassent les 60% de blocs seuillés. C'est insuffisant pour permettre une validation correcte de l'intégrité d'une image (moins de 40% des blocs portent

Image	Taux	Image	Taux
22	9.92258	11	21.9231
21	13.3244	10	23.203
20	13.8575	9	23.6799
19	17.1566	8	25.3072
18	17.9169	7	26.3849
17	18.7673	6	30.3436
16	19.8075	5	30.4639
15	20.8654	4	32.7614
14	21.3142	3	43.8609
13	21.359	2	64.258
12	21.6941	1	173.795

Table 2.2 – Taux de compression des images

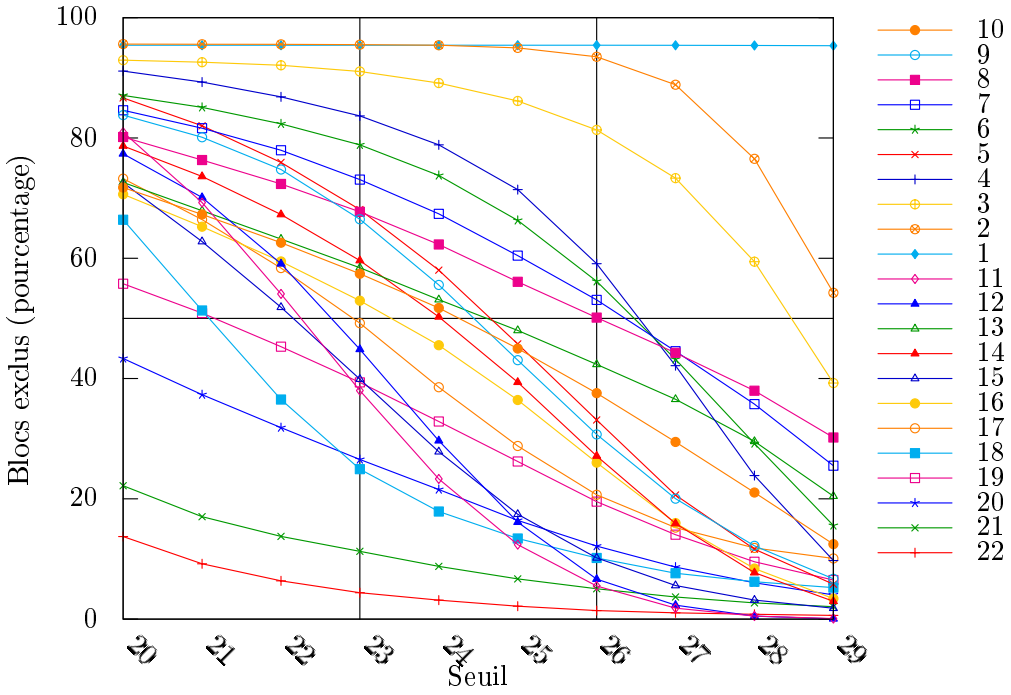


Figure 2.16 – Blocs seuillés pour les 22 images du jeu d'essai

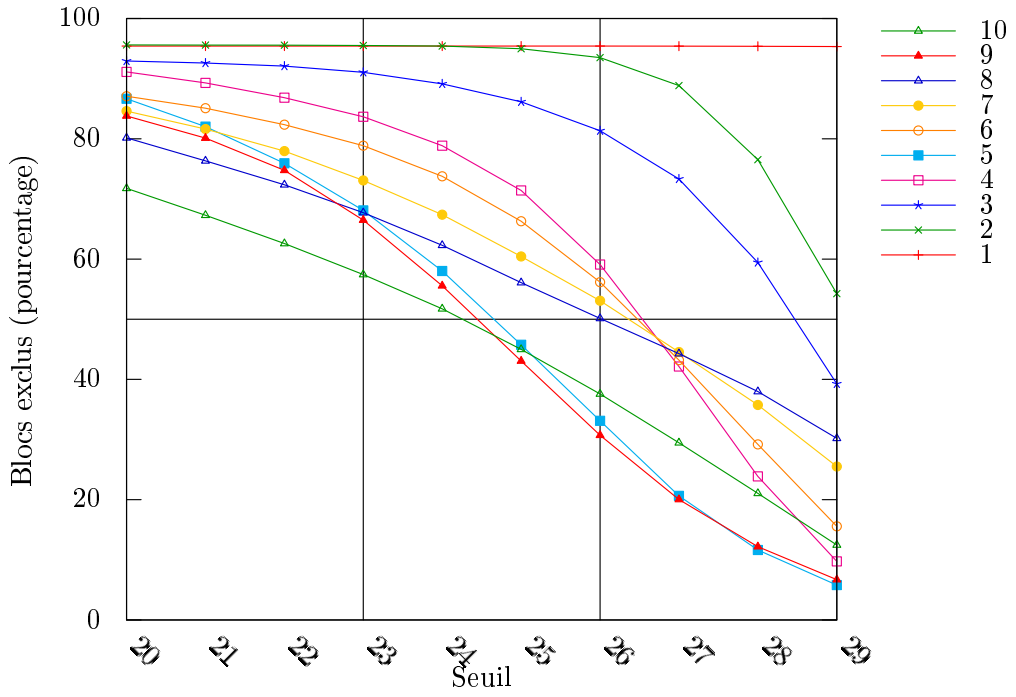


Figure 2.17 – Blocs seuillés pour 11 images à fort taux de compression

une information. En revanche, à la valeur 27, toutes les images sont sous la barre des 50%, même si certaines sont à peine inférieures à la limite. Pour les images présentant les plus fortes variations (images 4,5,6,7 et 9), les courbes présentent des points d'inflexion situés entre les valeurs 25 et 27. Au delà de cette limite, augmenter le seuil apportera des gains de moins en moins significatifs.

La figure 2.18 confirme nos observations précédentes. Pour ces images à fort taux de compression, les points d'inflexion apparaissent dès le seuil 23. Toutes les images ont un taux de blocs tatoués supérieur à 50% à partir de 26.

Globalement, entre les seuils 20 et 23, la majorité des images sont insuffisamment tatouées et ne présente aucun intérêt. L'intervalle $[23, 26]$ est certes plus intéressant, mais il est le lieu de très fortes variations du nombre de blocs tatoués. Il est difficile de définir une valeur de seuil satisfaisante pour toute les images, ou pour définir un modèle qui permette de l'adapter automatiquement. En revanche, l'intervalle $[26, 29]$ place la plupart des images à un niveau de tatouage acceptable, avec une variation plus linéaire.

Si l'on a mesuré l'impact du seuil sur l'intégrité de l'image à travers le nombre de blocs tatoués, il reste à évaluer l'impact du seuil sur l'invisibilité des marques. On a vu que plus le seuil est élevé, plus cette intégrité est garantie,

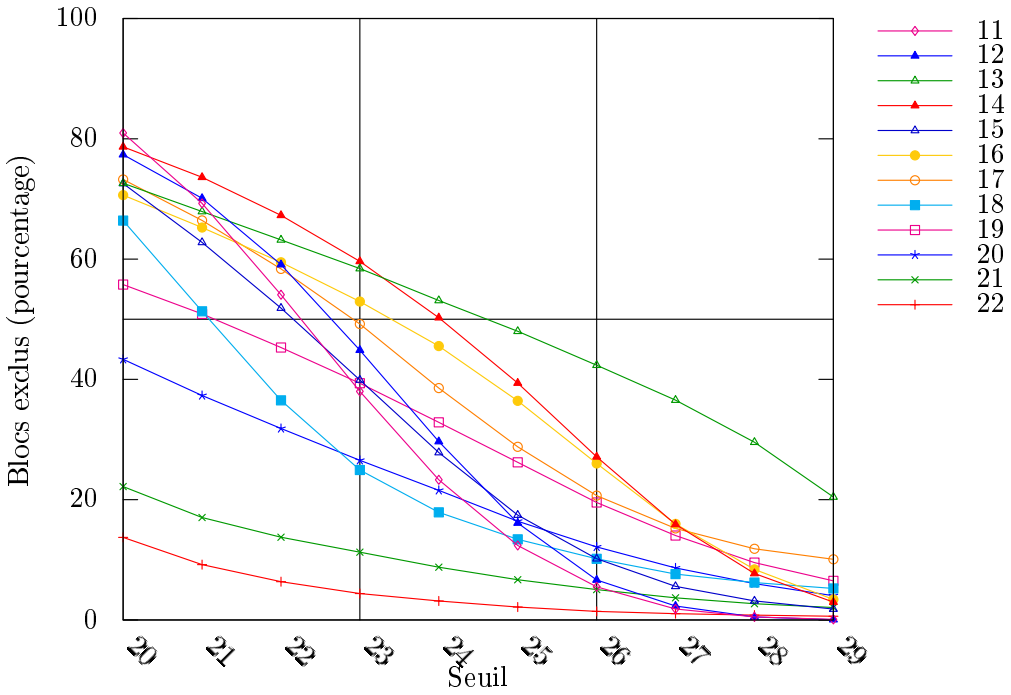


Figure 2.18 – Blocs seuillés pour 11 images à faible taux de compression

mais l'ont sait que plus ce seuil est élevé (donc plus le nombre de zéro est grand), plus l'information tatouée a un impact visuel. On ne peut donc pas simplement prendre la plus grande valeur. Pour affiner l'estimation du seuil idéal, nous devons examiner le critère d'invisibilité.

2.5.3 Invisibilité

L'impact visuel est mesuré par le PSNR. Cette mesure exprime le rapport entre la puissance maximale d'un signal et la puissance du bruit de corruption affectant la fidélité de la représentation. Les valeurs typiques trouvées en compression d'images se trouvent dans la fourchette de 30 à 50dB (Les valeurs hautes étant les meilleures). Nous avons calculé le PSNR de chaque image à chacun des seuils envisagés.

La figure 2.3 montre l'évolution du PSNR. Cette évolution est globalement identique et régulière. Elle ne montre pas d'inflexion significative qui mettrait en évidence un seuil caractéristique au delà duquel la dégradation est trop importante. Les images 2 et 3 possèdent quand à elles une inflexion, mais nous avons vu que ces images sont particulièrement homogènes, donc de toute façon peu favorables au tatouage de certification. Le PSNR minimal observé est d'environ 45dB. Aucune image, et quel que soit le seuil, n'est dégradée significativement du point de vue de l'invisibilité. Toutes les valeurs du seuil sont

acceptables.

L'étude de l'intégrité nous a conduit à privilégier les seuils les plus élevés. A l'inverse, l'étude du PSNR n'exclu aucun seuil et a plutôt tendance à privilégier les valeurs de seuil les plus petites, correspondant à une qualité d'image la moins perturbée. En conséquence, la valeur idéale que nous conserverons tout au long des tests présentés ci-après sera 27, qui établit le meilleur compromis entre intégrité et qualité.

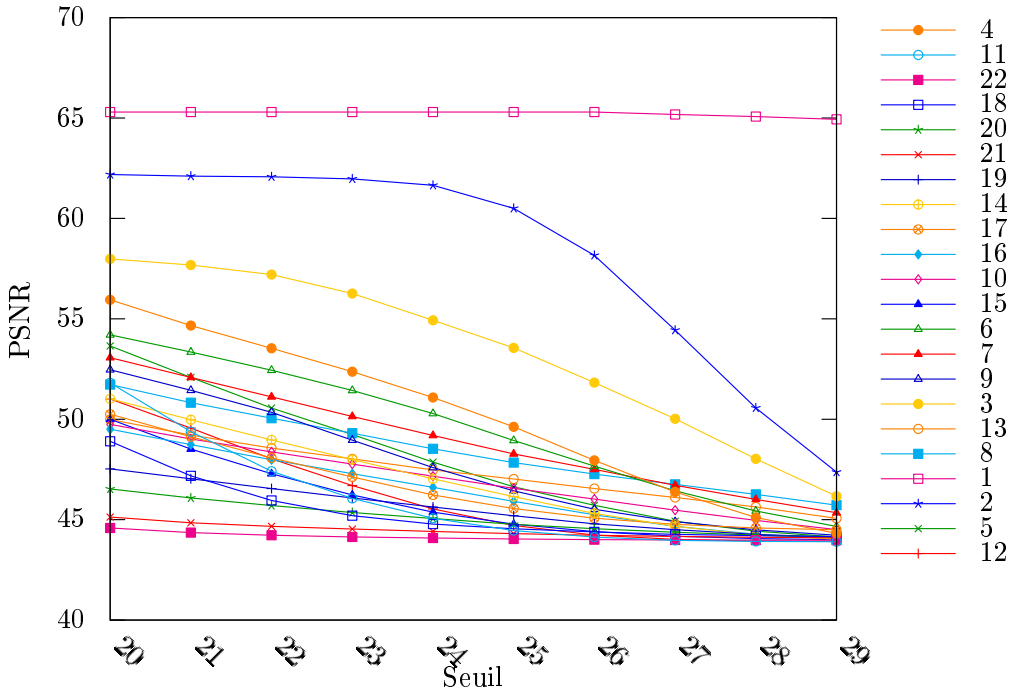


Table 2.3 – PSNR en fonction du seuil

Nous présentons maintenant une comparaison du critère d'invisibilité entre des images extraites du catalogue standard des images, utilisé en traitement d'image, avec des images capturées depuis un équipement mobile. L'intention est de vérifier s'il existe un écart dans le comportement de l'algorithme selon le procédé d'acquisition. Les figures 2.19 nous permettent d'apprécier de manière qualitative les performances en terme d'invisibilité. Les images de référence utilisées sont :

- Babouin
- Lena
- Poivrons
- Deux photo prise par un mobile.

Les paramètres de compression sont Q=85 sous-échantillonnage horizontal et vertical pour toutes les images. (Figures 2.19, 2.5.3 et 2.5.3)

Le calcul du PSNR (Tableau 2.4 - Formule 2.6) nous donne les résultats

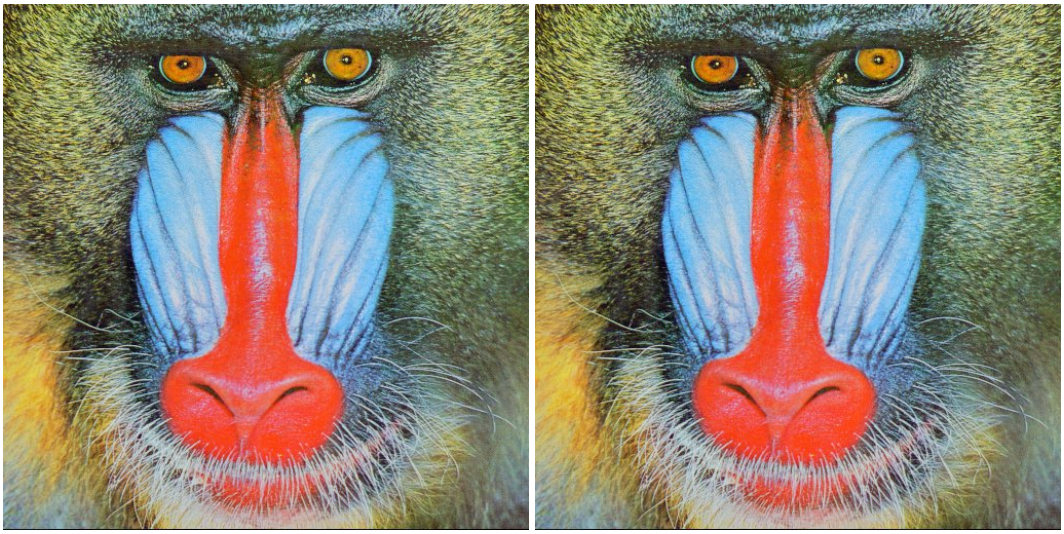


Figure 2.19 – Invisibilité (de gauche à droite) : image originale, image tatouée

Figure 2.20 – Invisibilité (de gauche à droite) : image originale, image tatouée

Figure 2.21 – Invisibilité (de gauche à droite) : image originale, image tatouée

suivants :

Nom	Babouin	Lena	Poivron	Image 4	Image 5
PSNR (en dB)	43.82	44.18	44.15	45.71	47.95

Table 2.4 – Tableau des PSNR

Le PSNR moyen est de : $45,16$ dB. Plus le PSNR est élevé moins la distorsion est importante. On remarquera donc que ce coefficient supérieur a 38dB est très bon pour un algorithme de tatouage. L'image 4, sans être anormalement dégradée, présente l'écart le plus important avec le PSNR moyen. Encore une fois, il s'agit d'une image assez homogène, dont la structure particulière explique ce comportement. En conclusion, l'algorithme de tatouage, avec un seuil idéal de 27, satisfait pleinement l'objectif d'invisibilité nécessaire à la préservation de la qualité de l'image exigée par la certification.

2.5.4 Contrôle d'intégrité

Le contrôle d'intégrité doit répondre à plusieurs critères. Il doit permettre la détection des changements, mais doit aussi révéler l'utilisation de mauvaises clés. Enfin, l'un des objectifs du tatouage de certification est d'autoriser la copie conforme à quelques défauts près. Le contrôle d'intégrité doit permettre d'identifier ces défauts et d'en fournir assez d'informations (quantité, localisation) pour distinguer une copie conforme avec des défauts, d'une copie malicieusement modifiée. Ce sont ces trois points que nous allons nous attacher à décrire dans cette section. Mais auparavant, il est nécessaire de détailler comment le contrôleur d'intégrité analyse une image tatouée.

2.5.4.1 Principe de fonctionnement du contrôleur d'intégrité

Nous rappelons que le tatouage consiste à traiter une image non tatouée et déjà compressée. Cette image est partiellement décompressée (sans perte d'information) puis est traitée bloc par bloc. Parmi ces blocs, l'algorithme de tatouage va distinguer deux cas : les blocs qui contiendront des informations sur les méta-données (blocs messages) et les blocs qui contiendront des informations sur l'intégrité. Pour les blocs destinés à l'intégrité, on distingue deux sous-cas : les blocs seuillés qui ne seront pas tatoués et les blocs tatoués qui contiendront une partie du motif utilisé pour le contrôle d'intégrité. Par ailleurs, différentes clés sont utilisées dans les opérations de tatouage .

Le contrôleur d'intégrité est une fonction qui reçoit en entrée une image tatouée et le système de clés utilisé lors de son tatouage. La détection de l'intégrité procède selon un raisonnement quasiment similaire à celui du tatouage. Le décodeur considère un bloc et sa catégorie (message ou motif). Dans le cas d'un bloc message, les cinq bits inclus dans ce bloc sont ajoutés à la suite du message déjà décodé. Dans le cas d'un bloc d'intégrité, le décodeur vérifie si ce bloc répond au critère de seuil. Si le bloc est seuillé, cela indique qu'aucun motif ne pourra être lu dans ce bloc. Il est donc ignoré. En revanche, s'il n'est pas seuillé, le décodeur extrait la marque et la compare avec le motif attendu. Si cela est vérifié, le bloc est intègre, sinon il est erroné. Ainsi, le décodeur est en mesure de fournir une image distinguant les blocs message, les blocs seuillés, les blocs vérifiant l'intégrité et ceux ne la vérifiant pas. Nous insistons sur le fait que ce décodeur particulier n'est disponible que pour le tiers de confiance (CodaSystem) ou le juge.

2.5.4.2 Intégrité du tatouage

Il s'agit ici de vérifier la sensibilité du contrôleur à différentes modifications d'une image. On suppose ici que l'attaquant qui modifie l'image ne dispose pas des clés. Les tests ont été effectués de la manière suivante :

- Tatouer une image ;
- Modification par ajout de logos sur la version certifiée et recompression de la nouvelle image ;
- Extraction et affichage des blocs d'intégrité de la nouvelle image.

On remarque dans les images présentée dans les figures 2.22, que différents cas se présentent. Le plus évident est la modification du ciel (logo A). On remarque ici un point important, car si l'on pouvait lever des interrogations sur le tatouage d'une zone seuillée, nous avons ici un contre exemple. En effet, la région de l'image où le logo A a été ajoutée est seuillée. Ces blocs ne contiennent pas de marque d'intégrité. Puis, l'ajout de la marque a introduit suffisamment de détails pour que cette même zone soit maintenant considérée comme non seuillée. Le contrôleur recherche une marque d'intégrité que l'attaquant n'a pas pu fournir sans la connaissance des clés.

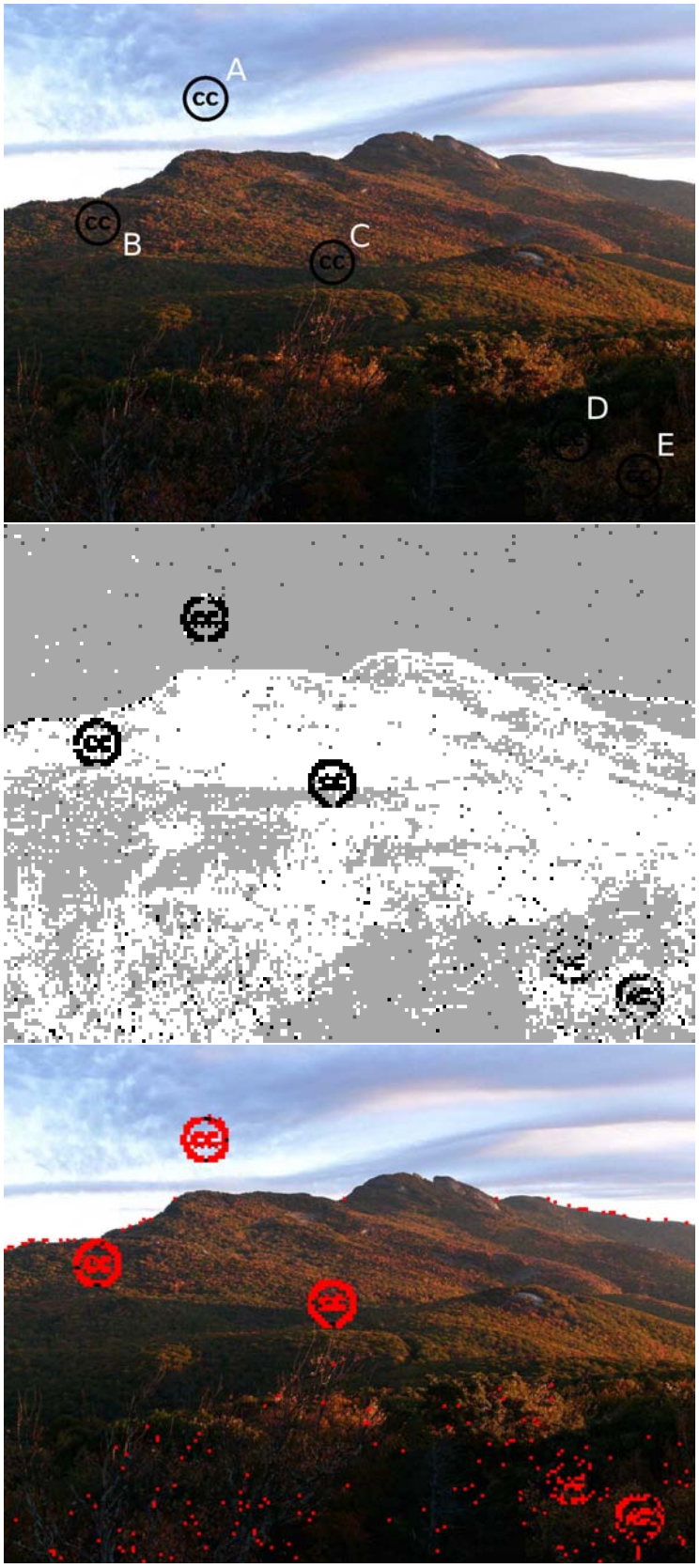


Figure 2.22 – Contrôle de l'intégrité (de haut en bas) : image modifiée, détail des blocs modifiées, superposition.
Légende : noir/rouge - non intègre ; blanc - intègre ; gris : exclus (NB : les ne sont là qu'a titre informatif et ne sont pas a considérer comme une modification)

Deux autres logos sont insérés dans des zones très détaillées (logos B et C). Ceux-ci sont détectés sans problème également. En revanche, on remarque que l'un des logos (logo D) n'est pas correctement détecté. Si l'on observe la carte des détails (Figure 2.22), nous remarquons que le logo se trouve dans une zone sensiblement homogène. De plus, cette zone est notablement foncée, en tout cas d'une couleur similaire à celle du logo inséré. Ces deux facteurs influent sur la détection partielle du logo à cette position. Le logo E se confond assez bien avec l'image, même s'il reste visible. Et même si l'œil ne les perçoit pas tous, les détails sont suffisants pour le contrôleur. Il détecte suffisamment de blocs non intègres et les fait apparaître nettement le logo. Par ailleurs, on observe quelques fausses détections, généralement en bordure, dues au réencodage de l'image tatouée (voir section 2.5.4.4).

Sur la figure 2.23, nous avons réaliser une insertion d'un log noir et blanc plus imposant (par rapport à la taille de l'image). On remarque que l'algorithme ne détecte pas l'intérieur du logo inséré. Cela est dû au fait que celui-ci est homogène. Cependant, les bords sont détectés car ils se trouvent entre plusieurs blocs. Le cas échéant, un logo étant parfaitement synchronisé avec les blocs et lui même homogène, ne serait pas détecté. De la même manière, remplacer des zones homogènes par d'autres zones homogènes est réalisable. Cependant, il y a peu de chances pour que cela soit réellement utile à un attaquant, car ne pas ajouter de détail n'apporte pas d'intérêt. On peut aussi tenter de gommer des détails, en les remplaçant par des blocs homogènes. Cette attaque peut-être mitigée en utilisant une technique de comptage, qui n'a pas été mis en œuvre ici, mais consistant à insérer après une succession de blocs homogènes un bloc contenant le nombre de blocs seuillés de la série. Ainsi, ajouter des blocs homogènes serait détecté par une incohérence lors du comptage.

Une attaque possible sur le tatouage est proposée par Holliman-Memon [Holliman 2000]. Elle consiste à composer une base suffisamment grande d'images tatouées, puis à composer une image en sélectionnant parmi cette banque de données des blocs ressemblants au bloc à insérer dans l'image à composer. L'idée est de ne prendre que des blocs connus pour posséder une marque pour tenter de tromper le détecteur. Dans notre cas, nous limitons l'impact de cette attaque sans pour autant la supprimer complètement. En effet, et dans un premier temps, la banque d'images doit être composée d'images tatouées avec les mêmes clefs. Dans un second temps, si l'on est à même de respecter cette condition, l'attaquant devra connaître la liste des blocs messages. Si ce n'est pas le cas, il ne pourra pas reconstituer des méta-données ayant un sens, sauf si ces méta-données sont exactement les mêmes dans chaque image. Or, pour ce dernier point, il lui faut un accès au codeur pour choisir son message. En effet, pour que la copie des blocs message ait un sens, il faut que ceux-ci contiennent un message connu. Dans le cas de notre banque d'images, il faudra donc que ce message soit le même pour toutes les images si l'on souhaite pouvoir intégrer ce message particulier dans l'image forgée. Dans un troisième temps, s'il est en mesure de distinguer les blocs messages, il devra,

pour respecter le motif du bloc i, choisir uniquement parmi les blocs de même numéro i dans sa banque d'images. En effet, un bloc j d'une autre image ne fonctionnera pas 1 chance sur 32 voir 2.5.6), puisque le motif est dépendant du numéro de bloc.

Le résultat se lit sur les images suivantes (Figure 2.23).

2.5.4.3 Décodage avec de fausses clés

Nous souhaitons tester et vérifier la validité des techniques utilisées pour la personnalisation du tatouage. Pour cela, nous allons reproduire une attaque fictive. Il suffit pour cela de changer les clefs utilisées lors de l'encodage ou du décodage.

Dans ce scénario, on suit la procédure suivante :

– Tatouer une image avec une fausse clé parmi les quatre ;
– On fait le contrôle d'intégrité avec les bonnes clefs.

Les résultats se lisent sur les figures 2.24.

Figure 2.23 – Contrôle de l'intégrité (de gauche à droite) : image certifiée, image modifiée, test de l'intégrité.
Légende : noir - bloc non intègre ; blanc - bloc intègre ; gris : bloc exclus

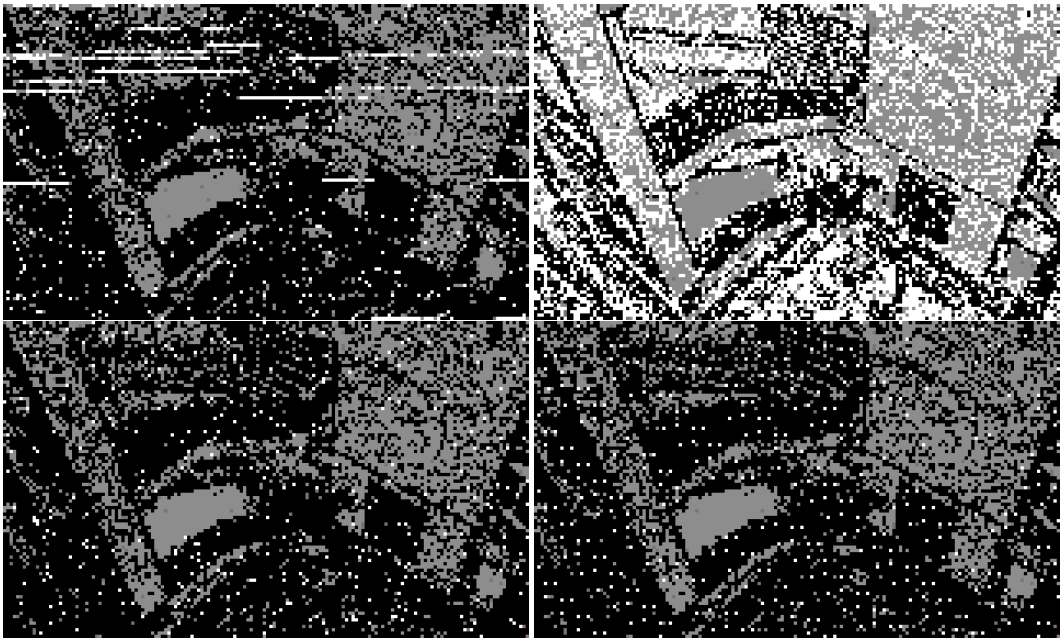


Figure 2.24 – Contrôle de l'intégrité avec de fausses clefs (de gauche à droite, puis de haut en bas) erreur sur la clef de : sélection pseudo-aléatoire, caractéristique, motif, insertion.
Légende : noir - bloc non intègre ; blanc - bloc intègre ; gris : bloc exclus

Ces contrôles nous montrent que le choix des systèmes implémentés se révèlent judicieux.

2.5.4.4 Réencodage

Dans cette section nous prenons une image tatouée, que nous allons éditer sans modification de pixels. En revanche, nous souhaitons étudier l'influence de la recompression sur le tatouage. Premièrement, faire une différence entre une image et une image réencodée est nulle. Cela s'explique facilement par le fait qu'une fois la qualité perdue, elle ne se recalcule pas dans le bitmap décompressé. En revanche, notre algorithme introduit une variation dans les coefficients DCT. Ainsi, nous allons décoder une image qui n'est pas exactement l'originale. Il se peut alors que lors de la recompression, une partie de l'information que l'on a insérée par tatouage soit perdue lors de la création du JPEG.

Le premier réencodage se fait sans modification de qualité. La figure 2.25 montre le réencodage de la l'image 6. La mesure du nombre de blocs non intègres est de 5%. La plupart des zones impactées sont des bordures. L'explication de cette particularité est logique. En effet, les bordures des éléments de l'image sont les lieux où les blocs présentent la plus grande variation. Ils sont donc les plus sensibles à la quantification et aux arrondis réalisés lors des

étapes de la compression.

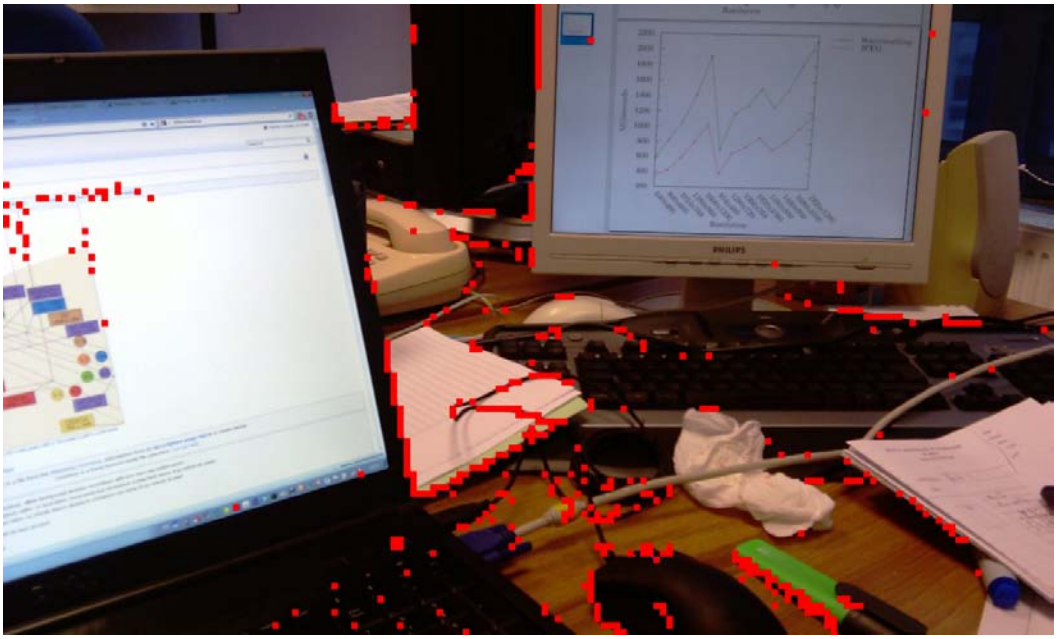


Figure 2.25 – Image 5 réencodée sans modification de qualité

Si l'on modifie cette image en augmentant de 1 le paramètre de qualité de compression, c'est-à-dire que l'on cherche à réduire encore la taille de l'image, les erreurs d'intégrité sont très importantes (figure 2.26). En l'occurrence, le nombre de blocs non intègres est de $90,35\%$. La figure 2.27 montre cette même image avec, à l'inverse, une diminution de la qualité, où le paramètre de compression est réduit de 1. Le nombre de blocs non intègres est de $48,70\%$. Même s'il est moindre, le taux d'erreurs est très élevé et disqualifie l'image. La figure 2.28 diminue le paramètre de compression de 2. La mesure du nombre de blocs non intègres est de $78,01\%$.

Le paramètre de compression de l'original nécessite un respect strict pour garantir au mieux la conformité de l'image. Mais la question de « la valeur d'une copie conforme » se pose malgré tout. Si le réencodage fidèle de l'image 6 présente un taux d'erreur relativement faible (5%), et une répartition de ces erreurs qui peut s'expliquer d'après la structure de l'image, nous vérifions le taux d'erreur sur les copies conformes des autres images de l'échantillon. Le tableau 2.5 présente les mesures effectuées. Les images 1 et 2 ne rentrent pas dans la mesure car leur blocs étant quasiment à 100% seuillés, aucun bloc ne peut servir à déterminer la validité du tatouage réencodé. Les mesures de ce tableau montrent que (excepté l'image 13 qui révèle une valeur extrême) le maximum est de 8%. Cette image comporte des zones à forte luminosité.

Figure 2.26 – Image 6 réencodée avec une qualité originale plus 1

Figure 2.27 – Image 6 réencodée avec une qualité originale moins 1

Figure 2.28 – Image 6 réencodée avec une qualité originale moins 2

Image	Pourcentage	Image	Pourcentage
2	N/A	21	2.395586
1	N/A	12	2.753146
11	0.007046	3	2.897238
5	0.276980	16	4.257271
10	0.417353	7	4.817650
9	0.789656	20	4.987853
19	0.985798	4	5.586000
18	1.015417	8	5.912776
6	1.279948	22	7.318709
15	1.391266	17	8.368734
14	1.408582	13	14.628040

Table 2.5 – Pourcentage de blocs erronés après réencodage simple

2.5.4.5 Bilan

Les expériences de cette sous-section nous ont permis de valider le contrôle d'intégrité mis en œuvre. Nous avons vu qu'il est possible de localiser les modifications apportées à une image, ainsi que les cas qui peuvent s'avérer peu efficaces. Nous avons ensuite valider l'utilisation de clés en montrant que le choix d'une mauvaise clé implique un décodage fortement perturbé. Enfin, nous avons porté notre attention sur le concept de copie conforme. Nous avons pu montrer que la réalisation d'une copie conforme est possible, lorsque l'on respecte les paramètres d'encodage de l'image initiale. Les tests ont cependant montré qu'une dégradation inhérente à la recompression est toujours présente et dépend de la structure de l'image. Ainsi, le taux d'erreur se disperse entre 0% et 8.5%. En considérant la moyenne de 3.57%, nous en concluons qu'une copie conforme peut présenter un taux d'erreur d'environ 5%. Il ne faut cependant pas perdre de vue que toute modification n'est pas anodine et que la validation de la copie conforme doit être faite après analyse visuelle venant confirmer l'analyse quantitative automatique. En effet, la localisation chaotique ou respectant un motif régulier permet au juge de décider du caractère non-intentionnel ou malicieux de ces défauts et de rejeter ou non l'image comme une preuve.

2.5.5 Rapidité d'exécution

Le tableau 2.6 présente les temps d'exécution du tatouage de certification sur un HTC HD2 pour les images du jeu d'essai. Le temps comporte trois mesures : total d'exécution, total utilisé pour le tatouage et temps utilisé pour la compression et décompression JPEG cumulée. Cette dernière mesure ne prend en compte que le codage et décodage entropique nécessaire pour passer de l'image JPEG (issue de l'encodeur du mobile) aux coefficients DCT quantifiés, et *vice-versa*. Le HTC HD2 prend des photos de 5 MegaPixels à une résolution de 2592 x 1944 pixels. Son processeur est cadencé à 1 GHz.

On constate très clairement que le tatouage représente une grande part du traitement total, mais reste sensiblement le même quelle que soit l'image traitée. On remarque aussi que le temps d'exécution total est lié au temps de compression, lequel est lui-même lié au taux de compression de l'image (les images sont triées selon le taux de compression croissant). Le tatouage, lui, est indépendant du taux de compression. Globalement, les performances de l'algorithme pour une résolution d'image relativement élevée est compatible avec l'interactivité attendue (environ deux secondes en moyenne).

2.5.6 Estimation de la sécurité

Par construction, le message tatoué dans l'image est chiffré. La confidentialité de ce message repose sur la sécurité de l'algorithme de chiffrement et de la protection de la clé qui sert à ce chiffrement. Comme l'algorithme de

Image	Total (ms)	Watermarking (ms)	JPEG (codage et decodage)
1	1787	1241	546
2	1855	1242	613
3	1891	1243	648
4	1936	1244	692
5	1929	1226	703
6	1930	1232	698
7	1960	1221	739
8	1985	1236	749
9	1973	1218	755
10	1972	1226	746
11	2074	1272	802
12	2009	1252	757
13	2019	1239	780
14	2003	1229	774
15	2014	1224	790
16	2013	1227	786
17	2014	1222	792
18	2039	1225	814
19	2041	1219	822
20	2089	1209	880
21	2134	1225	909
22	2194	1193	1001

Table 2.6 – Exécution comparée sur les images du jeu d’essai

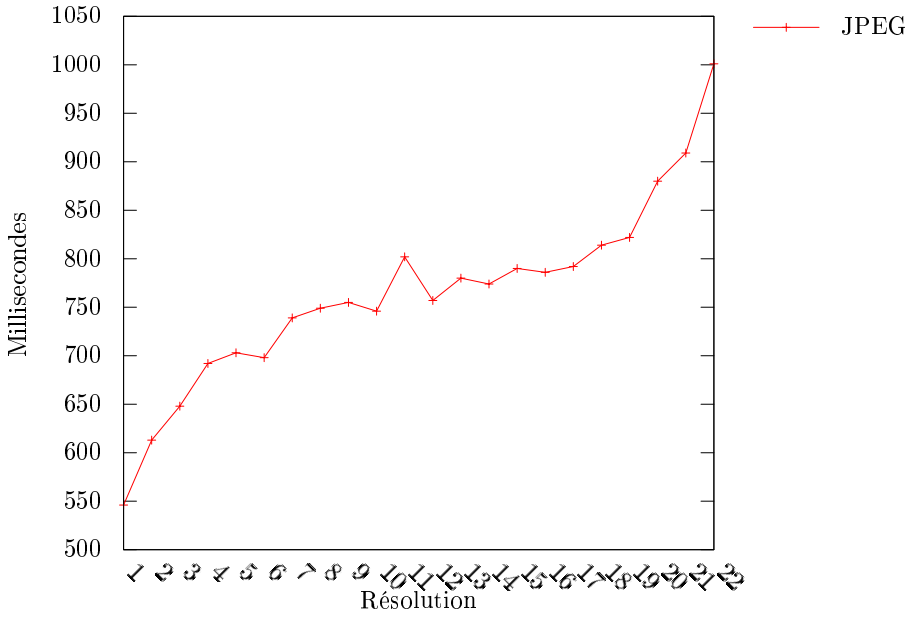


Figure 2.29 – Représentation des temps d'exécution du tatouage en fonction des images

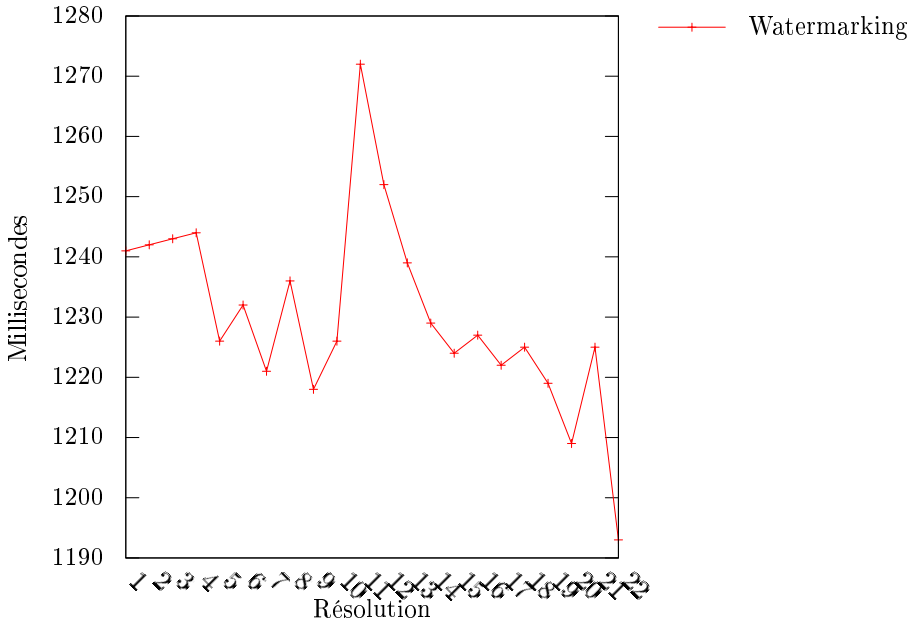


Figure 2.30 – Représentation des temps d'exécution JPEG en fonction des images

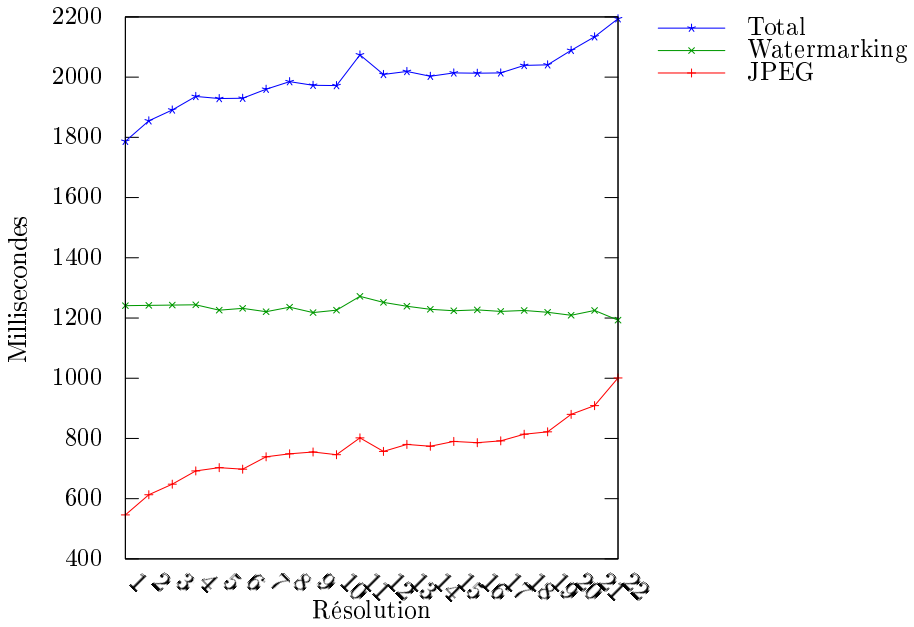


Figure 2.31 – Représentation des temps d'exécution combinés en fonction des images

chiffrement est un module indépendant dans la méthode proposée, alors, si l'on estime que la confidentialité nécessite une protection plus élevée, il suffit de choisir un algorithme de chiffrement plus sûr. La seule condition que ce dernier doit satisfaire est de respecter les contraintes de complexité induites par l'environnement mobile. Du point de vue de l'attaquant, vouloir prendre connaissance du contenu du message revient à décrypter le message. L'attaquant se heurte alors à la complexité des attaques connues sur l'algorithme de chiffrement choisi.

En revanche, l'attaquant peut avoir l'intention d'identifier uniquement quels sont les bits modifiés dans l'image. L'objectif étant d'utiliser cette information pour construire une nouvelle image certifiée. La vérification de l'authenticité d'une image repose sur la fonction $Auth$ définie dans la section 2.2.4.2. Cette fonction reçoit deux paramètres, l'image à authentifier et l'identifiant de son propriétaire. Grâce à ces informations, la fonction est capable de rechercher la clé secrète attachée à l'identité du propriétaire et de procéder au contrôle d'intégrité. Bien entendu, nous cette fonction $Auth$ ne doit pas révéler d'information quant à cette clé. Enfin, la fonction fournit un résultat booléen $Vrai$ si l'image est parfaitement authentique (sans la moindre erreur), ou $Faux$ si elle contient au moins un bloc non intègre. La localisation des blocs non intègres ne fait pas partie des informations restituées aux utilisateurs. La localisation des

erreurs est une fonctionnalité réservée à l'autorité de certification, et l'interprétation de ces erreurs est laissée à l'appréciation du juge en cas de besoin.

Dans ces conditions, l'attaquant ne réussit son attaque qu'à condition d'identifier tous les blocs d'intégrité d'une part et, d'autre part, de déterminer le bit qui a été modifié pour coder cette intégrité. L'attaquant connaît l'algorithme de tatouage et notamment la valeur du seuil utilisé pour exclure un bloc pour le tatouage d'intégrité (un bloc trop homogène). Il sait également que le bit à modifier dépend des MSB du bloc. Cependant, le choix de ce bit lors du tatouage dépend de clés dont l'attaquant n'a aucune connaissance. En définitive, pour un bloc donné, les 32 possibilités de modification des 31 LSB sont plausibles, avec équiprobabilité pour chacune des 32 possibilités. L'entropie est donc maximale.

La seule approche possible pour l'attaquant est une attaque exhaustive sur l'ensemble des blocs d'intégrité, avec pour chacun des blocs, les 32 possibilités. Le nombre N de possibilités est donc

$$N = 2^{5 \times nbi}$$

où nbi est le nombre de blocs d'intégrité. Il va de soi que plus le nombre de blocs d'intégrité est élevé, plus la complexité de l'attaque exhaustive est grande. Il convient donc d'assurer que ce nombre soit suffisamment grand pour résister suffisamment longtemps à la puissance de calcul des ordinateurs.

Nous avons considérer qu'au moins 50% des blocs d'une image devaient être tatoués. L'une des raisons de cette valeur est qu'elle garantit en l'occurrence une robustesse plus que suffisante.

On peut illustrer numériquement cette complexité sur une image de résolution 640×480 pixels, soit 4800 blocs de 8×8 pixels. Nous supposons qu'elle respecte à minima le taux de 50% de blocs potentiels pour le tatouage. Nous faisons également abstraction des blocs éventuellement utilisés pour insérer les métadonnées. Par conséquent, $nbi = 2400$. La complexité d'une attaque exhaustive est donc

$$N = 2^{5 \times 2400} = 2^{12000}$$

ce qui est très largement supérieur à la complexité suffisante pour conférer au tatouage la robustesse cryptographique nécessaire.

Si l'attaquant peut déterminer si un bloc est tatoué, il peut naturellement déterminer si un bloc n'est pas tatoué. Un bloc non tatoué est vulnérable. Mais à condition que la modification ne change pas le statut du bloc. En effet, si la modification apporte suffisamment de détails jusqu'à faire franchir à ce bloc le seuil d'exclusion, le détecteur contrôlera son intégrité qu'il ne trouvera évidemment pas. L'image ne sera pas authentifiée. Pour modifier un bloc non tatoué au delà du seuil, l'attaquant doit lui-même le tatouer, avec les clés dont il n'a pas connaissance par hypothèse. Néanmoins une modification minime est possible. Mais dans la mesure où elle n'ajoute pas de détail, les conséquences visuelles

sont elles aussi minimes, y compris pour l'attaquant. Nous faisons ici l'hypothèse que le sens d'une image, en tout cas sa valeur « légale », réside dans les détails qu'elle porte et non dans les zones homogènes.

La limite de 50% de blocs tatoués que nous avons fixé empiriquement est donc largement supérieure à ce que la robustesse cryptographique exige, et ceci est d'autant plus vrai que la résolution de l'image est élevée. Une limite inférieure peut mathématiquement satisfaire la robustesse cryptographique exigée. Mais compte tenu des modifications possibles sur les blocs non tatoués, il convient de fixer une limite moyenne à partir de laquelle l'utilisateur sera informé de la structure sensible de l'image. Encore une fois, il ne s'agit pas d'empêcher le tatouage en fonction du nombre de blocs tatoués, mais de donner les informations qui permettent à l'utilisateur de faire ses propres choix en toute confiance.

2.5.7 Conclusion de l'étude expérimentale

Nous avons vu plusieurs résultats dans cette section. Ainsi, nous avons d'abord regardé le taux de compression des images. En procédant à une classification selon le taux de compression nous avons pu montrer qu'il est possible de distinguer des images impropres au tatouage. Cependant, cette distinction se dégrade pour des images plus réelles. Nous avons donc ensuite détaillé l'influence du seuil sur la quantité de blocs tatoués et avons déterminé expérimentalement que ce dernier devait être situé dans l'intervalle $[27; 29]$. Ensuite, nous avons étudié la visibilité du tatouage en mesurant le PSNR. Les mesures recommandent un seuil le plus petit possible pour un PSNR maximum. En fixant le seuil à 27 nous arrivons à un compromis acceptable tant en terme d'invisibilité qu'en terme de blocs tatoués (la majorités des images ont plus de la moitié des blocs tatoués).

Nous avons ensuite cherché à caractériser une copie conforme. Plusieurs tests nous ont permis de voir que le changement des paramètres de quantification se traduis par une perte très importante de la marque. En revanche, utiliser le même paramètre de quantification nous donne des résultats exploitables. Les tests de réencodage ont permis de montrer que pour toutes les classes d'images acceptables, le taux d'erreur apporté par le réencodage varie de 0% à 8.5% avec une moyenne de $3,57\%$. Une valeur de 5% donne une bonne idée de la possibilité d'avoir une copie correcte. Bien entendu, toute déviation par rapport à l'originale est significative et doit être relevée.

L'analyse des temps d'exécution sur une image de 2592×1944, ont montré un étalement du processus de tatouage du jeu d'essai allant de $1,7sec$ à $2,2sec$. Cette variation étant apportée majoritairement par le processus JPEG, qui varie de $0,55sec$ à $1sec$ alors que le tatouage reste stable dans l'absolu avec un maximum de $1,2sec$.

Enfin, nous avons analysé la robustesse cryptographique. Nous avons vu que l'algorithme résiste de manière très satisfaisante à une attaque par force

brute. Nous avons aussi montré que certaines attaques connues sont réalisables, mais que les techniques utilisée dans ce tatouage imposent des conditions très particulières, ainsi que la réalisation de banques d'images dédiées aux utilisateurs, mobiles et/ou fournisseurs.

2.6 Conclusion

Il existe de nombreux travaux sur le tatouage, qui correspondent à des contextes très variés. Les algorithmes de tatouage sont caractérisés par de nombreuses propriétés que l'on a rappelé dans la section 2.2.1. Le contexte qui nous intéresse est celui de la certification d'image. Nous avons décrit les principales attaques qui menacent le tatouage. Nous avons proposé une classification des attaques en fonction du contexte et des objectifs de l'attaquant. Nous en avons déduit les propriétés du tatouage de certification. Nous avons proposé un algorithme qui satisfait ces propriétés. Cet algorithme est un algorithme semi-fragile basé sur l'algorithme F5, lequel résiste à la recompression JPEG. Deux types d'informations sont tatouées. D'abord les métadonnées contenant l'identité de l'auteur, l'horodatage et la géolocalisation de l'image. Ensuite, les données d'intégrité qui permettent de garantir autant que possible l'authenticité de l'image. Pour tatouer ces différents types d'informations, plusieurs clés de chiffrement sont utilisées et constituent l'essentiel de la protection. Nous avons réalisé des expérimentations sur un échantillon de 22 images. L'analyse de ces expérimentations a permis de préciser les paramètres de l'algorithme et renseigner au mieux l'utilisateur sur les conditions de ce tatouage.

En conclusion, l'algorithme de tatouage présenté garantit une sécurité nécessaire et suffisante à la certification d'une image. Mais il n'est qu'un maillon dans la chaine de certification. Ce maillon est consécutif à d'autres maillons aussi importants comme la collecte des métadonnées. En effet, tatouer, même de manière fiable, des informations erronées (fortuitement ou malicieusement) remet en question la force probante de l'image et rend caduque toutes les autres étapes de la chaine de certification. Il convient donc d'accorder un soin tout particulier à l'acquisition de ces métadonnées. En l'occurrence, on sait que les systèmes de navigation globaux GPS, GLONASS, etc.) ne garantissent pas actuellement une qualité d'information optimale. Ils sont vulnérables à différents types de perturbations et de manipulations. Or, comme nous l'avons montré au début de ce chapitre, l'attaquant peut éventuellement forger des alibis en tatouant une géolocalisation qu'il aurait malicieusement modifiée. Par conséquent, il convient de traiter cet aspect avec beaucoup d'attention pour là encore informer l'utilisateur sur les conditions de collecte, voire pour empêcher toute manipulation frauduleuse de la géolocalisation. La chapitre 3 décrit le fonctionnement du système GPS, ainsi que ses propriétés et ses vulnérabilités. Nous proposerons dans ce chapitre des solutions graduelles qui informeront l'utilisateur des anomalies détectées, et éventuellement qui pourront corriger

les défaillances du système.

Chapitre 3

Authentification de la géolocalisation

Sommaire

Nous avons décrit dans le chapitre précédent une technique de tatouage d'image qui nous permet de créer des photos certifiées, autrement dit qui ont la valeur de preuve. Cette force probante est le résultat de la combinaison de plusieurs éléments. Si l'identité de l'auteur et l'intégrité de l'image sont naturellement essentielles, la classification des attaques, décrite dans le chapitre précédent, met en évidence le rôle primordial du lieu et du moment où la photographie a été prise. Il reste alors à embarquer ces informations, les méta-données, à l'intérieur de l'image dans des conditions qui permettent de satisfaire les usages prévus dans le contexte industriel donné.

L'acquisition de la géolocalisation est une vulnérabilité du système de certification. Ne pas utiliser une donnée fiable rend le procédé caduque. Nous verrons qu'il est possible de maîtriser une partie de l'environnement de géolocalisation. Comme nous le disions au chapitre 1, avoir un contrôle total de la sécurité est impossible. Cela est d'autant plus vrai pour les Global Navigation Satellite System - Système de navigation globale par satellites (GNSS) dont les spécifications, logiciels et matériels embarqués dans les satellites sont totalement hors de notre contrôle. À défaut de tout contrôler, il nous faut alerter

l'utilisateur. La combinaison des contrôles - pour éliminer au mieux les cas d'erreurs - et la sensibilisation de l'utilisateur à sa situation sont deux méthodes qui nous permettront de renforcer la sécurité concernant la géolocalisation.

Pour ce qui est du lieu, les équipements mobiles actuels sont maintenant systématiquement munis de puces GPS . La démocratisation de cette fonctionnalité a conduit au développement des multiples applications basées sur l'utilisation de la géolocalisation. On peut citer, à titre d'exemple, la fonction "magasin le plus proche" de nombreux sites de commerce, la localisation des photographies en ligne, les réseaux sociaux (Facebook, LinkedIn, Foursquare), le micro-blogging (Twitter), des expériences scientifiques ou encore certains jeux en lignes.

Ces différents services ne requièrent pas le même niveau de fiabilité [Lorenzo 2009]. Certains ne requièrent aucune exigence alors que d'autre nécessitent une fiabilité très importante. En effet, lorsqu'un piéton se promène en ville et cherche son chemin, une position GPS erronée, voire pas de GPS du tout, ne provoquera qu'un désagrément passager. En revanche, un véhicule de secours devant se rendre précipitamment sur le lieu d'un accident, ne peut souffrir du moindre retard. Et toute erreur, même minime, peut se traduire par des dégâts matériels ou des conséquences dramatiques sur des vies humaines.

Les attaques sur les outils de géolocalisation par satellite comme le GPS ne sont pas une problématique nouvelle. En effet, les attaques sont multiples : brouillage, *meaconing*, *spoofing* sont autant d'exemples à faibles ou moyens coûts d'exploitation. Pour cela il devient primordial de mettre en œuvre des outils et des méthodes pour réduire le risque engendré par ces vulnérabilités inhérentes au système et assurer la fiabilité des méta-données.

Actuellement plusieurs systèmes de positionnement sont opérationnels. Cependant, seul le GPS est global. Le système Russe GLONASS possède actuellement dix-neuf satellites mais ne fournit à ce jour qu'une couverture régionale.

Ces systèmes sont appelés « Systèmes de positionnement par satelittes » ou encore « systèmes de navigation par satellites (satnav) ». Lorsque la couverture d'un tel système est globale, il n'est pas rare de voir apparaître le terme GNSS (Global Navigation Satellite System). Les principaux GNSS sont : GPS, GLONASS et Galileo. Bien que seul GPS soit complètement opérationnel, les deux autres sont prévus pour fournir à terme une couverture globale.

Cette étude se propose d'analyser l'environnement disponible pour la certification de photographie et de mettre en œuvre les contrôles *ad hoc* permettant l'amélioration du positionnement ou la détection des problèmes. Les contraintes industrielles de cette étude nous ont limité au potentiel logiciel du smartphone de l'utilisateur (excluant de fait toute intervention sur l'infrastructure GPS, ou sur l'électronique embarquée dans le smartphone). Malgré cette limite, il existe néanmoins un certain nombre d'anomalies détectables. En effet, les GNSS sont basés sur des infrastructures que l'on peut surveiller. De plus, les smartphones, disposant de plus en plus d'équipements spécifiques, nous permettent de mettre en place des systèmes de détection d'erreurs. En-

fin, des infrastructures particulières dédiées à la surveillance et l'amélioration des GNSS sont autant de moyens à notre disposition pour détecter ou corriger des informations erronées.

L'étude de ces besoins a conduit la société CodaSystem, Nottingham Scientific Limited et l'Université de Metz (Université de Lorraine) à l'élaboration d'un projet européen nommé ATLAS (septième programme cadre). Ce projet repose sur ce problème d'authentification des données provenant des GNSS. L'objectif est de proposer une suite de services destinés à accroître la confiance que l'utilisateur final peut accorder à la géolocalisation délivrée par son smartphone. L'amélioration de la fiabilité de cette position est un autre objectif du projet.

Dans la section suivante nous décrirons l'architecture des principaux systèmes de positionnement par satellites. Ainsi nous mettrons en évidence les aspects techniques de ces infrastructures qui conduisent aux vulnérabilités de la géolocalisation. Cette section intègrera notamment l'analyse des menaces qui pèsent sur ces systèmes. Une fois les menaces identifiées, nous présenterons un état de l'art des contre-mesures existantes qui atténuent ces risques. Finalement, nous détaillerons nos propositions et les spécificités qui ont été intégrées dans le projet européen FP7 ATLAS .

3.1 Les Systèmes Globaux de Navigation par Satellite

3.1.1 Architecture des systèmes de géolocalisation

La géolocalisation par satellites est une opportunité majeure pour le développement d'une solution globale de positionnement. En effet, avec suffisamment de satellites déployés, il est facile de calculer une position depuis n'importe quel endroit du globe. Voici comment s'effectue le calcul d'une position, autrement appelée *solution*. Un récepteur reçoit des signaux de plusieurs satellites dont il connaît la position exacte grâce à des almanachs. Chaque satellite diffuse la date courante. Ainsi à la réception, le récepteur calcul la différence de temps pour connaître sa distance d_i au satellite i. Avec le premier satellite, le récepteur peut se trouver n'importe où sur la sphère de rayon d_1. Dans des conditions idéales, les sphères de rayons d_2, d_3, d_4 calculées de la même manière que pour le premier satellite ont une intersection avec cette première sphère. Deux satellites permettent de réduire l'ensemble des positions à l'intersection des deux sphères, soit un cercle (figure 3.1). Avec une autre sphère, on réduit les possibilités à deux points (figure 3.2). Enfin, l'usage d'un quatrième satellite nous donne la dernière sphère permettant d'avoir un unique point qui satisfait toutes les équations. Ce point nous indique notre position (voir figure 3.3).

Les principaux GNSS sont, en réalité, composés de plusieurs infrastructures. La constellation de satellites fournit le service de base de la géolocalisation. Mais d'autres infrastructures terrestres peuvent la compléter afin de fournir une qualité de service supplémentaire. Une description très détaillée du

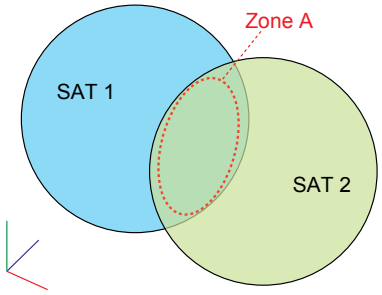


Figure 3.1 – Intersection de deux sphères

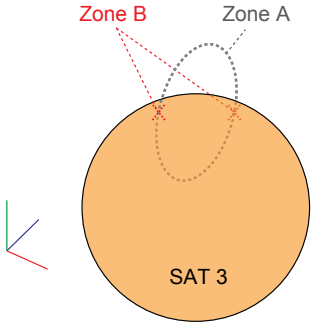


Figure 3.2 – Intersection de trois sphères

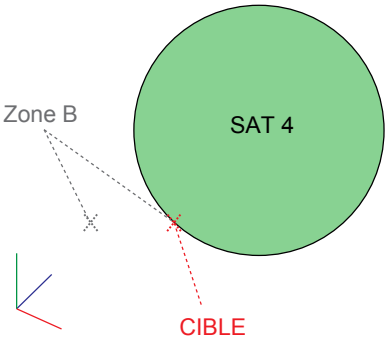


Figure 3.3 – Intersection de quatre sphères

fonctionnement de GPS, qui reprend les bases du positionnement par satellites, se trouve dans [Kaplan 1995] et [Hofmann-Wellenhof 1993].

3.1.1.1 Agencement de base

Chacun des trois systèmes de positionnement majeurs GPS, GLONASS, GALILEO) a été conçu pour des opérations particulières. Cela implique des architectures spécifiques que nous nous proposons de décrire maintenant.

GPS GPS a été mis en place pour le département de la défense américaine, qui reste encore à l'heure actuelle le responsable de la maintenance, de l'amélioration et de la gestion de la constellation. Le GPS Joint Program Office (JPO) est chargé d'assurer la disponibilité constante du service, principalement pour les militaires américains et leurs alliés, bien que la demande d'usages civils augmente.

GPS est constitué de trois structures matérielles, appelés segments : le segment espace, le segment contrôle, et le segment utilisateur.

Le segment espace, aussi appelé constellation, fait aussi partie intégrante d'un réseau de capteurs servant à la détection des explosions nucléaires. En dépit du succès et de l'âge du système, GPS n'a quasiment pas évolué depuis les années 1970. Les premiers satellites, formant le bloc I, furent lancés entre 1978 et 1985. Ces satellites de « développement » ont été lancés sur une orbite moyenne. Leur durée de vie était de cinq années. Le système opérationnel (bloc II), est composé de neuf satellites. Il a été renforcé par 19 satellites bloc IIA (Avancé), lancés entre 1989 et 1997. Ces satellites ont une durée de vie de sept ans. Le chemin au sol des satellites est illustré figure 3.4.

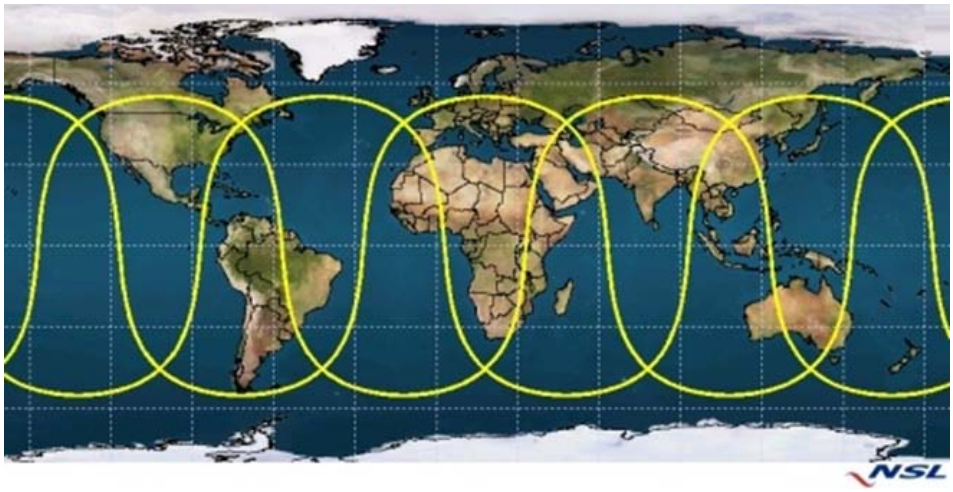


Figure 3.4 – Tracé au sol d'un satellite (1 jour) – Courtesy of NSL.

En juillet 1995, GPS a été déclaré complètement opérationnel, avec une

constellation de vingt-quatre satellites complétés par un segment au sol. Ce segment se compose de stations au sol. Elles surveillent l'état du segment espace. Un autre segment, appelle segment contrôle, se charge de faire fonctionner les deux autres. Des satellites (treize) de secours ont complété le système (bloc IIR, sept ans et demi de durée vie). Ce nouveau bloc intègre une fonction « autonav » qui améliore la précision et la validité des informations (jusqu'à plusieurs mois sans que des contacts au sol ne soient nécessaires).

Fréquence de transmissions Les blocs II (IIA et IIR) transmettent leur signal conformément aux spécifications du GPS originales. Sur les quatre fréquences autorisées, deux servent au positionnement : L1 et L2 (1572.42MHz, 1227.6 MHz). Les deux autres fréquences sont : L3 (1381.05 MHz) et L5 (1176.45 MHz) qui portent le signal pour les rôles alternatifs de détection de lancement de missiles, de détonations nucléaires et autres événements hauts en rayonnement infrarouge. GPS, dans sa conception originale, offre deux services, le « Standard Positioning System » (SPS) ouvert à tous les utilisateurs sur L1 et basé sur C/A avec les messages de navigation et le « Precise Positioning system » (PPS) un service évolué réservé aux personnes autorisées, sur les fréquences L1 et L2, basé sur P/Y et les messages de navigation.

Il existe trois codes de modulation sur les fréquences de positionnement. Le code *Coarse Acquisition* (C/A) modulé sur L1 et L2 est accessible à tous les utilisateurs. Le code *Precise code* (P/Y)est modulé sur L1 et L2. Il est chiffré et accessible uniquement aux utilisateurs autorisés (militaires). Ces signaux comportent aussi les messages de navigation. Le codage des informations pour le signal GPS s'effectue en plusieurs étapes. Premièrement, le message de navigation et de temps sont combinés avec le *Pseudo Random Noise* (PRN) du satellite. Cela permet de transmettre un identifiant unique au satellite qui permet au récepteur de savoir exactement de quel satellite il reçoit des informations. Ensuite, ce signal est mixé sur la fréquence L1.

Le récepteur détermine sa distance à un satellite en utilisant la formule suivante :

$$d_{sat} = \Delta t * s_{light}$$

Où Δt est le différentiel de temps entre l'émission et la réception du signal et s_{light} la vitesse de la lumière.

Un des problèmes de cette technique est que les horloges des récepteurs ne sont en aucun cas comparables aux horloges atomiques des satellites. Cependant il existe une méthode permettant de calculer sa propre erreur de temps. En effet, lorsque le récepteur calcule sa position, les différentes erreurs impliquent que le récepteur ne peut pas calculer un point unique d'intersection entre les sphères des satellites. En interpolant une position depuis l'aire commune qui est calculée, le récepteur peut en déduire son propre décalage temporel.

Actuellement GPS est critiqué par les utilisateurs avancés pour avoir une période de couverture trop faible avec peu de satellites visibles. Pour les utilisateurs standard, cela ne pose aucun problème car quatre satellites seulement sont nécessaires pour leurs applications. Mais pour la surveillance ou pour des applications Real Time Kinematic (RTK) comme les études hydrographiques, il faut six satellites qui ne soient ni bloqués ni inactifs.

Modernisation Il y a eu plusieurs améliorations notables du service utilisateur. En 1998, un second signal utilisateur, plus précis, a été ajouté. En 2000, la « disponibilité sélective [1] » qui dégradait volontairement le signal utilisateur a été enlevée (Plus de détails dans [Shaw 2000]). Cette dégradation du signal civil était une mesure pour empêcher un ennemi d'utiliser GPS pour le guidage de missile avec précision. Cette modernisation a pour objectif de donner une meilleure précision des signaux, à la fois militaire et civil. Il concerne le segment spatial ainsi que le segment contrôle.

Une deuxième phase de modernisation a débuté en 2005 avec le lancement des premiers satellites block IIR-M. Ces satellites transmettent sur la fréquence L2C, L1M et L2M (C- civil ; M- Military). L2C a pour but de donner une précision accrue et une meilleure traçabilité du signal. L1M et L2M sont les deux signaux contenant le code amélioré pour les usages militaires (M-code) et diffusent sur les fréquences L1 et L2. Ce signal militaire possède une résistance accrue contre les tentatives de brouillage, ainsi qu'une meilleure détection du multi-path (phénomène décrit dans une section ultérieure).

De nouveaux satellites ont été lancés courant 2010 (de Mai à août) puis en 2011. Ce bloc IIF déploie la fréquence L5, qui est utile pour la navigation aérienne.

Le gouvernement américain a prévu de continuer la modernisation de GPS avec GPS III. Ce projet répondra aux besoins civils et militaires des trente prochaines années. Les technologies qui seront intégrées permettront une meilleure capacité contre le *brouillage*, une meilleure *sécurité*, une plus grande compatibilité et inter-opérabilité avec Galileo et la « grille d'informations globale [2] ». Ce système inclut trente satellites, avec des satellites de secours [Hegarty. 2008].

Nous proposons ici une figure montrant les différents blocs de satellites, et les spectres de fréquences utilisés.

Le planning pour GPS III est prévu pour un premier lancement en 2014. Le programme mettra trois ans pour arriver à un premier état fonctionnel (dix-huit satellites) et deux ans supplémentaires pour arriver à une capacité maximale (30 satellites). Selon la plannification prévue, GPS III sera pleinement opérationnel en 2020.

1. Selective availability
2. Global information grid

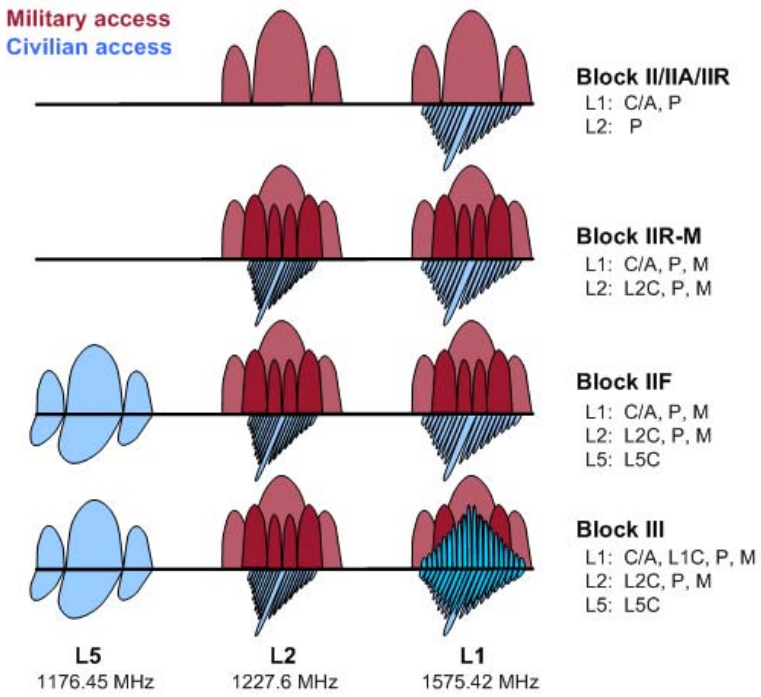


Figure 3.5 – Signal GPS et fréquences des codes (Courtesy of NSL)

GLONASS ROSKOSMOS est le contrôleur global du système GLONASS et de son développement. Cependant il y a aussi le comité inter-agences de coordination [3] qui organise le programme GLONASS selon les directives de différents ministères. GLONASS est un système dual, intégré dans une constellation Medium Earth Orbit (MEO) à une altitude nominale de 19150 kilomètres, à 64.8 degré d'inclinaison et sur trois plans orbitaux. Le parcours au sol est montré sur la figure 3.6. GLONASS diffuse sur une base d'accès multiples en fréquence divisée (Frequency Division Multiple Access (FDMA)) sur laquelle chaque satellite est identifié par une fréquence plutôt qu'un code unique comme c'est le cas pour GPS (Code Division Multiple Access (CDMA))

Les derniers lancements de trois satellites GLONASS-M (Modernisé) furent achevés le 1^{er} mars 2010. Depuis, 19 satellites sont actifs ainsi que trois satellites en orbite en "commission phase". Deux autres satellites sont en orbite mais en période de maintenance. Les dix-neuf satellites offrent une couverture totale de la Russie. Il faudrait vingt-quatre satellites pour obtenir une capacité opérationnelle complète, avec une couverture mondiale. Deux autres lancements étaient prévus pour 2010, pour délivrer les derniers satellites requis. Les premiers satellites GLONASS-K faisaient partie du premier lancement. Hélas, à cause d'un soucis technique avec la fusée de lancement, ces trois satellites

3. Interagency Coordination Board

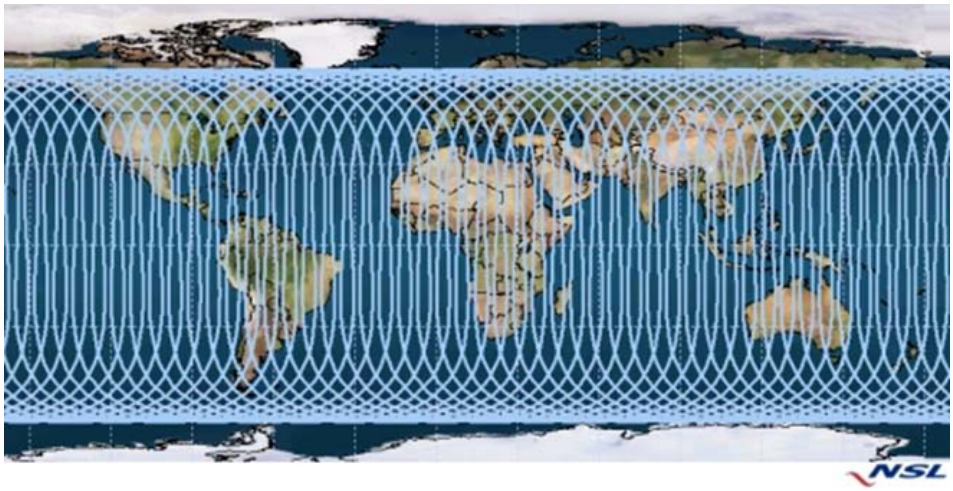


Figure 3.6 – Parcours au sol d'un satellite GLONASS (8 jours) – Courtesy of NSL

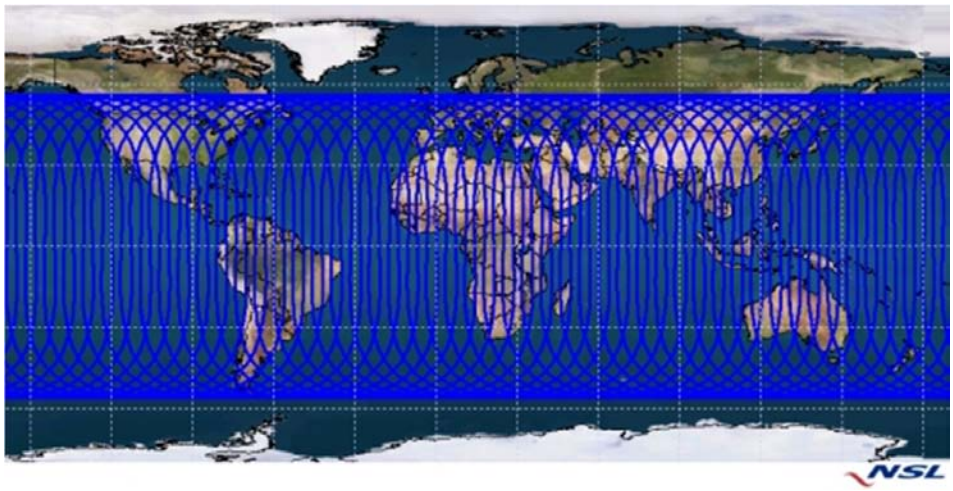
furent perdus. Un lancement a été correctement effectué en février 2011 et le satellite mis en orbite est en train d'être testé. Un autre lancement, prévu en août 2011 a été annulé du fait de problèmes liés aux précédents lancements. Aucune date n'est actuellement connue pour la reprise des lancements. Lorsqu'il sera complètement opérationnel, GLONASS devrait offrir un niveau de service comparable à GPS.

En juillet 2009, Dmitry Marareskul, chef du "onboard satellite navigation sector of Information satellite System Reshetnev Corporation in Zheleznogorsk" révéla les plans de la Russie en matière d'expansion de son réseau de suivi et de mesures de GLONASS. Le plan était d'inclure l'Australie, Cuba et l'Amérique du sud. La Russie a aussi signé des accords pour promouvoir GLONASS au Brésil, à Cuba et au Venezuela. L'agence souhaite établir une surveillance terrestre hors du territoire Russe pour améliorer la précision, l'intégrité et la fiabilité de GLONASS. D'autres stations de référence dans d'autres pays seront nécessaires pour renforcer le segment de contrôle de GLONASS et atteindre un niveau comparable à GPS.

Galileo Programme européen pour développer un système civil global de navigation par satellites, Galileo doit fournir un positionnement très précis et fiable. Il fournira aussi un positionnement instantané de l'ordre du mètre. Cela sera possible grâce à une meilleure orbite, de meilleures horloges et un système à double fréquences.

Galileo donnera aussi une assurance concernant la disponibilité du service et informera ses utilisateurs dans les six secondes en cas de défaillance d'un satellite. Cela doit permettre son utilisation dans des applications critiques impliquant des vies humaines, ou pour le commerce, . . .

Le système Galileo complètement déployé comprendra trente satellites,



dont vingt-sept opérationnels et trois de secours. En plus des satellites, une infrastructure au sol complète le dispositif (voir section 3.1.2.1 sur EGNOS).

Au total, dix signaux seront émis. Quatre sont destinés à des usages gouvernementaux et payants. Ces quatre signaux seront cryptés. Le système offrira les services suivants :

- Safety of Life (Systèmes vitaux) ;
- Open service (Service ouvert) ;
- Public Regulated Service (Service public régulé) ;
- Commercial Service (Service commercial) ;
- Search and Rescue (Recherche et secours).

Galileo inclut aussi des concepts additionnels, permettant entre autre, de prendre en compte les services de localisation régionaux et locaux. On les appelle les "éléments locaux de Galileo" [4].

3.1.2 Les systèmes d'amélioration de positionnement

De par la nature du signal satellite, le positionnement n'est jamais parfait. En effet, l'environnement, les conditions de réception influent énormément sur la qualité du signal. De plus, il n'est pas possible de se localiser dans les bâtiments. Et il faut pouvoir supporter le délai engendré par le parcours du signal depuis l'espace. Pour palier à ces problèmes, plusieurs systèmes d'amélioration du positionnement ont vu le jour.

3.1.2.1 Satellite-based augmentation system

Les systèmes d'amélioration de positionnement, Satellite-Based Augmentation System (SBAS), trouvent leur origine dans la volonté d'améliorer la navigation aérienne. En effet, GPS n'est pas suffisamment sûr et précis pour assurer

4. Galileo Local Elements

les opérations très techniques de vols comme les approches ou les vols à basse altitude. En effet, un pilote doit pouvoir être averti dans la seconde d'une défaillance du système de positionnement, ce que ne permet pas le signal satellite à l'heure actuelle.

WAAS Le *Wide Area Augmentation System* est un système de navigation aérienne. Il est géré par la FAA *(Federal Aviation Administration)*. Ce système est conçu pour améliorer la précision et la disponibilité de GPS pour permettre aux avions de l'utiliser lors de toutes les phases d'approches.

Sans WAAS, les perturbations ionosphériques, les décalages d'horloge et les erreurs dans les orbites de satellites créent trop d'erreurs et d'incertitudes dans le signal GPS pour garantir une approche précise. Ce système se compose majoritairement de stations terrestres (*Ground segment*), et de cinq satellites (*Space segment*).

L'avantage de WAAS par rapport aux techniques utilisées actuellement est que ce système ne nécessite que l'équipement des appareils et le déploiement des infrastructures. Il n'est pas nécessaire d'équiper chaque aéroport. De plus, il apporte un avantage considérable pour les vols domestiques (d'un aéroport à un autre dans le pays), car les avions ne sont plus obligés de suivre des routes balisées à l'avance. Ils peuvent utiliser un système accessible quelle que soit leur position. Enfin, la précision supplémentaire leur permet plus de liberté, notamment la possibilité de voler avec sûreté à basse altitude.

EGNOS EGNOS fournit le même type de service que WAAS dans la zone d'espace aérien du ECAC (European Civial Aviation Conference). C'est un projet conjoint entre l'Agence Spatiale Européenne (ESA), la commission européenne et l'EOSAN (European Organisation for Safety of Air Navigation). Ces trois organisations sont responsables du développement et des tests initiaux du système.

EGNOS comprend trois satellites géostationnaires et offre deux services : le *Open Service* et le *Safety of Life service*. L'Open Service se compose de signaux de positionnement et de temps, accessibles librement depuis les satellites EGNOS. Ce service est accessible à tout utilisateur équipé d'un receveur compatible SBAS dans la zone de service EGNOS. Aucune garantie n'est offerte pour ce service. Après la qualification du système EGNOS le signal sera utilisé pour les services de Safety of Life (SoL) dans la majorité des applications de transport. EGNOS utilise une valeur appelée UERE pour corriger :

- Les horloges des satellites
- Les distorsions des signaux
- les erreurs de positionnement satellite
- les effets ionosphériques
- les effets troposphériques
- le bruit thermique, les interférences et "user receiver design"

Comme pour la zone ECAC, EGNOS fournit ses services pour la MEDA (Mediterranean Area) qui sera complétée par le déploiement de stations de référence supplémentaires en Égypte, Mauritanie, Maroc et Algérie. Des investigations ont été menées pour l'Afrique, l'Amérique latine et la région ACAC (Arab Civil Aviation Commission).

L'infrastructure *EGNOS Data Service* (EDAS) permet d'avoir accès aux données lorsque l'on n'a pas de vue nette avec les satellites EGNOS. Ces données sont collectées par les stations RIM (Voir 3.7). Ainsi EDAS fournit l'opportunité à chacun d'accéder aux données EGNOS en cas de rupture de connexion avec les satellites Galileo (En ville, tunnels, ...). Cela permet aussi de surveiller les performances d'EGNOS sur une zone sans avoir besoin d'un récepteur.

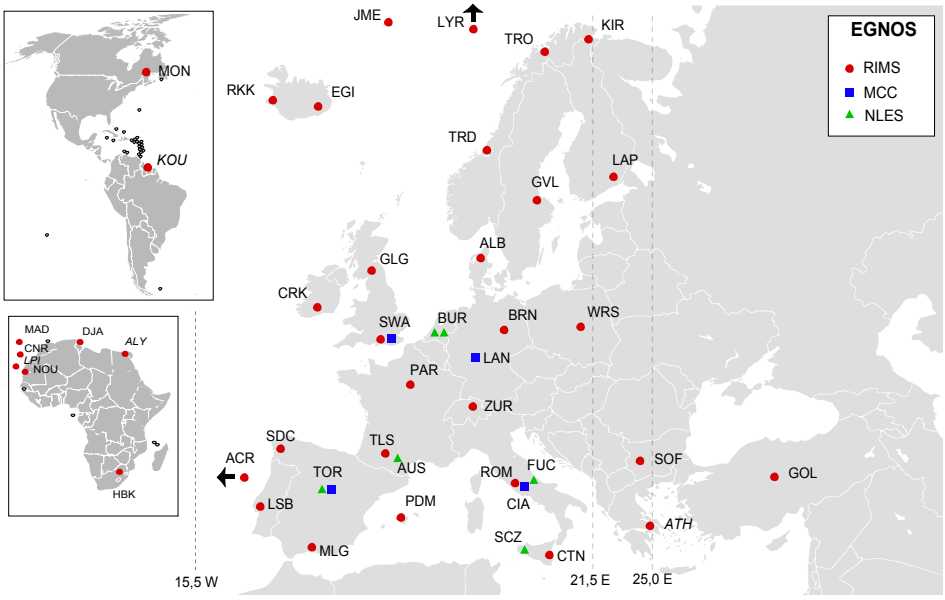


Figure 3.7 – Listes de stations RIM de EGNOS

Les informations d'EDAS se composent :

- des observations satellites GPS
- des observations satellites GLONASS
- des observations satellites GEO
- des éphémérides GPS
- des éphémérides GLONASS
- des éphémérides GEO
- du feedback cyclique NLES
- des données ATC
- des données RIMS APC

Ces données sont envoyées depuis toutes les stations RIM à travers le EDAS *control center*. La phase de test du système EDAS a débuté en mars 2009 et s'est terminée en mars 2010. Les messages EGNOS envoyés sur le système sont les mêmes que ceux reçus au niveau des RIM. Actuellement, EDAS souffre

encore d'une piètre disponibilité due au manque de bande passante au niveau du centre de contrôle.

SDCM Le système SBAS proposé par GLONASS, SDCM, sera le pendant russe du GPS WAAS d'EGNOS pour Galileo. En utilisant le réseau terrestre et les satellites géostationnaires, SDCM pourra donner en temps réel des corrections différentielles avec des précisions de l'ordre du mètre, et de 1 centimètre dans un rayon de 200 kilomètres autour des stations de référence. Deux satellites ont été lancés en 2010 et 2011 dans ce but.

3.1.2.2 Ground Based Augmentation Systems

Les Ground Based Augmentation Systems (GBAS) sont complémentaires des GNSS autour des aéroports (environ 30kms de rayon). Ils diffusent des corrections sur de très hautes fréquences. GBAS fournit initialement un support pour les approches de précision des appareils de CAT I et, au final, atteint les fortes exigences de précision, disponibilité et intégrité nécessaires aux approches de précision des CAT II et III. De plus, il fournit une précision accrue pour des approches plus flexibles, en courbe, dans la zone du terminal et autres zones où cela serait indiqué. La précision actuelle démontrée par GBAS est inférieure à 1 mètre.

3.1.3 Vulnérabilités des systèmes de positionnement par satellites

Comme nous l'avons vu, les signaux reçus par le receveur peuvent être de faible puissance. Ils sont donc sensibles à différentes sortes de perturbations, intentionnelles ou non. Ces perturbations sont dites *non-intentionnelles* lorsque des phénomènes extérieurs viennent perturber les transmissions (ionosphère, troposphère, perturbations RF), et *intentionnelles* lorsque l'utilisateur est délibérément privé du signal ou que les receveurs sont trompés pour indiquer de fausses positions.

3.1.3.1 Perturbations non intentionnelles

Nous allons développer dans cette partie les perturbations naturelles (ionosphériques majoritairement) et celles dues à d'autres facteurs.

Perturbations ionosphériques La ionosphère est une couche qui entoure la Terre entre 85 km et 500 km d'altitude. À cette distance, la densité de l'air est si faible que des électrons libres peuvent exister un bref instant du fait des radiations solaires. La plus grande densité d'électrons libres est relevée dans la couche dite F située entre 120 et 400 km d'altitude. De ce fait, c'est cette couche qui cause les plus grosses perturbations sur les GNSS.

L'activité ionosphérique est diurne par nature, elle est à son plus haut degré d'activité lorsque les rayons solaires frappent l'atmosphère, causant le phénomène dit d'"ionisation". Lorsque les rayons arrêtent de traverser l'atmosphère, l'ionosphère redevient inerte.

Pendant les périodes de forte activité solaire (voir les effets similaires engendrés, sur la figure 3.8), l'intensité des radiations augmente ainsi que le degré d'ionisation. Cela cause des perturbations encore plus prononcées. Mais en général cela n'affecte qu'un satellite ou deux tout au plus. Il a aussi été enregistré un cas de scintillation ionosphérique déclenchant le mode "détonation nucléaire" d'un satellite, causant ainsi une manœuvre orbitale non annoncée.

Facteurs environnementaux Ces erreurs sont celles qui sont le plus susceptibles d'être rencontrées par les récepteurs GNSS. Les deux facteurs les plus évidents sont le *multipath* et le masquage de signal.

Le Masquage du signal peut être causé par plusieurs facteurs : traversée d'un pont, d'un tunnel, présence d'arbres ou de bâtiments.

Le *multipath* se décrit bien par la réflexion ou la diffraction du signal. Comme le chemin parcouru par le signal diffracté ou réfléchi est toujours plus long que le chemin direct, ces multi-signaux arrivent toujours après le vrai signal. Or, la position calculée par le receveur utilise en particulier le temps de propagation des signaux émis par les satellites. Ces rebonds qui rallongent le temps de réception du signal contribuent à dégrader la qualité du positionnement. Si les longs retards peuvent facilement être détectés et n'auront pas d'influence sur le positionnement, en revanche, les retards courts sont plus compliqués à gérer. Ce sont eux qui génèrent le plus d'erreurs de *multipath*.

L'atténuation du signal, qui n'est pas exactement un effet de *multipath*, a également des effets dans des scénarios de *multipath* selon l'environnement, et peut causer le verrouillage sur un signal réfléchi plutôt que sur le signal direct.

Interférences des radios fréquences Des perturbations naturelles des radios fréquences (RF) peuvent se produire à cause de l'activité solaire, en plus de l'ionisation déjà évoquée. Elles sont créées par les pics de radiation solaire qui causent une élévation du bruit lorsqu'elles entrent en contact avec le champ de plasma de la Terre. Ce bruit empêche les récepteurs GPS de suivre les satellites. En décembre 2006 une telle éruption solaire a été détectée. La figure 3.8 montre l'effet maximum de la radiation. La réception du signal GPS a été fortement perturbée dans la zone du pacifique sud (zone bleue).

D'autres interférences peuvent provenir de multiples sources liées à l'activité humaine telles que les appareils de communication mobiles. La ITU (Internaltional Telecommunications Union) et ECPT (European Conference of Postal and Telecommunications Administrations) minimisent les effets en allouant sélectivement les différentes bandes de fréquences. Cependant les harmoniques

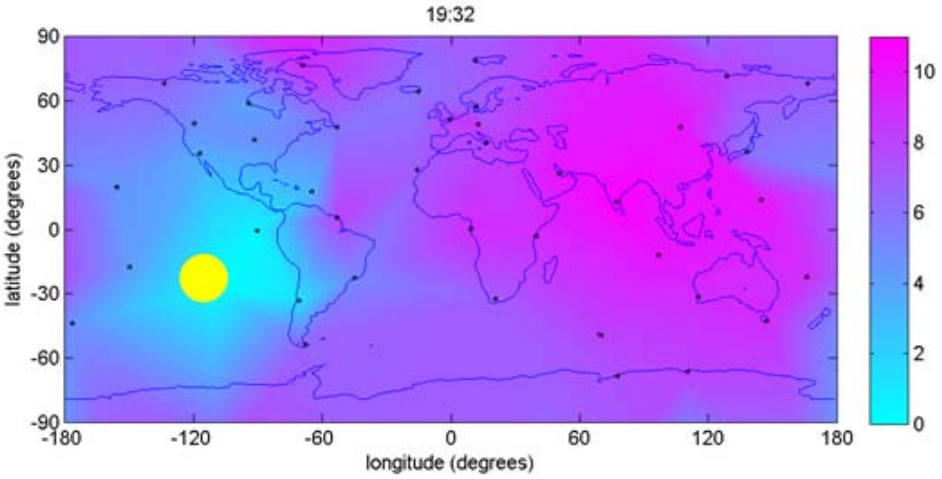


Figure 3.8 – Impact de l'éruption solaire du 6 décembre 2006

peuvent également créer des interférences. De la même manière, les antennes TV sont susceptibles de générer des interférences, notamment lorsque le poste inclut un amplificateur interne. Cependant le changement du signal analogique vers le signal numérique devrait modifier cet effet.

Facteurs humains Tous les GNSS sont susceptibles de connaître des problèmes dus à des erreurs humaines. Notamment des erreurs dans le *firmware* du récepteur, des erreurs dans le hardware des satellites, ou encore des erreurs dans le segment de contrôle menant à des transmissions incorrectes de données. Elles sont difficilement détectables et encore plus difficiles à corriger.

3.1.3.2 Perturbations intentionnelles

Des perturbations intentionnelles peuvent se produire pour deux raisons :
- Un déni de service militaire (guerres ou autres raisons d'états)
- Un déni de service civil.

Déni de service militaire : La plupart des systèmes de navigation par satellites consistent en deux composantes, une militaire et une civile. L'accès à la composante militaire est strictement contrôlé et restreint alors que la composante civile est librement accessible. Les signaux de ces composantes sont de natures différentes. Ainsi le signal militaire est plus résistant aux interférences que le signal civil.

Ce type de déni de service est à traiter de manière marginale. En effet, il est peu probable qu'une guerre conventionnelle, utilisant les systèmes de navigation par satellites comme guidage de munition, se déclare en Europe. Dans le cas où un tel événement se produirait, les technologies de brouillage

sont suffisamment avancées pour que les zones non impliquées directement restent non affectées. De plus, l'implémentation de solutions pour contrer ce type d'attaque n'est pas à notre portée, si les autorités décident de brouiller ou couper les signaux civils, il ne sera pas possible de contourner ce déni de service sans l'aide de moyens militaires.

Déni de service civil : L'introduction de l'utilisation des technologies de GNSS pour suivre, calculer des frais, effectuer des surveillances personnelles, amène les gens à s'équiper de technologies pour brouiller les signaux. Les raisons incluent la malveillance, l'appât du gain (en se jouant du système) ou encore la protection de la vie privée. Une recherche rapide montre que de petits brouilleurs peuvent être achetés pour un faible coût.

Figure 3.9 – Brouilleur GPS

Le brouillage consiste en la génération d'une vague continue d'interférences qui augmentent significativement le bruit sur la bande de fréquences des satellites. Cela empêche les récepteurs d'acquérir le signal et de suivre les satellites. Plusieurs techniques de brouillage peuvent être employées. Brouillage à longue portée sur une seule fréquence, courte portée sur toutes les fréquences. Le brouillage peut s'effectuer sur une faible bande de fréquences ou à plus large échelle. Des sauts de fréquences sont aléatoirement générés dans la bande de fréquences à altérer dans le but de rendre la détection et l'inhibition du brouillage plus difficile. Tout dépend du prix de l'investissement. Le brouillage en lui-même est une menace peu importante, et est en général bien détectée. Cependant il est souvent le prélude de falsifications plus coordonnées comme le *spoofing* et le *meaconing* qui, eux, ont pour objectif de ne pas être détectés (voir paragraphe suivant). Il existe de multiples références citant

les effets du brouillage sur le GPS. [Papadimitratos 2009] indique qu'un simple brouillage aérien de 1W peut empêcher l'acquisition des signaux GPS. Il est aussi possible de faire perdre le verrouillage des appareils. Un brouilleur sophistiqué pourrait étendre ces effets jusqu'à 350km. Ces effets sont maximaux lorsque le terrain ne vient pas interférer.

Falsifications La falsification consiste à leurrer le récepteur avec un signal qu'il considèrera comme authentique mais qui le conduira à fournir à l'utilisateur une mauvaise position. Il y a deux types de falsifications.

Le *spoofing* consiste à émettre un signal GPS avec une puissance légèrement supérieure à celle du signal GPS légitime. Cela amène les récepteurs à privilégier ce faux signal lors de l'acquisition. Ce type de brouillage est plus efficace dans le domaine civil où les codes PRN sont accessibles et connus. L'objectif de cette attaque est de fournir une information de temps erronée à un récepteur GPS. Puisque celui-ci le considère comme vraisemblable, il va tenter de l'inclure dans le calcul de sa position. Ainsi la solution trouvée sera fausse. Il est ainsi possible de tromper l'utilisateur. Les systèmes d'attaque les plus sophistiqués sont en mesure de prévoir la solution que va calculer un récepteur *spoofé*. Mais cela relève de techniques avancées et nécessite du matériel haut de gamme.

Le *Meaconing* est une réception d'un vrai signal GPS rediffusé depuis une nouvelle position (avec une puissance légèrement supérieure). C'est une sorte de spoofing qui utilise un vrai signal. Le problème du meaconing est qu'une fois identifié, on peut déterminer la source du brouillage.

3.2 État de l'art des techniques d'atténuation

3.2.1 Techniques d'atténuation d'effet

Pour pallier les attaques que nous avons décrites précédemment, certaines méthodes existent déjà. Il y a trois niveaux auxquels on peut intervenir :

- Le niveau de la constellation : interventions lourdes et coûteuses car elles impactent du matériel de pointe (satellites) et pas forcément accessibles ;
- Le niveau intégrateur : modification de l'équipement de réception pour détecter des attaques,
- Le niveau utilisateur final : utilisation des APIs mises à dispositions.

Au niveau système, plusieurs étapes de modernisation de la constellation GPS ont eu lieu. Originellement, seuls deux signaux étaient émis par les satellites. Depuis 2005 de nouveaux signaux se sont ajoutés à la liste. Ces signaux permettent une précision accrue et une résistance supplémentaire face aux

brouillages. Un code spécifique aux militaires a aussi été ajouté. Ce dernier apporte des méthodes encore plus évoluées pour parer les attaques de signaux.

Au niveau intégration, le multi-path est la perturbation contre laquelle il est le plus aisé de se protéger. La plupart des constructeurs ont développé des algorithmes dans le but de minimiser ces effets. Les fruits de leurs recherches sont inclus dans les récepteurs.

Les algorithmes développés par les constructeurs de récepteurs pour se protéger contre le multi-path sont transparents à l'utilisateur, mais la majorité est basée sur une réduction des écarts grâce à des corrélations qui permettent d'aligner le signal reçu avec une référence générée en interne. D'autres techniques connues sous le nom *pulse aperture* et *strobe correlators* utilisent un nombre important de corrélateurs pour différencier les combinaisons de signaux et réduire encore plus les effets du multi-path. Ces types d'algorithmes sont en général présents dans les systèmes haut de gamme.

Les options pour contrer les problèmes de masquage sont limitées. Cependant, certains récepteurs haut de gamme peuvent utiliser des techniques avancées pour suivre les satellites dans des environnements difficiles, même si la précision restera toujours dégradée. Une seconde technique pour régler le problème du masquage serait la compatibilité du récepteur avec d'autres signaux GNSS, ou l'utilisation de signaux terrestres qui fournissent une position à court terme, le temps de retrouver un signal satellite.

Les techniques anti-brouillage de niveau équipement consistent en de multiples méthodes, dont certaines sont réservées à l'usage militaire. Mais toutes requièrent l'amélioration des équipements en place. Les méthodes tombent dans deux catégories selon qu'elles concernent les antennes ou les récepteurs.

Des technologies d'antennes peuvent être utilisées pour atténuer les interférences. Elles utilisent des matrices d'antennes et utilisent quatre techniques différentes, allant du filtrage à des techniques plus avancées qui calculent les positions en fonction de l'orientation du récepteur.

Les utilisations civiles de la majorité de ces techniques d'atténuation sont largement restreintes par les autorités militaires du globe. Seule l'aviation civile utilise ces technologies, les appareils restant très coûteux. Au niveau opérationnel, les méthodes pour réduire les effets des interférences sont limitées. Des notifications concernant les probabilités d'interférences dues à l'activité solaire pourraient être annoncées par les observatoires astronomiques similaire aux systèmes *Notices to Airmen* (NOTAM) et *Notice to Mariners* (NTM) utilisés actuellement pour l'aviation et la marine. On peut se référer à [Group 2004] pour plus de détails sur les impacts potentiels liés au manque d'intégrité de GPS dans la sécurité de l'aviation.

3.2.2 Solutions logicielles basées sur des techniques de détection d'anomalies

La section a décrit des techniques souvent liées aux technologies pour éviter que le signal ne soit perturbé, volontairement ou non. Mais aucune de ces techniques n'empêche complètement les perturbations d'avoir lieu. Et les attaques décrites dans la section 3.1.3 restent possibles, même si de fait, elles nécessitent un effort supplémentaire de la part de l'attaquant.

Une autre voie, complémentaire à l'approche technologique, consiste à détecter les anomalies qui peuvent être contenues dans le signal ou les données utilisées pour calculer la géolocalisation. Les méthodes que nous décrivons ici n'apportent aucune correction à la géolocalisation d'une part et d'autre part, ne sont pas exhaustives. Cependant, elles permettent malgré tout d'augmenter le degré de confiance que l'utilisateur peut avoir dans le système. Lorsqu'aucune alerte n'est signalée, la probabilité que tout aille bien est plus élevée.

Dans [Wen 2004] les auteurs présentent dix méthodes pour contrer ou détecter les attaques contre le *spoofing* GPS. Des améliorations de ces méthodes similaires étaient déjà évoquées dans [Warner 2003]. [Papadimitratos 2009] complète ces méthodes avec de nouveaux contrôles que nous allons décrire dans le paragraphe suivant.

3.2.2.1 Contrôles d'intervalles

La première méthode se base sur le fait que la puissance du signal GPS pour la fréquence L1 ne peut dépasser -153dBWatt et -158dBWatt pour L2. Même si la valeur peut varier, elle doit rester dans un intervalle raisonnable. Si le seuil est dépassé, nous sommes alors en présence d'un signal suspect.

Si le signal entre L1 et L2 montre des variations significatives, on peut en conclure que ce canal est attaqué. Avec une surveillance constante des deux puissances relatives, il est possible de détecter les attaques qui ne ciblent pas tous les canaux.

La troisième méthode se base sur la propriété que les signaux RF irradiants d'une source non parfaite satisfont une équation particulière entre la distance au satellite et la puissance du signal. $Pr^2 = constante$, où P est la puissance reçue et r est la distance entre le récepteur GPS et le satellite. Les satellites GPS sont à une altitude de 20000 km. Malgré le déplacement du récepteur, cela ne doit pas changer la puissance du signal de manière significative. Il s'agit donc de vérifier à tout moment que la valeur du produit Pr^2 est toujours égale à une constante que l'on aura bien choisie.

La quatrième méthode implique la mesure du rapport selon lequel l'intervalle de code et l'intervalle de phase varie. Ces mesures sont appelées *range rate* et *phase range*. L'intervalle de code et de phase doivent être consistants. Cependant il est plus facile de *spoofer* l'intervalle de code car il est impossible de contrôler l'intervalle de phase lors du déplacement d'un récepteur. Ainsi le

spoofer sacrifiera généralement ce dernier au profit de l'intervalle de code. Détecter cette inconsistance permet de signaler une attaque.

À cause du mouvement de satellite, le signal GNSS est soumis à l'effet Doppler. Comme pour le son, la fréquence du signal change. La cinquième méthode se base sur l'hypothèse qu'un *spoofer* unique ne peut pas simuler cet effet, car cela nécessiterait de moduler sur plusieurs fréquences simultanément. Il faudrait autant d'équipements de perturbation que de satellites en vue.

3.2.2.2 Intra-corrélations

Il s'agit d'utiliser les corrélations entre les deux fréquence du GPS, L1 et L2. Ces deux fréquences partagent un même code, le code P(Y). Si on calcule la corrélation entre les deux, nous ne devons pas trouver de décomposition en deux classes.

Les amplitudes des signaux sur les fréquences L1 et L2 sont liées. Il est primordial de vérifier la consistance de ce lien. Une partie des différences entre ces amplitudes sont dues à l'ionosphère. Un signal contrefait partant du sol n'est pas affecté par l'ionosphère. Un signal qui ne souffrirait pas de ces délais est suspect.

3.2.2.3 Analyse résiduelle

Quand un récepteur acquiert un signal (S_{ant}), un filtre de fréquence peut être utilisé pour soustraire de ce signal la porteuse GPS (S_{GPS} et S_{bruit}). Si on arrive à régénérer un autre signal à partir de ce résidu (S_{spoof}), cela indique que plusieurs signaux étaient superposés et donc qu'il existe un signal de *spoofing*). L'équation suivante illustre ce principe :

$$S_{ant} - S_{GPS} - S_{bruit} = S_{spoof}$$

Cette régénération peut échouer si le bruit est trop important.

3.2.2.4 Inter-corrélations

Quand un signal est reçu, il est possible de le valider. Les positions des satellites peuvent être déterminées à partir du signal. Pour vérifier ces positions, il est possible d'utiliser les éphémérides. Il est difficile de *spoofer* les éphémérides sur une longue période. Les éphémérides (ou almanachs) sont des informations génériques sur la constellation de satellites. Lorsqu'un récepteur est allumé, il met à jour son almanach. Grâce à lui, le récepteur connait sa position initiale (estimation) ainsi que la position du premier satellite qui a envoyé le signal. Avec ces informations, un appareil peut prédire la position des satellites même en tenant compte des mouvements. Une bonne pratique consiste à vérifier la cohérence de ces prédictions avec la réalité.

3.2.2.5 Détection de saut

Toutes les données mesurables peuvent être utilisées pour détecter un *spoofing*. La plupart du temps, les valeurs fournies par les capteurs changent progressivement. En s'inspirant de la dixième méthode de [Wen 2004], tout changement brusque d'une de ces valeurs est un saut. C'est un indicateur de tentative de *spoofing*.

3.2.2.6 Analyse du signal

De lui-même, le capteur de signal est capable de détecter des problèmes lors de la géolocalisation. Peu de ces mesures sont actuellement disponibles, car les constructeurs ne les implémentent pas. Leur stratégie consiste à donner à l'utilisateur un résultat quoiqu'il arrive. Quand bien même ces mesures seraient effectuées, leurs API ne fournissent pas les données aux programmeurs tiers.

3.2.2.7 Corrélations

Les équipements personnels sont maintenant bardés de capteurs de différents types. Certes, leur fonction primaire n'a pas nécessairement vocation à localiser l'utilisateur. Cependant, certains équipement le font dans le cadre du fonctionnement de leur service associé (Antenne GSM, carte Wi-Fi). D'autres ne permettent pas de déterminer une position absolue de l'utilisateur, mais sont capables de calculer une variation de celle-ci (pré-senti par [Warner 2003] concernant la boussole et l'accéléromètre et étudié dans [Papadimitratos 2009] pour l'accéléromètre). Ces différentes informations peuvent nous permettre de montrer des incohérences dans les mesures GPS, pouvant indiquer un mauvais signal ou une attaque. [Zirari 2010] proposent une méthode dite hybride, permettant de palier les défauts de satellites grâce à une infrastructure 802.11. Il est nécessaire de disposer de quatre satellites pour calculer une solution de géolocalisation. Cependant, si leur nombre n'est pas suffisant, il est possible de calculer une solution grâce aux points d'accès composant le réseau 802.11, pour compléter le nombre de points de référence jusque quatre. Dans le cas où le nombre de satellites ou de points d'accès sont déjà en nombre suffisant, cette méthode permet d'améliorer la précision par rapport à la mesure originale, en choisissant la configuration points d'accès-satellites la plus avantageuse.

Surveiller la puissance relative Si le rapport des puissances des deux canaux L1 et L2 change de façon significative et qu'une seule des deux puissances augmente brutalement, cela signifie que le canal dont la puissance a subitement changé subit une attaque de *spoofing*. Cette méthode permet donc de détecter les attaques de *spoofing* qui ne couvrent pas tous les canaux.

Corrélation croisée entre L1 et L2 Les porteuses L1 et L2 possèdent le code P(Y) en commun. Sur ces deux codes le message GPS doit être identique.

Il faut donc vérifier que le message du P(Y) est le même sur les 2 porteuses. Même si le Y-code est activé, il est possible de faire ce test sur un récepteur qui ne ferait que le test de corrélation entre le P code de L1 et le P code de L2. Cette technique permet de détecter des tentatives d'usurpation sur une des deux porteuses. Donc avec cette méthode, pour qu'une attaque d'usurpation réussisse il faut que le signal d'usurpation couvre les 2 porteuses.

Différence de range entre L1 et L2 Les amplitudes des signaux sur L1 et sur L2 sont liées entre elles et il est nécessaire d'en vérifier la cohérence. Les différences d'amplitude entre L1 et L2 sont dues aux effets de l'ionosphère. Un signal d'usurpation ne sera pas altéré par l'ionosphère puisqu'il est émis au sol. Les différences entre l'amplitude de code et de phase entre L1 et L2 ne seront donc pas visibles et ne seront pas affectées par ce qu'on appelle le retard ionosphérique. Il suffit donc qu'un récepteur vérifie que le signal reçu a bien subi ce retard. Cette contre-mesure permet de vérifier que la source d'un signal reçu est bien un satellite.

Vérifier les données de position des satellites Lorsqu'un signal GPS est reçu par un récepteur, il est possible de vérifier la validité de ce signal. En principe les positions des satellites peuvent être déterminées par un signal. Pour vérifier la position des satellites, il est possible d'utiliser les éphémérides. Il est assez difficile pour un *spoofer* d'usurper ces données pendant un temps très long. L'almanach consiste en une information générale sur la constellation des satellites. Quand un récepteur est allumé, il met son almanach à jour. Grâce à cet almanach, il connaît sa position initiale et la position initiale du satellite qui a envoyé le premier signal. Grâce à cette information, le récepteur est capable de déterminer à peu près la position du satellite même si celui-ci se déplace en orbite. Il suffit de dire : les positions des satellites doivent être cohérentes avec les éphémérides et les almanachs qui sont des bases de données publiques recensant les positions des différents satellites.

3.2.2.8 Analyse des résultats

Parfois les signaux peuvent sembler cohérents. Cependant, les résultats dérivés de l'analyse de ces signaux peuvent révéler des problèmes.

Borner et comparer la vitesse radiale La vitesse radiale est la vitesse à laquelle la distance entre l'appareil de mesure et la source du signal mesuré change. Cette vitesse peut être mesurée et doit donner des résultats cohérents.

Le code range et le phase range doivent être *spoofés* de façon cohérente. Le code range est plus facile à *spoofer*. Pour un récepteur statique le problème ne se pose pas et il est facile de *spoofer* la phase range. Mais dans le cas d'un récepteur en mouvement le *spoofer* ne peut pas avoir de contrôle sur la

phase range. Donc la phase range sera en permanence sacrifiée par un *spoofer* au profit du code range et la cohérence sera perdue. Donc en comparant la vitesse radiale du code et de la phase il est possible de détecter une attaque de *spoofing*. De plus, la vitesse radiale entre un signal émis par un satellite et un *spoofer* au sol est totalement différente.

Vérification de l'effet Doppler Il n'est pas possible de reproduire toute la constellation avec un *spoofer*. En effet, même s'il est possible de moduler les codes PRN de différents satellites sur une même porteuse cela ne permet pas d'éviter le test Doppler. Il faut donc contrôler à tout moment que les équations de l'effet Doppler sont vérifiées.

Détection de saut Toute mesure observable peut être utilisée pour déterminer une usurpation. Généralement la valeur de ces informations change de façon cohérente. Un brusque changement de la valeur de ces informations s'appelle un saut et permet de suspecter une usurpation.

Les systèmes de géolocalisation sont soumis à des perturbations dont certaines sont des menaces réelles. Nous allons maintenant montrer comment tirer parti des informations autour de la géolocalisation dans le but de fournir des éléments visant à augmenter la confiance accordée à cette géolocalisation. Pour cela, nous nous appuierons sur de nouvelles sources d'informations, car le signal à lui seul n'est pas prévu pour être authentifiable. Ainsi, grâce à l'environnement de l'utilisateur et aux capacités des téléphones mobiles, nous serons en mesure de détecter, voire corriger, certaines perturbations.

Les faiblesses des architectures de géolocalisation sont bien connues. On a distingué deux types d'approche pour circonscrire ces faiblesses. Mais, dans la plupart des cas, ces méthodes impactent la conception même des services ou des équipements de l'utilisateur. Dans notre étude, l'une des contraintes est de fournir des solutions qui augmentent la confiance de l'utilisateur sans l'obliger à changer d'équipement mobile. Nous allons décrire, dans la section suivante, d'autres approches pour augmenter cette confiance.

3.3 Contribution à l'amélioration de la géolocalisation

Nous venons de décrire les infrastructures de géolocalisation avec leurs faiblesses. Assurer une qualité optimale ainsi que l'authenticité de la géolocalisation, qui sont les qualités exigées pour la certification d'une image, nécessiterait la modification de ces infrastructures. Or, les contraintes industrielles sont très restrictives. Il ne nous est pas possible naturellement d'agir au niveau des infrastructures GPS, ni même au niveau de l'équipement de l'utilisateur. Notre

axe de travail s'est porté sur le niveau le plus accessible, c'est-à-dire sur la partie logicielle de son smartphone. A défaut de pouvoir agir sur les propriétés de la géolocalisation (précision, authenticité, disponibilité), l'objectif minimal est d'informer l'utilisateur sur l'état de ces propriétés. La sécurité a été définie dans le chapitre 1 par l'« État d'esprit confiant et tranquille qui résulte du sentiment, bien ou mal fondé, que l'on est à l'abri de tout danger. » Considérons un automobiliste qui tombe malheureusement en panne sèche en rase campagne. S'il veut se rendre à la station d'essence plus proche, les moyens de s'y rendre sont très limités. Si la densité du trafic est nulle, il ne peut guère compter que sur la marche à pied. En revanche, s'il tombe en panne en agglomération, non seulement la station sera vraisemblablement plus proche, mais il aura à sa disposition différents moyens de transports aussi efficaces que rapides (taxis, bus, métro, etc.). On peut tout à fait transposer cette situation dans le domaine des télécommunications. Un utilisateur perdu dans une zone mal couverte par les infrastructures de télécommunication, ne pourra compter que sur l'information transmise par les satellites qu'il capte. Dans une zone urbaine, il sera plongé dans un environnement plus chargé en informations (GPS, GSM, Wi-fi, etc.), lesquelles pourront se pallier ou compléter les lacunes les unes des autres. A partir de ce constat, nous proposons différents scénarios qui, selon l'environnement de l'utilisateur, pourront au minimum l'informer de la qualité de sa géolocalisation (s'il n'utilise que les capacités standard de son smartphone). Dans le meilleur des cas, cette qualité pourra même être améliorée (s'il s'équipe d'antennes spécifiques, ou accède à des services dédiés).

Les contrôles que nous envisageons se décomposent en quatre catégories :

- **Les contrôles *a posteriori*** : la géolocalisation est effectuée normalement avec l'équipement standard de l'utilisateur (son smartphone) et l'on cherche à qualifier cette information. Cela peut être réalisé sur l'équipement lui-même, ou en transmettant la position à des infrastructures externes plus compétentes.
- **Les contrôles *a priori*** : avant de géolocaliser, l'environnement de l'utilisateur (a-t-il accès ou non à d'autres infrastructures comme le GSM ou le Wi-Fi, par exemple) et son comportement (son trajet, sa vitesse, etc.) sont utilisés pour repérer des anomalies et donc accorder plus de confiance à la géolocalisation qu'il effectue avec son équipement standard.
- **Les contrôles *correctifs*** : dans le cadre de certains GNSS, les infrastructures GBAS permettent de détecter et éventuellement de corriger les positionnements erronés, notamment ceux dus à des perturbations naturelles.
- *Les contrôles externes* : si l'on considère que le récepteur GPS d'un *smartphone* n'est pas d'une qualité suffisante, des équipements externes de réception peuvent compléter voire remplacer le récepteur standard.

Ces contrôles sont compatibles et peuvent être combinés pour obtenir un niveau de confiance maximal. Ils couvrent une palette de situations relativement large. L'utilisateur "isolé" et "négligeant" ne pourra s'appuyer que sur un

contrôle *a posteriori*, et sera contraint d'utiliser une géolocalisation de base, dont il n'aura la qualité qu'après le tatouage. A l'autre extrémité, l'utilisateur "prévoyant" qui s'est muni d'un récepteur spécialisé, et qui est plongé dans un environnement dense en terme d'infrastructures de communication, pourra connaître la qualité de la géolocalisation avant son utilisation. Il peut, de fait, combiner tous les contrôles pour augmenter le taux de détection des anomalies (perturbations ou attaques) et anticiper la force probante de l'image qu'il va tatouer. Naturellement, le coût de ces différents scénarios n'est pas le même.

Nous allons, dans les sections suivantes, décrire les solutions que nous proposons pour chacun de ces contrôles. Une partie de ce travail a été réalisé dans le cadre du projet européen FP7 ATLAS. La première section présentera plus en détail la structure, les objectifs et le déroulement du projet. Les sections 2 à 5 reprendront chaque type de contrôle à travers le scénario utilisateur et l'implémentation réalisée pour le projet.

3.3.1 Le projet ATLAS

Le projet ATLAS est né des constatations présentées ci-dessus. Il répond à trois besoins :

- Amélioration de la localisation
- Mesure de la qualité
- Rapport de preuve

ATLAS est un projet européen de petite taille dont le consortium réunit trois partenaires. La société NSL (leader du projet), la société CodaSystem et le LITA.

CodaSystem était une petite société d'une dizaine de personnes centrée sur son produit phare Shoot&Proof. Ce produit est un outil intégré de capture et gestion de photographie pour des agents de terrain. Le logiciel est déployé sur smartphone ainsi que sur des plateformes web. Ce système permet d'assurer une valeur légale à la photographie. CodaSystem a été racheté par STS-Group en 2011.

La société anglaise NSL, composée d'une vingtaine de collaborateurs, est spécialisée dans le monitoring de constellations satellites. *GNSS Integrity and Signal Monitoring Observatory* (GISMO) , l'un de ses produits majeurs, est un service d'analyse des systèmes de positionnement par satellites. Cet outil récupère les données brutes de mesure des sites de référence via Internet et calcule une position avancée à partir de ces données. L'analyse inclut :

- Observation des satellites de la Dilution of Precision (DOP)
- Observation de la précision du GPS par rapport aux stations de référence
- Observation des niveaux d'assurance et de protection
- Observation de l'environnement de réception du signal
- Observation de la qualité du signal
- Observation des erreurs
- Identification des fautes et échecs

GISMO opère en continu. Il a été testé et validé sur différents réseaux d'amélioration de positionnement dont EGNOS RIMS via EDAS.

Les résultats de GISMO se présentent sous deux formes :

- Un rapport journalier détaillant les analyses des observations
- Une base de données de statistiques qui est utilisée avec d'autres outils de dissemination. Le www.gnssassurance.org et www.edasmonitoring.org sont deux sites qui ont été développés pour analyser EDAS. Ces sites analysent les positions corrigées d'EGNOS/EDAS et comparées analytiquement avec les données GPS non corrigées.

Concrètement, les réalisations que nous avons effectuées au sein du projet ATLAS a permis d'intégrer autour et dans Shoot&Proof différents moyens d'augmenter la confiance de l'utilisateur. Ces intégrations peuvent être matérialisés sous forme de scénarios utilisateurs que nous allons présenter dans la section suivante.

3.3.2 Scénarios utilisateurs

Nous avons défini quatre scénarios. Ces scénarios s'appliquent dès lors qu'un utilisateur souhaite réaliser une photo certifiée. Quel que soit le scénario, l'objectif est toujours de fournir à l'utilisateur un maximum d'informations sur les conditions d'acquisition de la géolocalisation. Plusieurs types d'informations peuvent lui être fournis : des informations qualitatives, des corrections ou des alertes.

Dans la mesure où, pour augmenter le niveau de confiance de l'utilisateur, on ne peut pas prendre pour argent comptant l'information délivrée par le récepteur de son équipement, les scénarios proposés s'appuient tous sur des services externes au récepteur.

Le premier scénario consiste à confronter la géolocalisation du récepteur à GISMO. GISMO dispose d'un ensemble de paramètres environnementaux sur les conditions de réception, et permet donc de dire dans quelles conditions la géolocalisation qu'on lui soumet a été réalisée. Ce scénario s'inscrit dans les contrôles *a posteriori* et n'exige qu'une géolocalisation ponctuelle.

Le second scénario cherche à informer *a priori* l'utilisateur des conditions de réception. Pour ce faire, on considère qu'une géolocalisation ponctuelle est insuffisante. L'historique des positions de l'utilisateur est une source d'information importante, qui peut révéler des anomalies. Si l'historique des géolocalisations contient intrinsèquement les éléments révélateurs (anomalies sur la vitesse, l'accélération, etc.), cet historique peut être confronté à d'autres historiques qui révèleront d'autres anomalies (cohérence de la trajectoire GPS avec la trajectoire GSM par exemple).

Le troisième scénario se base sur un autre type d'informations externes. L'utilisation des GBAS permet de mesurer les déviations des positionnements dues aux perturbations non intentionnelles. Dès lors, il est aussi possible de

corriger la position des récepteurs mobiles dans les environs de la station de référence considérée.

Sans pouvoir maîtriser les conditions d'émission des signaux GNSS, il est en revanche possible d'améliorer la quantité d'informations disponibles à la réception. Dans le quatrième scénario, un récepteur dédié permet d'avoir accès à toutes les informations captées, sans les restrictions dues aux API mises en place par les constructeurs et intégrateurs de téléphones. De plus le récepteur dédié n'est pas nécessairement limité à une constellation particulière. Cela permet de corréler ces multiples constellations pour vérifier les positions.

Ces quatre scénarios sont tous développés en respectant l'intégrité de l'équipement de l'utilisateur final. Les composants logiciels nécessaires à ces scénarios ont la forme d'applications développées avec les outils standard des plateformes cibles (Android, iOS, etc.). Le quatrième scénario fait appel à une antenne externe, laquelle est totalement indépendante des équipements et s'appuie sur des modèles de communication standard (Bluetooth). Les sections suivantes décrivent en détail les modèles et traitements mis en œuvre pour ces quatre scénarios.

3.3.3 Modules de détection et correction de signaux

Les quatre scénarios ont donné lieu à la réalisation de quatre modules, qui couvrent l'ensemble des contrôles que nous avons présentés. Nous commencerons par décrire le *Sanity Check Module*, qui présente un ensemble de contrôles *a priori* embarqués sur le téléphone. Nous continuerons avec le module de qualité de service (QoS), contrôle *a posteriori* permettant de évaluer une position obtenue sans outils particuliers. Le module d'amélioration de position (EHL) permet quand à lui d'implémenter un contrôle *a priori* et correctif en utilisant les informations d'EGNOS. Finalement, le dispositif GAM propose un équipement dédié, utilisé pour réaliser des contrôles externes et *a priori* de la géolocalisation.

3.3.3.1 SCM

Le *Sanity Check Module* est l'implémentation du second scénario. Il est chargé de contrôler l'état du téléphone lorsqu'il est sous tension. Il est conçu pour fonctionner de manière autonome (excepté pour les bases de données trop volumineuses pour l'appareil).

Son objectif n'est pas d'améliorer la précision du signal ou sa réception. En revanche, il cherche à réaliser un ensemble de contrôles pour détecter une modification dans le signal GNSS : puissance qui change trop vite, trop grande, vitesse ou accélération humainement impossible...

Les techniques utilisées se divisent en trois types :

- Les méthodes basées sur les satellites
- Les méthodes basées sur les données

– Les méthodes basées sur le corrélations.

La plupart de ces techniques définissent un intervalle de valeur acceptable. Chaque fois qu'une mesure ne rentre pas dans son intervalle, l'utilisateur est informé qu'une attaque est potentiellement en cours.

Les approches basées sur les satellites sont utilisées pour détecter les cas de brouillage. Un récepteur peut surveiller le rapport signal bruit (SnR). Pour le signal GPS, on considère que le SnR doit se situer entre 19 et 47dB-hertz.

Les approches basées sur les données GNSS utilisent des configurations et des outils adaptés pour surveiller l'activité GNSS. Être en mesure de garder une trace des différentes positions calculées par le passé ainsi que leur date associée, permet de calculer la vitesse (deux points au moins) et l'accélération (trois points au moins). La vitesse est limitée par le moyen de locomotion utilisé. Quelqu'un voyageant à bord d'un véhicule standard ne sera pas en mesure d'aller plus vite que $160km.h^{-1}$ environ. Ainsi, en ayant connaissance du contexte de l'utilisateur, il nous est possible de fixer une limite de vitesse entre deux points de passage et des moyennes sur des sections de routes. Les mêmes limitations s'appliquent à l'accélération. Le véhicule n'est pas en mesure de délivrer de trop forte accélérations. En fixant le type de moyen de transport, il est donc possible de déterminer un intervalle d'accélération acceptable.

Les méthodes basées sur la corrélation tirent profit des différentes sources d'informations fournies par le mobile. Les antennes GSM ont une position et une zone de couverture unique et identifiable. Connaissant ces positions et ces zones de couverture, il est possible de calculer une position approximative du téléphone. Ce type de base de données est disponible sur Internet, comme par exemple le service en ligne Google Gears. Tout le temps qu'un mobile est connecté à une antenne GSM donnée, on peut supposer que celui-ci se trouve dans la zone de couverture de cette antenne. Une première méthode pour s'assurer de la fiabilité de la position acquise grâce au signal GNSS est de le comparer à la zone de couverture actuelle. Si l'on dispose de plusieurs zones de couverture en intersection (équipement en mouvement), il est alors possible de déterminer une distance maximale parcourues par le mobile durant un certain laps de temps (Voir la ligne bleue sur la figure 3.10). Cette distance maximale permet de calculer une borne supérieure pour la vitesse de l'équipement :

$$v_{max} = d_{max}/t$$

Si le capteur GNSS indique une vitesse supérieure, il y a suspicion d'attaque.

Sur la figure 3.11, un autre type de mesure est effectué. Celui-ci concerne l'orientation du déplacement. En effet, dans le cas où l'utilisateur traverse deux zones non superposées, l'angle de direction est borné par les tangentes à ces deux zones de couverture.

Sur la figure 3.12, on observe plusieurs courbes. La courbe bleue est le trajet réel suivi par l'opérateur. La courbe brisée vert clair est la géolocalisation

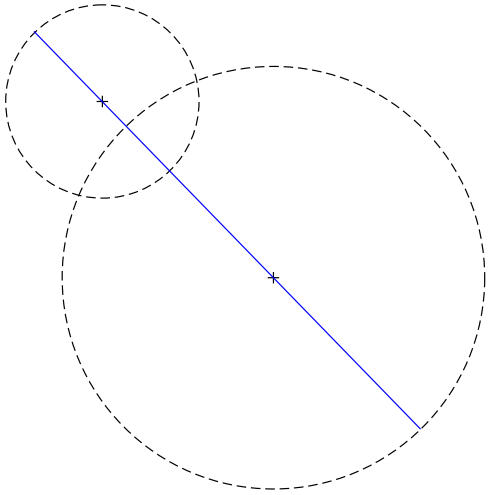


Figure 3.10 – Distance maximale parcourue dans deux zones de couvertures

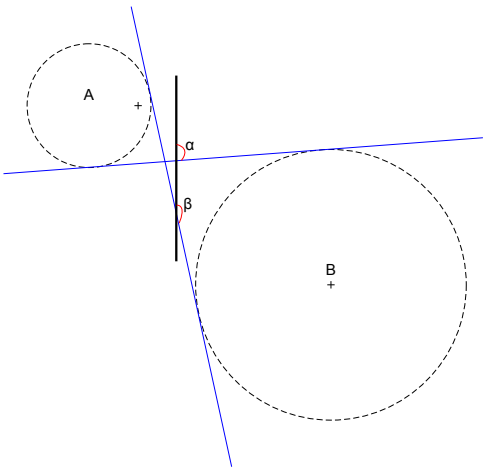


Figure 3.11 – Angles de directions possibles entre deux zones de couverture

GSM, qui se divise en plusieurs segments dont les extrémités sont une station radio. La courbe rouge est le trajet *spoofé*. Les points bleus représentent des alertes. Ceux-ci sont levés lorsqu'une anomalie est détectée. La courbe vert foncé est une courbe de Bézier. Cette courbe permet d'approximer le changement de position de l'utilisateur, en fonction du temps de connexion à une antenne donnée. En effet, l'utilisateur ne se trouve jamais au niveau d'une antenne mais dans son rayon d'émission. Le passage par la courbe de Bézier permet donc d'adoucir le chemin polynomial formé par les antennes. Cette courbe n'est pas fixe, elle se met à jour en fonction des changements de station GSM. Ainsi, le trajet rouge n'est pas indiqué comme mauvaise position pendant un certain laps de temps. En effet, la dernière station radio étant marquée au point vert, la localisation ne s'éloigne pas trop de la courbe de Bézier à ce moment là. Dès lors que la dernière antenne est connue, la courbe s'infléchit suffisam-

Figure 3.12 – Comparaison du trajet réel et *spoofé*

ment pour détecter l'anomalie. Il est même possible de recalculer les alertes rétroactivement.

3.3.3.2 QOS

Nous avons vu qu'il existe des infrastructures pour surveiller les signaux GPS (SBAS, *etc.*). Ces entités permettent de garder une trace des changements dans les informations du signal satellite. L'utilisation de ces données peut se faire *via* des services dédiés (comme EDAS). À l'aide de ces services, nous sommes en mesure de réaliser des analyses de signal, pour en déduire un niveau de fiabilité. Le module de qualité de service est un outil d'analyse *post-mortem* .

Le service QoS calcule les statistiques pour toutes les stations de référence situées à une distance de $1000km$ maximum (voir figure 3.7). Quelle que soit la position du récepteur en Europe, celui-ci devrait se situer dans le champ d'au

moins deux de ces stations, si ce n'est plus. Plus on est proche des stations, meilleure est la comparaison avec les statistiques du signal GPS.

Les mesures de la qualité s'intéressent plus particulièrement aux informations comme la dilution de précision horizontale, qui décrit la précision de la composante horizontale d'une solution GPS. On dispose aussi du pourcentage d'erreur horizontale en dessous de 5m, l'erreur horizontale pour le site à 95% de confiance. Le niveau de protection horizontale (HPL) est un cercle centré sur le récepteur qui indique une zone dans laquelle on est sûr que la position calculée se trouve pour le site à 95% de confiance ainsi que le pourcentage de temps pour lequel le nombre de satellites en vu est supérieur à 6.

Le tableau 3.1 indique quelle est la qualité pour un signal donné. La moins bonne mesure déterminant la qualité.

Table 3.1 – Tableau de calcul de la qualité

Maximum DOP	Maximum HPL	Maximum 95%	Qualité
N/A	N/A	N/A	faible
3	30	6	medium
2	20	4	haute
1	10	2	très haute

L'intérêt de ce service est de pouvoir qualifier la géolocalisation malgré l'absence d'infrastructures particulières (GSM, Wi-Fi, *etc.*) au moment de la capture. En effet, lors de la capture d'une image, il n'y a pas de garantie que l'une de ces infrastructures soit disponible. Pouvoir effectuer un contrôle *a posteriori* permet d'obtenir tout de même un niveau de confiance. L'intégration de cet élément à la chaîne de certification est aisé. Toute photographie est présentée avec un rapport qui indique les actions qui ont été menées depuis sa création. Les éléments de QOS y sont intégrés pour donner une meilleure indication de fiabilité.

3.3.3.3 EHL

Le module *Enhanced Location Module* est conçu pour améliorer les positions GNSS calculées. Contrairement à QoS, ce module s'utilise de manière instantanée. De fait, les informations utilisables ne sont pas les mêmes que le module précédent.

Pour ce faire, il est indispensable de surveiller en permanence le signal GNSS. Des stations de références du réseau *SBAS* sont situées à des points connus sur la planète. Ce réseau permet de surveiller plusieurs éléments relatifs à la position des satellites. Il est possible de réaliser une carte des per-

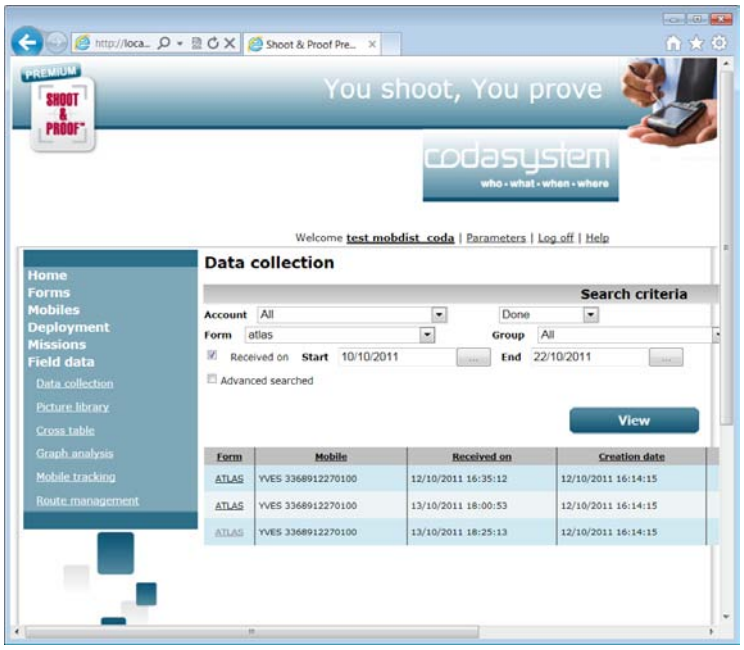


Figure 3.13 – Interface utilisateur web 1

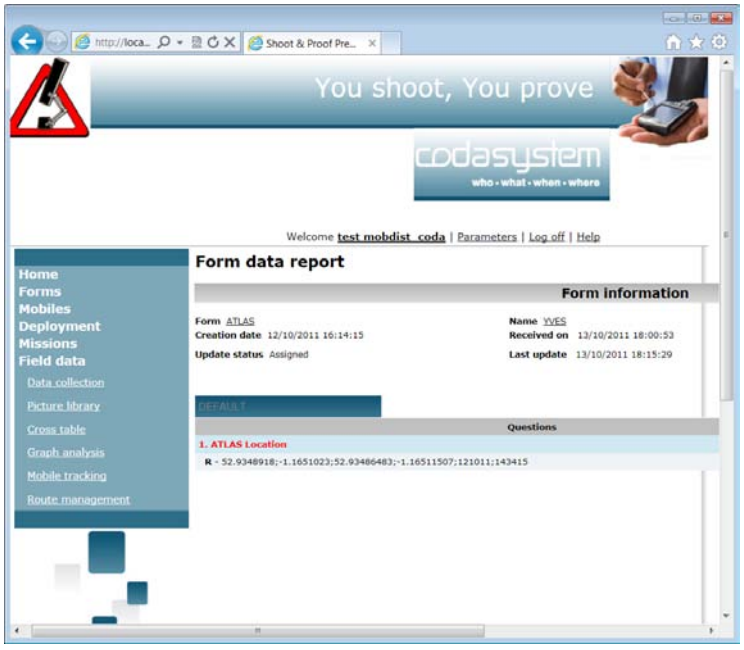


Figure 3.14 – Interface utilisateur web 2

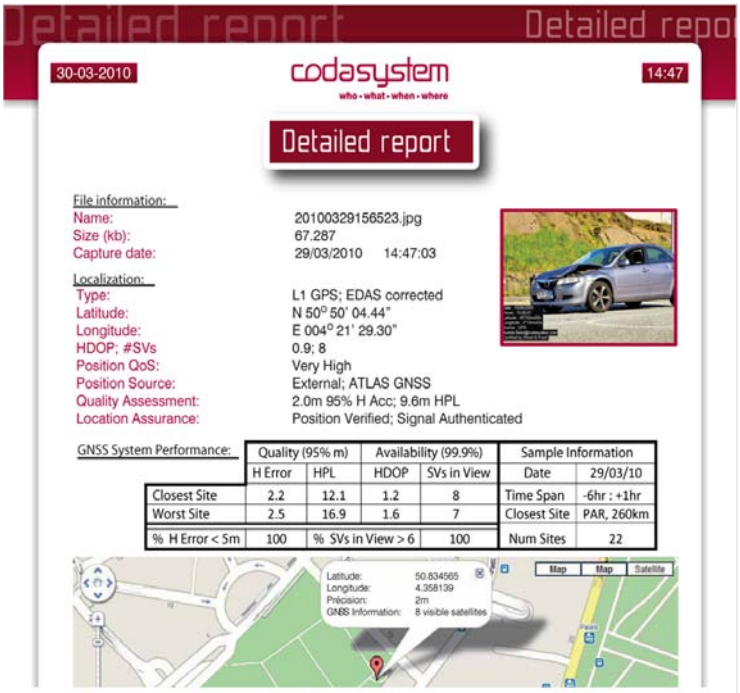


Figure 3.15 – Rapport de photographie 1

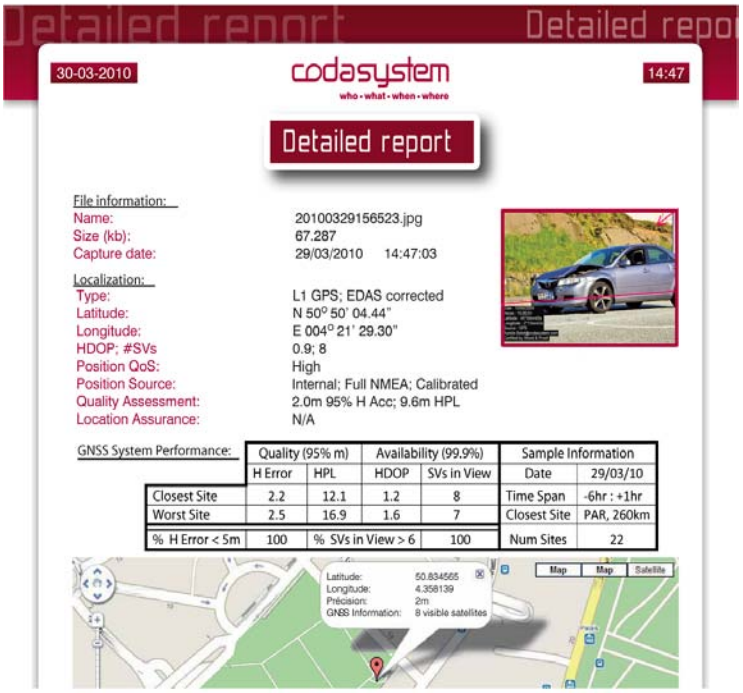


Figure 3.16 – Rapport de photographie 2

turbations ionosphériques grâce aux données reçues sur les deux fréquences L1 et L2. Le regroupement des données collectées par l'ensemble des stations permet de calculer plusieurs informations. Tout d'abord, des corrections à long terme sur les erreurs d'orbites, dérivée des éphémérides et des calculs locaux. Ensuite, des corrections long et court terme sur les erreurs d'horloge des satellites, le temps dans l'espace s'écoulant différemment. Il est donc nécessaire de réaliser de petites corrections pour obtenir une mesure exacte.

Voici la liste des informations que l'on peut obtenir du service EHL :

- La précision à 95%, qui indique à partir de quelle marge (en mètre) on dispose d'une certitude à 95% d'être à cette position.
- Le marqueur d'assurance qui indique si c'est l'appareil dédié qui fournit le signal
- Le marqueur de certification qui indique si l'appareil dédié à la même position que le téléphone.
- Drapeau de correction EDAS, qui indique si la position a fait l'objet d'une correction où non.
- La dilution de précision horizontale (DOP) qui indique si les satellites sont suffisamment éloignés pour calculer une bonne position.
- Le niveau de protection horizontale, qui donne un cercle centré sur la position réelle, dans lequel on est sur de trouver la position horizontale indiquée.
- Le marqueur d'interférence lorsque l'on a détecté une interférence.
- Le nombre de satellites.

L'utilisation de ce service se fait par l'appel à un *web service*. Ce service utilise en entrée une trame NMEA. Cette norme est une spécification pour la communication entre équipements marins, dont les équipements GPS. Elle permet de spécifier un grand nombre de message relatifs à la géolocalisation (plus d'une trentaine de trames GPS différentes). Seule une partie de ces types de trames (GPGGA par exemple) est utilisée pour ATLAS, parfois en combinaison. En fournissant une(des) trame(s) NMEA suffisamment complète(s), le serveur peut effectuer des mesures, puis interroger les données des *GBAS*. Cela permet de donner une indication de qualité de service d'une part et, d'autre part, de réaliser les corrections nécessaires le cas échéant. Ces corrections portent sur la longitude et la latitude. Les mesures sont susceptibles d'être corrigées d'une amplitude de 0 à 1,5 mètres.

Dans la chaîne de certification, ce contrôle est effectué à chaque géolocalisation. Une géolocalisation peut être faite régulièrement pour obtenir les données nécessaires au rafraîchissement de l'indicateur de qualité. La position corrigée est intégrée directement dans le message qui sera ensuite tatoué. En revanche, le niveau de qualité de service est affiché sous la forme d'une icône, visible à tout moment par l'utilisateur. Cela lui permet de choisir, en connaissance de cause, s'il désire faire une capture de photographie certifiée avec cette position. On peut voir les icônes de notification sur la figure 3.17 ainsi que sur une capture d'écran du mobile comme la figure 3.18.

Figure 3.17 – Liste des icônes de notifications

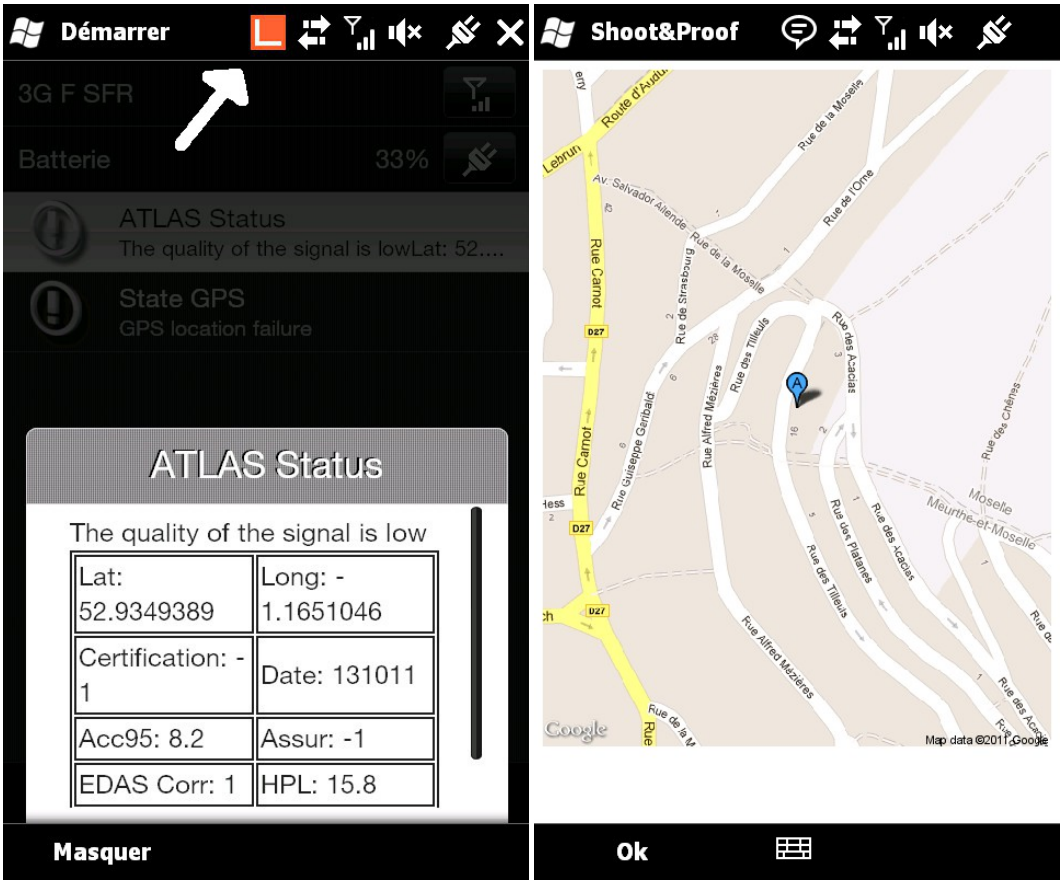


Figure 3.18 – Interface mobile

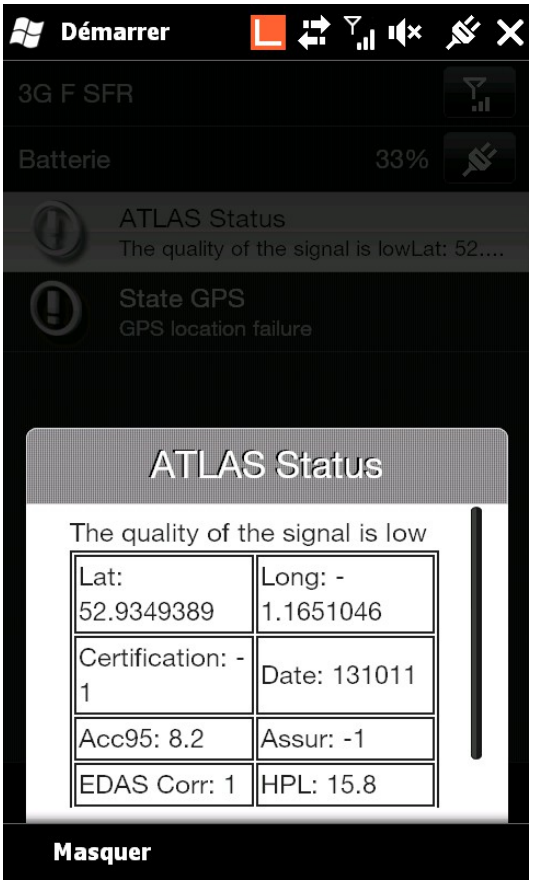


Figure 3.19 – Icone "L" d'indication "Low quality" en rouge

3.3.3.4 GAM

Construire un appareil de capture de signaux satellites présente deux avantages. Premièrement, le contrôle de la fabrication de l'appareil permet de choisir avec précision la manière de capturer et traiter le signal. Les données qui sortent de l'appareil ne sont pas limitées par les APIs des constructeurs et intégrateurs. Deuxièmement, nous ne sommes pas limités à une constellation unique. C'est avec ces considérations que la société NSL a développé son propre équipement.

Ce dispositif intègre une antenne, un module de décodage radio-fréquence ainsi que la capacité à stocker de l'information sur un support externe. Il implémente également le protocole Bluetooth qui lui permet d'échanger ses données avec un autre équipement (ordinateur, tablette, smartphone, etc.). Il est prévu pour être en mesure de lire les informations provenant des satellites GPS, Galileo et Glonass (Seulement deux utilisables en simultané). Les spécifications de ce module requièrent la portabilité, une grande autonomie et la non perturbation des opérations du téléphone mobile associé.

Le développement du composant se base sur la technologie SDR (*Software-defined radio*) car celle-ci apporte une grande flexibilité. La philosophie de cette technologie est de minimiser la quantité de matériel nécessaire et d'exporter les fonctions complexes au niveau logiciel (serveur) dès que possible.

Dans la stratégie SDR, le serveur s'occupe de l'acquisition, de la mesure et la détermination de la position. Le serveur *RetroFix* opère donc sur des données brutes provenant du récepteur. Ce service apporte des bénéfices substantiels : consommation de l'équipement mobile très faible, positionnement à la demande (service web), mise à jour continue des algorithmes de vérification d'intégrité sur le serveur, possibilité de calcul de solutions multi-constellations.

Les données envoyées et reçues par le récepteur se composent des messages suivants :

```
Date: [current date from smartphone]
UtcTime: [current time from smartphone]
PosLlh: [latitude longitude height from GSM data]
Binary: [payload length]
[payload from receiver data]
```

```
Status: [OK or FAILED]
Utc: [date and time determined by Retrofix]
GpsTime: [gps time]
PosXyz: [geocentric coordinates determined by Retrofix]
PosLlh: [latitude longitude height determined by Retrofix]
```

L'ensemble des échanges du scénario GAM est présenté dans la figure 3.21. Dans un premier temps, le smartphone est appairé avec l'antenne externe. L'antenne joue le role de serveur et attend les requêtes du smartphone. Lorsque le smartphone demande une géolocalisation (par exemple avant le tatouage), l'antenne transmet, sous forme d'un fichier binaire, le signal brut capturé pendant un intervalle de temps de 200ms. Le smartphone transmet alors tel quel ce fichier au serveur RetroFix via une requête https classique. Une fois que le serveur a calculé la meilleure position à partir des informations brutes, il retourne la géolocalisation au smartphone, qui peut dès lors l'utiliser (par exemple la tatouer effectivement) avec une qualité et une authenticité améliorée.

L'intérêt de ce récepteur est double. Il permet de passer outre les restrictions des constructeurs, intégrateurs et développeurs des technologies mobiles qui parfois ne laissent pas accès à tout les composants de géolocalisation. Or, les améliorations présentées dans nos expérimentations nécessitent un minimum de données techniques recueillies par le récepteur GPS. Ensuite, ce récepteur combine la réception du signal de plusieurs GNSS. De ce fait, il est possible de corréler les différentes informations des constellations pour vérifier la cohérence des données.

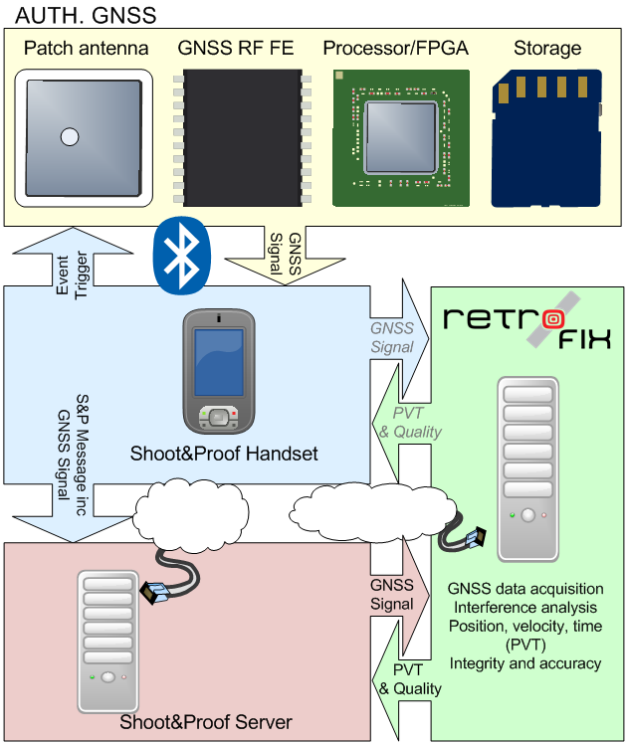


Figure 3.20 – Architecture du scénario récepteur dédié

3.3.4 Résultats

Cette partie de l'étude a comporté deux phases. La première, la plus fondamentale, a porté essentiellement sur le choix et la conceptualisation des scénarios avec la prise en compte constante des contraintes liées à la chaine de certification (sécurité, efficacité, etc.). La contribution de chaque partenaire dans le développement des composants est variable d'un scénario à l'autre et dépend des compétences principales de chacun. Notre participation aux développements s'est clairement effectuée dans le cadre de la société CodaSystem.

L'ensemble des composants logiciels nécessaires aux quatre scénarios décrits dans les sections précédentes, ont été développés collaborativement par les partenaires du projet ATLAS. Le composant SCM a été développé essentiellement par un ingénieur d'étude du LITA avec notre collaboration. Nous avons réalisé l'essentiel des composants des scénarios QOS et EHL avec la collaboration des ingénieurs de NSL pour adapter les protocoles de communications entre nos composants est les serveurs de NSL. Enfin, le développement (physique) de l'antenne du scénario GAM a naturellement été assuré par NSL. Le composant logiciel assurant, sur le smartphone, la passerelle entre l'antenne et le serveur *RetroFix* a été réalisé par l'ingénieur d'étude du LITA. Si tous les composants sont opérationnels indépendamment les uns des autres, l'ef-

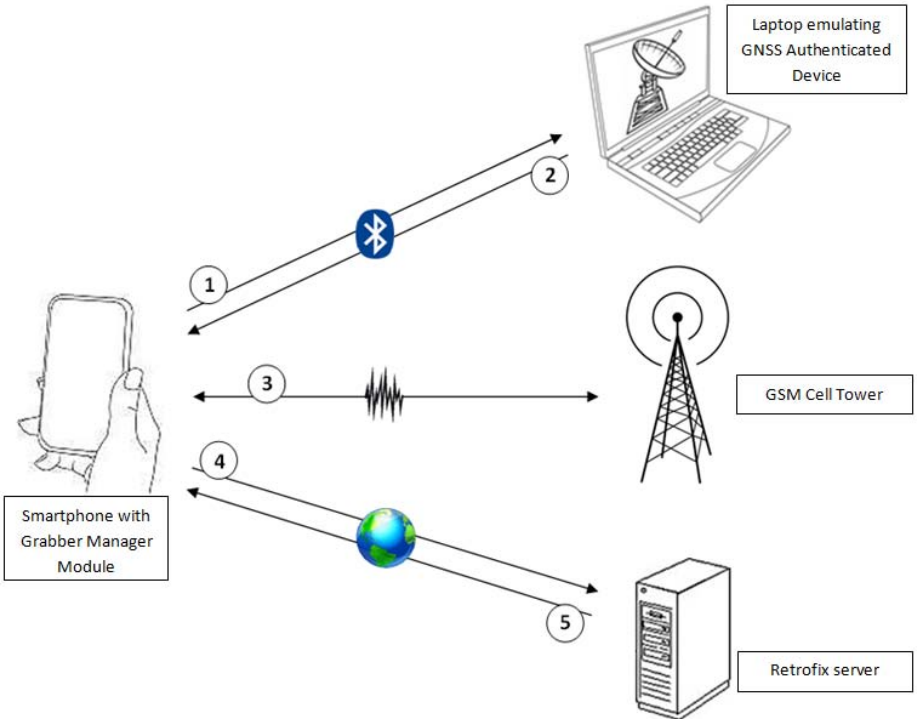


Figure 3.21 – Protocole GAM

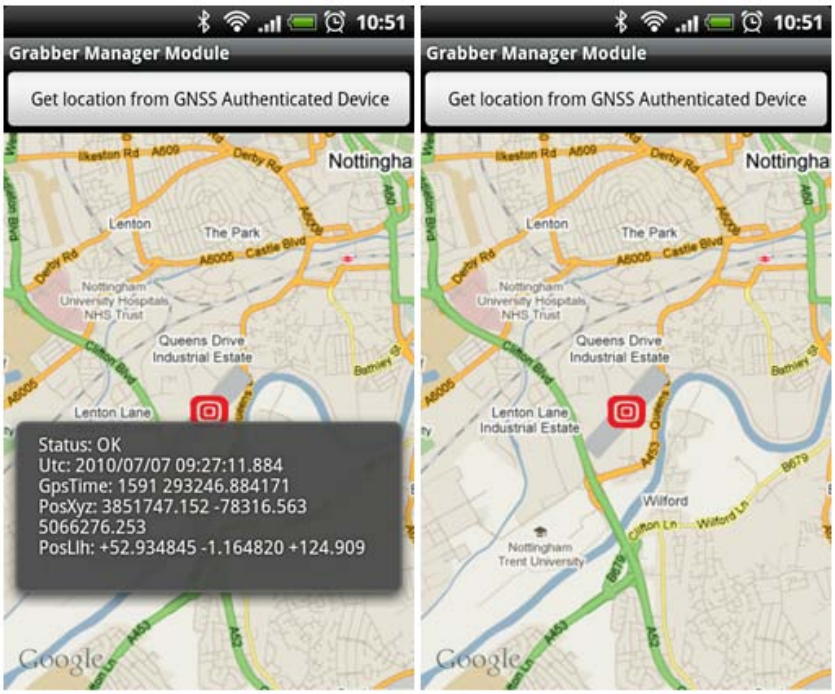


Figure 3.22 – Interface GAM sur mobile

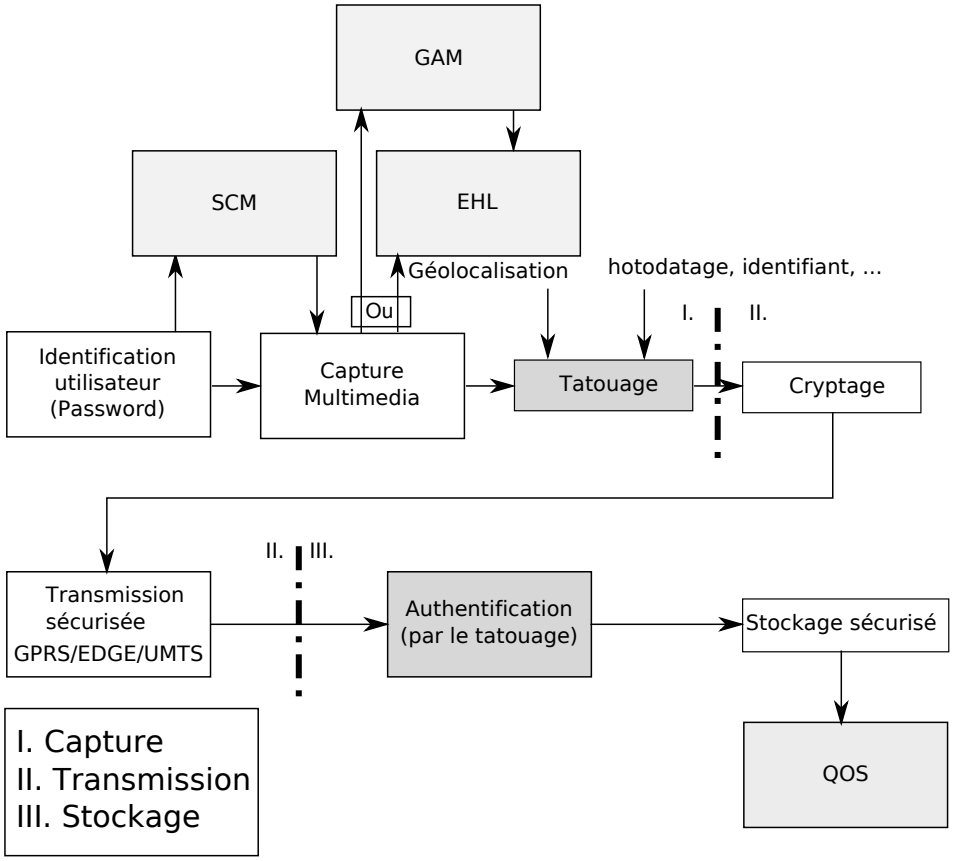


Figure 3.23 – Modules intégrés à la chaîne de certification

fort d'intégration de ces composants dans la chaine de certification (voir figure 3.23) n'a pas encore été réalisé et sort du cadre de notre travail.

Le caractère original de ce travail sur la géolocalisation sur smartphone rend toute comparaison avec d'autres solutions existantes impossible. De plus, les protocoles de tests prévus dans le projet n'ont pu être déroulés complètement. Il est à noter que ces tests sont particulièrement exigeants. Il est bien entendu exclu d'attendre une éruption solaire pour valider la détection d'une perturbation involontaire. La détection d'une usurpation est en revanche techniquement possible. Mais en pratique, elle requière des équipements spécifiques (*spoofer*) qui n'ont pas été disponibles, et dont la mise en œuvre n'était pas incluse dans les actions de CodaSystem. Les seuls tests réalisés sont ceux qui illustrent les différents scénarios. Ils restent très localisés. Leur nombre et leur échelle sont insuffisants pour constituer une preuve expérimentale valable. Une perspective à notre travail serait de combler cette lacune en procédant à un ensemble de simulations et de tests réels pour confirmer en pratique la validité théorique des scénarios proposés.

Néanmoins il est possible depuis les résultats collectés à partir de ces tests,

d'identifier certaines limites. Le SCM présente plusieurs inconvénients. Tout d'abord, il nécessite une surveillance constante (Wi-Fi, GSM, GPS), qui peut être gourmande en batterie. Ensuite, et par construction, la plupart des algorithmes du SCM détectent uniquement les changements d'environnements. Or, il peut être difficile de définir la situation de l'utilisateur au moment du démarrage du système, notamment s'il est déjà en situation d'attaque. Concernant les méthodes de corrélation, il existe une certaine latence. Par exemple, le changement d'antenne GSM ne suit pas un algorithme prévisible et dépend de nombreuses de conditions dépendantes de l'opérateur téléphonique. De plus, le calcul des courbes de Bézier peut potentiellement être fortement modifié sur les derniers segments lors de l'ajout d'un nouveau point. Cela implique que des alertes qui auraient pu être levées plus tôt (si on avait eu connaissance de ce nouveau point) seront retardées. Le SCM a cependant des avantages. C'est un composant qui est quasiment autonome. Seules les données des réseaux GSM sont nécessaires pour faire les corrélations. Ce composant est compatible avec un grand nombre de terminaux actuels, dont les APIs fournissent suffisamment d'informations. Le principe de corrélation du signal GPS avec l'infrastructure GSM peut être transposé dans d'autres infrastructures. En particulier, des services de géolocalisation existent pour les bornes Wi-Fi. Du point de vue technique, le principe est rigoureusement le même que pour l'infrastructure GSM. Mais, cette infrastructure présente au moins deux inconvénients. Le premier est la faible densité de la couverture Wi-Fi. En effet, la portée d'une antenne Wi-Fi est nettement plus limitée que celle d'une antenne GSM. Et même si le nombre d'antennes Wi-Fi est beaucoup plus important, ces antennes sont très inégalement réparties. De nombreuses antennes se recouvrent (ce qui n'apporte aucune plus value pour la corrélation), et de nombreuses zones ne sont pas couvertes correctement. Le deuxième inconvénient est la volatilité des données dédiées à la géolocalisation des antennes Wi-Fi (par exemple Google Gears API, ou encore Combain location API). Elles sont difficiles à tenir à jour par rapport à la réalité. Les antennes Wi-Fi sont liées aux utilisateurs lesquels déménagent, se désabonnent puis se réabonnent chez un autre fournisseur d'accès, *etc*. Cela n'apporte pas la fiabilité recherchée dans notre contexte. Néanmoins, comme on l'a déjà observé, l'évolution des comportements des utilisateurs induit une importance plus grande de la géolocalisation dans leur vie quotidienne. Cela devrait conduire à une amélioration de la maturité de ces bases de données. Elles pourraient devenir une source de plus en plus fiable.

Les informations fournies par le module QOS sont importantes, mais cette approche reste limitée. En effet, elle informe l'utilisateur quelques heures après que la photo a été réalisée. Cela n'est pas adapté à tous les usages. Les résultats obtenus sont d'ordre technique et ne peuvent pas être interprétées par des utilisateurs néophytes. Ce n'est pas autonome et nécessite le recours à un service tiers.

Le module EHL présente aussi quelques inconvénients. Les corrections ne peuvent être effectuées que dans un rayon de 1000 kilomètres d'une station de

référence. Au delà, celle-ci n'est plus valable. En Europe, la couverture GBAS est suffisante pour pallier ce problème, mais ce n'est pas le cas sur les autres continents. De plus la qualité du service nécessite un appel avant la première géolocalisation. Comme pour QOS, ce module requiert l'accès à un service tiers NSL) pour obtenir les analyses des données EDAS.

L'inconvénient du module GAM est l'obligation pour l'utilisateur de gérer un autre équipement que son smartphone habituel. De plus, cet équipement n'est pas autonome et requiert la disponibilité du service de géolocalisation tiers.

3.4 Conclusion

Dans ce chapitre, nous avons abordé trois axes. Premièrement, nous nous sommes attachés à décrire les GNSS : architecture, systèmes d'améliorations ainsi que les vulnérabilités. Concernant ces dernières, nous avons pu les classer en perturbations intentionnelles et non intentionnelles. Les perturbations non intentionnelles sont largement dues aux phénomènes naturels comme les perturbations ionosphériques, mais aussi à cause des interférences d'origines humaines comme la radio-fréquence. Les perturbations intentionnelles sont liées en général, à une volonté de nuire. Ainsi, on peut organiser le brouillage d'un espace, ou encore aller jusqu'à modifier les signaux des GNSS pour falsifier les positionnements des utilisateurs.

Ces perturbations ne sont pas inéluctables. Il existe des moyens pour les réduire, voire les éliminer. Dans la seconde partie du chapitre, nous avons mis en évidence les différentes techniques d'atténuation d'effet connues. Nous nous sommes attardés plus précisément sur les méthodes logicielles, domaine qui nous est accessible (à la différence des méthodes matérielles). L'état de l'art a montré que la plupart des techniques sont basées sur la surveillances du signal GNSS. Les variations trop importantes sont un signes de perturbations. Certaines techniques permettent de corréler les différentes fréquences de signal utilisées pour une même constellation. Peu de travaux portaient sur l'utilisation de moyens externes aux signaux des GNSS, comme en particulier d'autres infrastructures (autres constellations, infrastructures GSM, etc.).

La troisième section présente notre contribution. Nous avons identifié quatre types de contrôles logiciels pouvant améliorer la confiance dans le positionnement par satellites. Les contrôles *a posteriori* sont les contrôles effectués lorsqu'aucune mesure n'était possible au moment de l'acquisition du signal. Ils permettent à un utilisateur de disposer d'une méthode qui reste disponible quelles que soient les conditions dans lesquelles il peut se trouver. En revanche, les contrôles *a priori* permettent d'alerter un utilisateur des problèmes détectés bien avant l'acquisition d'une position. L'utilisation de méthodes de surveillance de signal, combinée avec des outils de corrélation de position GSM ou Wi-Fi, permet en effet de fiabiliser le positionnement GNSS en ajoutant une validation du signal ainsi qu'une seconde source de positionnement. Les

contrôles *correctifs* sont effectués lors d'une capture de position. Ils permettent de mettre à profit les infrastructures d'amélioration de position GBAS) pour corriger les positions reçues par un récepteur civil. Enfin, les contrôles *externes* utilisent un équipement dédié permettant un contrôle total sur les signaux reçus par les constellations. Le projet ATLAS nous a permis de développer et intégrer chacun de ces types de contrôles sous la forme de modules dans la chaîne de certification mobile que nous avons présentée au chapitre 2.

Ces contrôles ne sont pas infaillibles. Mais, en donnant à l'utilisateur un maximum d'information, il est en mesure de savoir avec plus de détails à quoi il est confronté. Le niveau de confiance dépend aussi des précautions que va prendre l'utilisateur. En effet, s'il se trouve dans une zone non couverte par les infrastructures d'amélioration de position, ou par les services GSM/Wi-Fi, une partie des contrôles restera inopérante.

La mise en œuvre de ces modules a aussi permis de remarquer l'importance du tiers de confiance. En effet, la plupart des contrôles ne sont pas autonomes. Ils se basent sur l'utilisation d'infrastructures (SBAS, GSM, ...) et de savoir-faire ou algorithmes particuliers (serveur Retrofix par exemple). Il est donc impératif de pouvoir faire confiance à ces différents fournisseurs de services, sans quoi les contrôles sont vains. Ces protocoles ne sont pas sans rappeler d'autres protocoles en sécurité informatique. Notamment la cryptographie à clef publique, ou par extension les certificats utilisés dans les PKI. Dans ces protocoles, il est nécessaire de disposer d'une tierce partie qui fait office de garant à un moment clef du processus : la création et la gestion des certificats. Ce recours à des tiers de confiance est indispensable compte tenu des scénarios d'attaque que nous avons identifié dans la chapitre 1. En effet, dans certaines attaques comme la construction d'un alibi, l'attaquant est l'utilisateur lui-même ! De fait, dans le cas général, on ne peut pas le considérer comme un maillon fiable de la chaîne de certification. C'est ce qui explique que chaque donnée doive être validée de manière externe.

Nous allons retrouver cette nécessité pour réaliser un horodatage fiable d'une photographie. Le temps, comme le lieu, est une information vitale pour la preuve. En effet, nous allons voir au chapitre A, que la mesure du temps est plus complexe qu'il n'y paraît. Dans la mesure où l'horloge du smartphone est sous le contrôle de l'utilisateur, son utilisation pour la certification est d'emblée interdite.

Annexe A

Authentification des méta-données

Dans le chapitre 2, nous avons étudié les propriétés d'un tatouage d'image. La classification des attaques nous a permis de déduire les propriétés particulières que doit posséder un « tatouage de certification » : imperceptibilité, capacité, performance, semi-fragilité et localisation. Un protocole cryptographique robuste a été élaboré pour agencer ces marques évitant qu'un attaquant ne puisse retrouver et évaluer ces marques sans se heurter à un problème calculatoire difficile.

En définitive, l'algorithme de tatouage a deux entrées, une image et une suite d'octets. Il garantit que ces octets seront correctement tatoués dans l'image en vertu des propriétés de certification et que l'intégrité de l'image sera vérifiable et sûre. Les autres éléments de la chaîne de certification vont à leur tour garantir la transmission sécurisée, puis le contrôle et le stockage sécurisé de cette image. Mais le tatouage n'a pas la responsabilité de vérifier la qualité des informations qu'il place dans l'image. Il effectuera le même traitement, que les octets correspondent à des données aléatoires ou à des données réelles qui composent une preuve (auteur, lieu et date). L'authenticité de ces données est donc primordiale pour amorcer la suite des opérations qui conduit à la production puis à la gestion d'une image probante.

Les protocoles d'authentification se décomposent en plusieurs types. Les protocoles *arbitrés* impliquent le recours à une autorité externe aux parties souhaitant s'authentifier. Cette autorité est généralement appelée *tiers de confiance*. Un tiers de confiance est un organisme agréé par l'autorité gouvernementale d'un pays après un audit très strict des procédures suivies par cette organisation. En général, ce type de protocoles offre une sécurité maximale. En contrepartie, ils sont lents et coûteux. Les protocoles avec *juge-arbitre* limitent l'intervention de l'autorité, celle-ci n'étant sollicitée que lorsque qu'un conflit intervient. Ces protocoles sont plus rapides que les protocoles arbitrés. Enfin, les protocoles à *discipline intrinsèque* n'utilisent pas d'autorité pour gérer l'authentification. C'est la construction du protocole lui-même qui garantit aux participants une authentification mutuelle fiable. Ces protocoles sont les moins coûteux, mais ils ne sont pas applicables dans toutes les situations.

Nous décrivons dans ce chapitre comment l'authenticité des données constituant la preuve est prise en compte. Dans tous les cas, l'authenticité est le résultat d'un protocole arbitré, c'est-à-dire recourant à un tiers de confiance. Dans

le contexte de notre étude, l'authentification de l'auteur et de l'horodatage n'a pas fait l'objet de contributions particulières. Nous rappelons simplement les dernières avancées et tentons de donner un cadre cohérent pour traiter cette problématique. L'authentification de l'utilisateur est un problème ancien qui a fait l'objet de nombreux travaux. Nous rappelons dans la section A.1 les principales approches, notamment celles qui prennent spécifiquement en compte la problématique des *smartphones*. La section A.2 décrit l'authentification de l'horodatage. Enfin, dans la section A.3 nous reviendrons sur la géolocalisation. Nous résumerons brièvement nos contributions et les re-situerons par rapport à la notion de tiers de confiance.

A.1 Authentification de l'auteur

Il est important de distinguer l'identification de l'authentification. Identifier une personne consiste à déterminer quelle est la personne avec qui l'on communique, alors qu'authentifier une personne vérifie cette identité.

Les méthodes permettant d'authentifier une personne se basent sur le fait de posséder un secret. Pour vérifier cela, la personne souhaitant authentifier présente un *challenge* à l'utilisateur. Si celui-ci est en mesure de réussir le challenge, on considère qu'il connaît le secret. Ce secret peut avoir plusieurs formes. Il peut être quelque chose que l'on sait (mot de passe), que l'on a (clé, jeton, badge) ou que l'on est (biométrie). On considère qu'une authentification est forte lorsque celle-ci se base sur l'utilisation d'au moins deux des types mentionnés.

L'authentification sur mobile s'est bien entendu inspirée des méthodes utilisées dans le contexte traditionnel de l'informatique. Cependant, il faut tenir compte de, voire exploiter, l'arrivée d'un nouvel acteur qui est l'opérateur téléphonique. Par exemple, il est courant de recevoir un code de confirmation sur son téléphone mobile (via un SMS) pour valider un achat par carte bancaire sur un site de commerce en ligne. Cet exemple est une technique d'authentification forte. Le premier challenge est la saisie du mot de passe lors de l'identification sur le site de commerce électronique (réseau Internet). Le second challenge est la saisie du code de vérification lors de la validation de l'achat (réseau GSM).

Nous allons aborder quatre types d'authentification pour lesquels le mobile apporte une nouvelle dimension :

- La PKI et les certificats
- Kerberos
- Les mots de passe
- Le comportement et la biométrie

A.1.0.1 PKI et certificats

La PKI est basée sur la cryptographie asymétrique pour laquelle l'utilisateur (ou une autorité) émet une clef publique et une clef privée. La clef publique doit

être exportée, pour que le monde extérieur puisse communiquer et authentifier la personne émettant la clef. Cependant, le problème de l'authentification est alors reporté sur la clef : comment être sûr que cette clef est bien celle de la personne que l'on veut authentifier. Les certificats émis par les PKI permettent de résoudre ce problème.

Une PKI (Public Key Infrastructure, figure A.1) est un ensemble de composants physiques et logiciels ainsi que des procédures pour la manipulation des certificats digitaux (voir figure A.1). Ces certificats sont actuellement aisément déployables sur tous les systèmes mobiles majeurs (iOS, Android, Windows Phone).

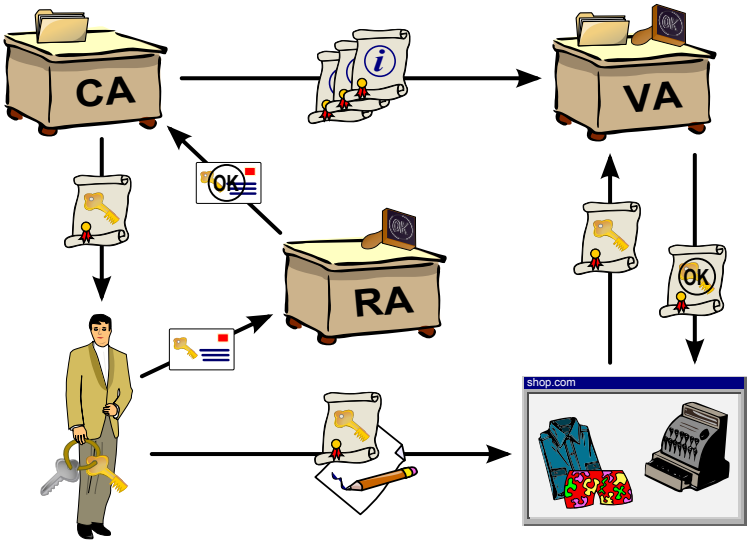


Figure A.1 – Schéma d'une PKI [Source : Wikipédia]

Les PKI, plus précisément la Autorité de certification - *Certification Authority* (CA), émettent des certificats contenant les clefs publiques de ses clients. Les certificats permettent de lier la clef à un utilisateur. Pour que la confiance puisse être établie, la PKI signe le certificat qu'elle émet, pour certifier que les informations qu'il contient sont fiables et vérifiées. Il existe donc une chaîne de confiance qui s'établit entre l'utilisateur et la CA ainsi que les sous-autorités de certification si elles existent. Pour que le système fonctionne, il faut que les utilisateurs accordent leur confiance à la CA.

La PKI se compose de plusieurs acteurs, remplissant différents rôles. La Autorité d'enregistrement - *Registration Authority* (RA) est en charge de l'inscription des utilisateurs dans la PKI. Elle peut aussi avoir une autorité en charge du stockage des clefs.

En 2009, Hassinen *et al.* de l'University de Kuopio, ont réalisé un paiement mobile via PKI [Hassinen 2006]. Les clefs privées sont stockées dans la carte

Subscriber Identification Module (SIM), qui n'est pas falsifiable, rendant ainsi les clefs non extractibles. Toutes les clefs publiques sont disponibles dans un annuaire et librement accessibles. Ainsi on peut signer avec la carte SIM et vérifier la signature par ailleurs.

Dans un cadre mobile, la PKI existe sous le nom de WPKI (WAP PKI - [Group 2009]). La PKI regroupe les même acteurs que la PKI avec, en plus, l'opérateur mobile pour l'intégration à la carte SIM. Ce processus implique des protocoles particuliers pour l'enregistrement des utilisateurs. L'opérateur mobile s'occupe en général de la création et distribution des certificats inclus dans la carte SIM. Une fois en possession de la carte, l'utilisateur peut s'inscrire à la RA/CA. L'usage des certificats se fait *via* deux canaux. L'un est le canal d'information, par lequel le service souhaitant authentifier l'utilisateur va communiquer le challenge. L'autre est le canal sécurisé. Il est utilisé en collaboration avec l'opérateur mobile pour transmettre la demande de signature. L'utilisateur signe ensuite la requête, qu'il renvoie par ce même canal sécurisé. Dans ce scénario, l'opérateur va jouer un rôle critique d'intermédiaire permettant d'établir le canal sécurisé avec le possesseur de la carte SIM.

[Moon 2009] indique que les PKI ne sont pas très adaptées pour l'environnement personnel. En effet, celles-ci sont basées sur l'utilisation de certificats qui, bien qu'efficaces dans une entreprise disposant d'un environnement homogène, ne sont pas adaptées à un ensemble d'équipements divers. L'article reprend des méthodes développées dans [Hwang 2006a] et [Hwang 2006b]. Le *inter-domain device authentication/access protocol* propose différents schémas pour décentraliser la PKI standard, grâce à des méthodes de délégation ou des PKI partagées permettant d'avoir des acteurs locaux. La méthode utilise des clefs publiques basées sur l'identité pour les usages mobiles.

A.1.1 Kerberos

Kerberos est un protocole réseau d'authentification développé par le Massachusetts Institute of Technology (MIT). Il est défini dans sa version 5 par la RFC 4120 [Neuman 2005]. Kerberos utilise un système de ticket en lieu et place des mots de passe en texte clair. Cela renforce la sécurité des systèmes et empêche l'interception des informations critiques par des tierces personnes. Kerberos est basé sur un ensemble de clefs secrètes partagées (chiffrement symétrique).

Le protocole fonctionne comme suit : le client envoie une demande d'authentification au serveur d'authentification (AS) pour une identité I et un service S. L'AS répond en renvoyant un ticket pour le serveur et une clef temporaire de session, le tout crypté par la clef du client. Le client transmet ce ticket au serveur. Le ticket contient l'identité du client et une copie de la clef de session, cryptée avec la clef du serveur. Cette clef de session peut désormais être utilisée par le client et le serveur pour communiquer. Dans ce protocole, le tiers de confiance est le serveur d'authentification (AS) qui émet les tickets.

Dès 2004, une équipe de l'université d'Australie occidentale a utilisé Ker-

beros pour des authentifications *ad-hoc* [Bouillon 2004]. De nos jours, Kerberos est disponible pour les appareils mobiles, dont le système d'Apple. Les bibliothèques pour développer les applications utilisant ce protocole son incluses dans le SDK. Bien qu'il n'y ait pas d'application officielle Kerberos sur l'*Apple store*, il existe une application Kerberos développée par Jay 'Saurik' Freeman (programmeur connu pour son travail sur les iPhones 'jailbroken' et sont alternative à l'Apple store nommée Cydia). Cette application fonctionne sur des iPhones 'jailbroken' et permet l'usage du protocole Kerberos. Il est aussi possible d'utiliser Kerberos sur le sytème d'exploitation Windows Mobile 5.0, cependant il semble que le protocole soit un consommateur excessif de ressources. En fait, cela prendrait quinze secondes au mobile pour recevoir le ticket original.

A.1.2 Mot de passe

Le challenge du mot de passe consiste à demander un secret à l'utilisateur sous la forme d'une chaîne de caractères. Certains mots de passe sont parfois stockés sous forme de texte clair, le vérificateur regarde alors si la chaîne est identique ou non. Cependant, on utilise maintenant les propriétés des fonctions d'empreinte (hachage) pour éviter de stocker les secrets sous leur forme visible. Ainsi on trouve souvent l'utilisation d'algorithmes comme anciennement MD5 et plus récemment les fonctions de la famille SHA. En général, le service qui requiert une authentification par mot de passe doit mettre en œuvre des procédures de gestion strictes. Il doit notamment s'assurer qu'un mot de passe respecte des critères de robustesse (nombre de caractères, présence de symboles de ponctuation, etc.). La politique de gestion des mots de passe doit aussi inclure des règles concernant le stockage, le renouvellement et la révocation des mots de passe. Définir et respecter ces règles de gestion des mots de passe, place le gestionnaire du service d'authentification dans le rôle d'un tiers de confiance. Lorsque l'on protège par un mot de passe ses coordonnées personnelles, voire ses coordonnées bancaires, il est bien entendu primordial d'avoir confiance dans le système.

Les mots de passe à usage unique, *One Time Password* (OTP), sont aussi de plus en plus utilisés. OTP permet l'usage de mots de passe dynamiques et est souvent utilisé en association avec les téléphones mobiles pour l'accès à des sites sécurisés ou à des fonctions critiques comme le paiement en ligne. Le serveur envoie un message (via SMS en général) à l'utilisateur. Celui-ci peut alors copier le challenge contenu dans ce message pour confirmer son identité. Cette méthode est considérée comme une authentification forte car le mobile joue le rôle d'un jeton de sécurité.

On trouve les mots de passe pour différents accès sur les mobiles. Tout d'abord, la carte SIM, véritable carte à puce, dispose d'un code PIN pour l'accès aux fonctions GSM standard et parfois d'un code PIN2 utilisé pour les fonctions avancées comme le transfert d'appel. Le téléphone lui-même peut être

protégé par un code (numérique, parfois alphanumérique ou encore un motif à tracer sur l'écran pour les nouveaux systèmes, figure A.2). Ce mot de passe permet de limiter l'usage du téléphone dans sa globalité (y compris les appels). Lorsque ces mots de passe sont égarés, il existe un système de récupération. Ainsi le code PUK permet de réinitialiser le code PIN. Ces codes sont en général d'une complexité plus importante que le code initial et ne sont connus que des personnes autorisées : détenteur de l'abonnement mobile pour la SIM (PIN, *PIN Unlock Key* (PUK)), gestionnaire du téléphone pour les codes maîtres de l'appareil.

Dans ce cas, la gestion du mot de passe revient à l'utilisateur. C'est lui qui décide s'il veut l'utiliser, s'il choisit un code PIN, un mot de passe alphanumérique ou un motif. L'avantage de ces mécanismes est leur simplicité. L'utilisateur choisit une méthode d'authentification qu'il peut retenir facilement, aisément interchangeable et qu'il peut ignorer. En contrepartie, la sécurité en est réduite. Ce n'est pas de l'authentification forte et les protections sont souvent désactivées ou les mots de passe trop peu compliqués. De plus, il est facile de récupérer l'information par *social engineering*.

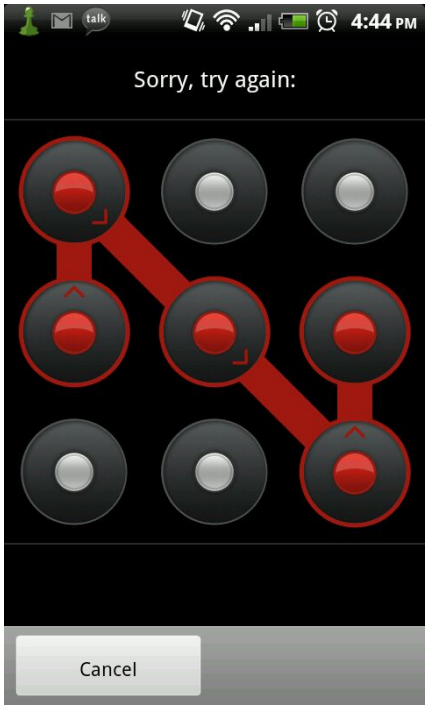


Figure A.2 – Motif de déblocage sur Android

Le plus gros problème des mots de passe sur *smartphone* tient au fait que la manipulation du téléphone n'est pas très aisée. Certes, les plus récents disposent maintenant d'un clavier complet, mais ils restent petits et mal adaptés à la saisie de mots de passe compliqués et donc suffisamment sécurisés. De

plus, l'utilisation des téléphones est discontinue. Demander un mot de passe à chaque sortie du mode veille va ennuyer l'utilisateur qui utilise son téléphone très souvent sur de courtes périodes de temps.

A.1.3 Comportement et biométrie

La biométrie permet de reconnaître de manière unique un individu. La technologie est déjà fort utilisée hors du contexte de la mobilité. En revanche, elle n'est que très peu présente sur mobile. On peut croiser des ordinateurs portables équipés de lecteur d'empreintes digitales, mais encore peu de téléphones disposant de la même technologie. Il existe aussi des prototypes de téléphones capables de reconnaître un utilisateur par sa voix ou par lecture de l'iris ou encore de la géométrie de la main.

Depuis quelques années, la recherche s'est portée sur des techniques plus avancées. La reconnaissance de comportement permet d'ajouter un facteur difficilement falsifiable à l'authentification. C'est un apport certain pour aller vers une authentification forte.

La reconnaissance de l'utilisateur par ces méthodes comportementales peut se faire sur divers critères :

- la voix
- la manipulation du téléphone (vibration, orientation)
- la manipulation des touches ([Buchoux 2008])

Toutes ces méthodes servent à réaliser une authentification implicite. En effet, l'utilisateur ne donne jamais d'information permettant de l'identifier de manière volontaire. C'est l'avantage principal de cette méthode, car elle est non intrusive. En revanche, la fiabilité est à relativiser : les résultats ne sont pas immédiats, l'authentification peut provoquer des faux-positifs ou des faux-négatifs.

Pour un service de réalisation de preuves, il est important de pouvoir fournir une authentification forte. Celle-ci doit être demandée pour chaque accès au service, garder en mémoire l'information d'authentification pourrait s'avérer problématique. Or nous avons vu qu'il n'est pas possible de faire confiance au téléphone lui-même. En effet, un service n'aurait aucun contrôle sur la gestion du secret et donc aucune fiabilité. Les meilleures méthodes seraient donc un challenge par mot de passe venant du service, un accès avec jeton ou encore la biométrie.

A.2 Authentification de l'horodatage

De nos jours, la planification des activités de la vie est une préoccupation importante. L'horodatage permet d'obtenir une date pour un événement. Horodater une image consiste donc à enregistrer le jour et l'heure précise où l'image a été capturée. Son utilité dans le contexte d'un tatouage de certification est

liée à la confiance que l'on peut y accorder. La qualité de cette preuve est donc basée sur les processus utilisés pour créer et manipuler les structures de données qui représentent l'événement, mais aussi des données qui permettent de situer l'évènement dans le temps.

Différents algorithmes existent pour la collecte, la transmission et la sécurisation d'une marque de temps. Les méthodes d'acquisition de temps sont nombreuses et variées. Elles peuvent se faire grâce à des horloges atomiques, prévues pour ne dériver que de quelques fractions de seconde toutes les décennies, comme c'est le cas dans les satellites. Mais, elles peuvent aussi être basées sur la lecture d'une horloge interne bien moins précise, mais accessible, comme pour un téléphone mobile.

A.2.1 Définitions et concepts

L'unité de base du système international d'unité (SI) est la seconde. Les minutes et heures, bien que non reconnues par le SI (non décimal, secondes de décalage) sont cependant acceptées dans les usages.

Le temps universel a été défini en 1972 pendant une conférence du Bureau International des Poids et Mesures (BIMP). Le NIST (US), NPL (UK), PTB (DE) et le CRL (JA) ont décidé de remplacer le système GMT (Greenwich Mean Time) par l'UTC (temps universel coordonné). UTC divise le temps en journée, heure, minute et seconde.

C'est la convention UTC qui est maintenant la plus utilisée dans les différents protocoles qui impliquent la gestion du temps. Par exemple, la RFC1305 ([Mills 1992]) définit le *Network Time Protocol* qui permet de synchroniser les horloges d'une multitude d'ordinateurs sur Internet.

Pour les constellations de satellites, GPS en particulier, le temps n'est plus synchronisé sur UTC. Il existe une divergence (*offset*) avec le temps atomique international (ITA). Les messages de navigation de GPS incluent la différence entre les temps GPS et UTC.

En avril 2009, le BIMP a donné une liste de quarante trois autorités habilitées à disséminer le temps UTC. En général, ces autorités fournissent le temps grâce au protocole NTP de la RFC 1305 ([Mills 1992]). Ce protocole est basé sur une architecture hiérarchique dans laquelle les données sont propagées par des requêtes NTP du plus haut au plus bas niveau (figure A.3). Dans les cas où la mesure de temps n'est pas critique, on peut aussi trouver le *Simple Network Time Protocol* de la RFC 2030.

Ces protocoles ont pour seul objectif de standardiser les échanges du temps pour, par exemple, synchroniser les horloges entre différents ordinateurs où qu'ils se trouvent. Ils n'assurent aucune fonction de sécurité. Ces requêtes sont transmises en clair sur le réseau et peuvent éventuellement être interceptées. Pour rendre ces échanges plus sûrs, en tout cas confidentiels, il convient naturellement d'utiliser des canaux de transmissions chiffrés (via des connexions SSL par exemple).

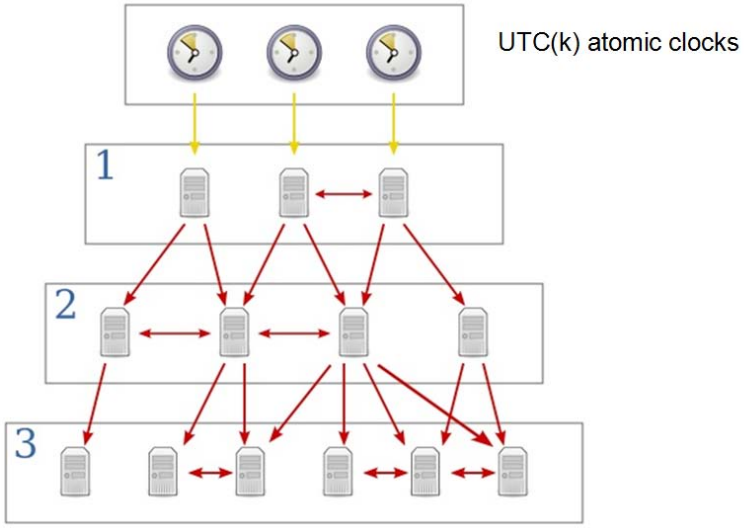


Figure A.3 – Hiérarchie des serveurs NTP [Domaine public]

A.2.2 Horodatage sécurisé

Nous savons qu'il est maintenant possible d'obtenir un temps fiable. Ce temps peut ensuite être utilisé pour horodater un document. Cependant, ces informations sont très facilement modifiables. De plus, ce temps est absolu. Il n'est lié à aucun contexte ou événement. Il est donc nécessaire de prendre des précautions avant d'utiliser une date comme preuve. En utilisant un service fournissant une marque de temps qui ne peut être forgée et qui lie la date à un contexte particulier, on obtient une preuve de l'existence d'un document à un temps donné.

L'horodatage d'un document numérique suivant un protocole particulier à été proposé dans [Haber 1991]. Les auteurs avaient pour objectif de construire un protocole qui ne permette pas à l'autorité d'horodatage de construire un faux, même si celle-ci le voulait. La première contribution apportée est l'utilisation d'une empreinte (*hash*). L'autorité ne manipule jamais le document original, mais uniquement son empreinte. L'association de cette empreinte avec l'horodatage signé par l'autorité, constitue un jeton. Celui-ci permet au receveur de vérifier que l'autorité a bien pris en compte et traité la demande. Pour éviter que l'autorité puisse effectuer des attaques lors du traitement d'une demande, les auteurs introduisent deux mécanismes : la liaison et la confiance distribuée. Le mécanisme de liaison permet de rendre les signatures dépendantes de la séquence des demandes faites à l'autorité. La confiance distribuée consiste à faire signer la demande par plusieurs autres clients. Forger un faux nécessite alors la coopération de tout le réseau.

Plusieurs RFC décrivent la manière d'obtenir un "horodatage sécurisé".

En particulier la RFC 3161 ([Adams 2001]). Ces RFC font aussi partie d'une spécification technique (TS) de l'Institut Européen des Standards de Télécommunications (ETSI). Le *Time Stamp Protocol (TSP)* décrit dans la RFC 3161 ([Adams 2001]) correspond à la TS 101 861 ([ETSI 2006]). On retrouve d'autres points de sécurité dans la RFC 3628, ([Pinkas 2003]) correspondant à la TS 102 023 ([ETSI 2003]). Le protocole TSP définit tout d'abord ce qu'est une *Time-Stamping Authority* - Autorité d'horodatage (TSA). La TSA est un *Trusted third party* - Tiers de confiance (TTP) (qui utilise une source de temps fiable (horloge atomique par exemple). Lorsque la TSA effectue une opération, elle inclut une date fiable. De même, elle utilise un nombre unique dans le jeton qu'elle délivre. Les messages envoyés se présentent sous la forme d'une requête *Time-Stamp Query* - Demande d'horodatage (TSR), générée par le client et contenant le *hash* du document à dater, ainsi que d'éventuels paramètres pour la TSA. La TSA répond avec un message *Time-Stamp Response* - Réponse (de demande) d'horodatage (TSR) contenant le jeton. Le fonctionnement de la TSA est représenté sur la figure A.4.

A.3 Conclusion

Le problème rencontré avec la fabrication de preuve est la fiabilité des informations qui la composent. Pouvoir authentifier ces données est donc primordial.

À l'heure actuelle, il n'existe pas d'attaque connue sur les protocoles d'horodatage. L'utilisation des protocoles présentés ici est donc plus que recommandée. La mise en œuvre des protocoles est simple et un prototype a été réalisé dans le cadre du projet ATLAS. En ce qui concerne l'authentification, cela n'est pas aussi clair. De multiples moyens d'authentification existent, mais nous souhaitons avoir une authentification forte, dans notre cas. Or il n'est pas possible de faire confiance au téléphone seul. Pour cela, nous recommandons soit l'usage de plusieurs challenges, mot de passe et jeton par exemple, soit l'usage de la biométrie.

Dans le chapitre 3, nous avons fait différentes propositions pour mieux authentifier la géolocalisation. Elles ont en commun de s'appuyer sur différents services regroupés au sein du même prestataire, jouant dès lors le rôle de tiers de confiance de géolocalisation. Bien entendu, à la manière des autorités de certification ou des tiers horodateurs (TSA), le prestataire délivrant un service de géolocalisation sécurisé devra satisfaire de nombreuses contraintes pour protéger son activité de toute négligence ou attaque, qu'elle soit interne ou externe. En particulier, les réponses contenant les informations de géolocalisation doivent naturellement être signées.

En définitive, chaque information utilisée composant le preuve est authentifiée par un tiers de confiance spécialisé. La nature de ces informations étant différentes, les procédures associées sont différentes. Chaque tiers de confiance

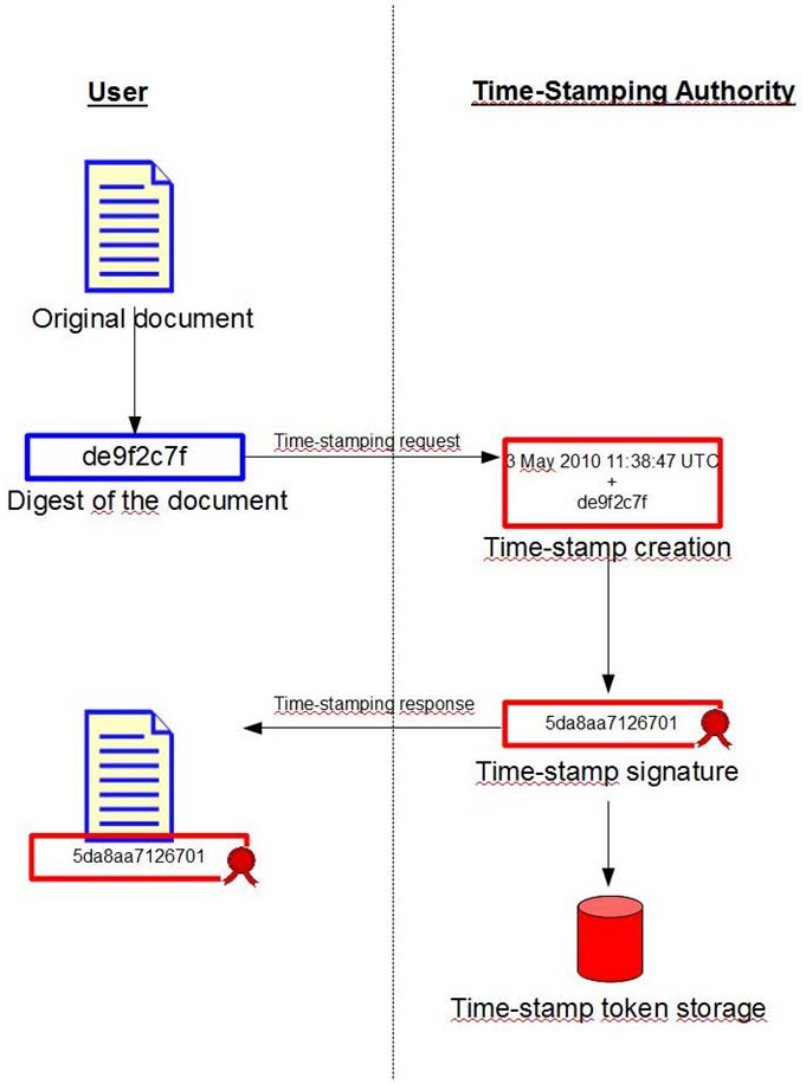


Figure A.4 – Fonctionnement d'une TSA [Source : projet ATLAS]

doit respecter un cahier des charges strict concernant la réalisation de ce procédures. Si ce cahier des charges existe pour l'horodatage (RFC) et les autorités de certification (CA) pour l'authentification de l'utilisateur, ce cahier de charge n'est en revanche pas définit dans le cadre de la géolocalisation.

Conclusion

Les équipements numériques ont subi de très nombreuses évolutions technologiques en un temps très court, tant sur l'aspect matériel que sur les capacités logicielles. Cette évolution s'est accompagnée d'un changement, également important, dans l'usage qu'en font leurs propriétaires. La possibilité de produire de l'information instantanément et de la diffuser massivement interpelle en particulier sur la valeur de ces informations. Dans le cas spécifique de la photographie numérique, il est de plus en plus nécessaire de pouvoir lui (ré)attribuer une valeur garantissant son originalité. Ce travail présente une technique de tatouage permettant d'accorder une valeur probante à une photographie capturée depuis un *smartphone* agissant dans le contexte d'une chaîne de certification. Cette chaîne de certification a pour but la création, protection et conservation d'une preuve ayant comme support une image. Elle est constituée d'une succession d'opérations protégeant l'image et les éléments de la preuve.

Dès le chapitre 1, nous avons défini les attaques pouvant cibler une preuve numérique. Cela nous a permis de réaliser ensuite une analyse de risques de la chaîne de certification selon la norme ISO27001. Cette norme débouche sur une certification, qui est un élément clef pour une entreprise qui se veut être un tiers de confiance. Dans le cadre du processus de normalisation, l'analyse de risque est une étape préliminaire cruciale. Cette analyse décrit les vulnérabilités, les impacts et les menaces qui sont pertinents dans le contexte du système d'information. Elle a mis en évidence des points critiques concernant différentes phases de la certification d'image. Des risques comme l'effacement de la marque incluse dans l'image, l'insertion d'une marque choisie ou encore la recopie de marques entre images restaient élevées. L'analyse a aussi mis en évidence le fait suivant : même avec un tatouage fiable, la preuve peut être contestée. En effet, il faut s'assurer de l'authenticité et de la fiabilité des informations qui sont tatouées dans l'image. Cela inclut l'authentification de l'utilisateur, l'authentification de l'heure de capture (l'horodatage), ainsi que l'authentification du lieu (la géolocalisation).

Dans le chapitre 2, nous avons étudié les propriétés d'un tatouage d'image. Nous avons ensuite réalisé une classification des attaques possibles sur le tatouage. Toutes les attaques ne sont pas pertinentes dans le contexte des images probantes. Aussi, avons nous proposé une nouvelle classification plus sémantique, qui identifie les scénarios d'attaque importants pour une image probante. Nous en avons déduit les propriétés particulières que doit posséder un « tatouage de certification » : imperceptibilité, capacité, performance, semi-fragilité et localisation. L'algorithme que nous proposons est dérivé de l'algorithme F5 que nous avons modifié pour répondre au mieux aux propriétés précédentes. L'image est tatouée bloc par bloc entre l'étape de quantification

et l'étape de compression du schéma JPEG. Pour obtenir le meilleur compromis entre invisibilité et capacité, nous avons choisi d'utiliser la méthode des *encoding matrix* en choisissant uniquement le bit le moins significatif des coefficients DCT. Le bit de message tatoué peut soit appartenir aux métadonnées, soit appartenir à l'empreinte assurant l'intégrité. Un protocole cryptographique solide a été élaboré pour agencer ces marques évitant qu'un attaquant ne puisse retrouver et évaluer ces marques sans se heurter à un problème calculatoire difficile. Tous les blocs ne sont pas propres au tatouage. Lorsqu'un bloc présente trop peu d'informations visuelles (un bloc de couleur unie), toute modification devient très sensible et peut constituer une fuite d'informations indésirable. Définir le seuil de tolérance pour le tatouage d'un bloc est un problème non négligeable. Un ensemble de tests valide l'implémentation de l'algorithme et permet d'établir de manière expérimentale le paramétrage approprié. Des pistes supplémentaires restent à étudier pour améliorer le tatouage. Tout d'abord, il serait judicieux d'étudier en détail la possibilité de réaliser un circuit électronique embarquant ce savoir-faire. Par cette méthode, on éliminerait la nécessité de travailler avec un fichier JPEG intermédiaire créé par le téléphone lors de la capture d'image . Ensuite, nous avons vu que les blocs homogènes peuvent poser problème. Notamment, remplacer un bloc quelconque par un bloc homogène risque de ne pas être détecté. Cependant on peut tenir un compte de leur nombre. Pour une succession de blocs homogènes, on inscrit le contenu de ce compteur en lieu et place de la marque d'intégrité du premier bloc de la série. Lors du décodage, on vérifie la cohérence du nombre de blocs homogènes successifs avec la valeur lue dans le bloc compteur. Si la photo reste une fonction importante des *smartphones*, les utilisateurs utilisent de plus en plus la fonction vidéo. Le problème de la force probante se pose également dans ce contexte. Certaines des idées proposées dans notre étude restent valables pour la force probante d'une vidéo. La chaine de certification, l'analyse des risques, l'authentification des informations sont des éléments communs. Mais la transposition du tatouage de certification pour la vidéo est beaucoup plus complexe. Les défis techniques sont exacerbés par le volume d'informations à traiter au fil de l'eau. Non seulement les images arrivent en nombre, mais les données de la preuve évoluent également. Par définition, l'horodatage change d'une image sur l'autre. Mais la géolocalisation peut également varier si l'utilisateur filme en se déplaçant. Il serait aussi intéressant de raffiner la métrique de détermination des blocs homogènes. En effet, la méthode actuelle ne traite que du cas des coefficients à zéro et, bien que cela soit suffisant d'un point de vue sécurité, une mesure plus complète pourrait prendre en compte l'amplitude des coefficients non nuls . Dans le même ordre d'idée, le PSNR bien qu'étant une mesure utilisée *de facto* dans la plupart des articles, ne reflète pas tous les aspects de dégradation perçus par le système visuel humain. La validation des mesures en utilisant la mesure SSIM [Wang 2004] pourrait également apporter de nouveaux éléments pour le raffinement des paramètres du tatouage. Enfin, le jeu d'essai pourrait être étendu (par exemple la base BOWS ou en-

core un modèle statistique comme [Srivastava 2003]). Toutefois, il conviendrait de tenir compte du fait que les images qui nous intéressent sont issues de la compression faite par un équipement mobile. De plus, tous les mobiles ne travaillant pas avec les mêmes paramètres de compression ou capacité de calcul, il conviendrait d'intégrer cette disparité dans la configuration de l'algorithme. Au final, l'algorithme proposé garantit une sécurité nécessaire et suffisante à la certification d'une image. Comme le montre l'analyse de risques, il est aussi important de collecter les informations à tatouer de manière fiable. En effet, tatouer, même de manière fiable, des informations erronées (fortuitement ou malicieusement) remet en question la force probante de l'image et rend caduques toutes les autres étapes de la chaîne de certification. Cela permet, par exemple, de créer des alibis en changeant les données de géolocalisation. Par conséquent, il convient de traiter cet aspect avec beaucoup d'attention pour, là encore, informer l'utilisateur des conditions de collecte, voire pour empêcher toute manipulation frauduleuse de la géolocalisation.

Dans le chapitre 3, nous décrivons l'architecture des systèmes de positionnement par satellites (GNSS). Puis nous détaillons les différentes vulnérabilités et attaques auxquelles la géolocalisation est soumise. Celles-ci sont d'ordre non intentionnel comme les effets dus à l'ionosphère ou les éruptions solaires, ou intentionnel comme le brouillage ou encore l'usurpation du signal qui cherche à positionner l'utilisateur à un endroit choisi. Ce chapitre s'attache à étudier les méthodes permettant de rendre la collecte plus fiable. La méthode la plus efficace pour authentifier la géolocalisation, aurait été de pouvoir authentifier les signaux par rapport à leur source. Cette possibilité n'existe qu'à des utilisations très restreintes et très contrôlées (militaire, secours, *etc.*). Pour le grand public, une des seules manières d'agir sur le téléphone mobile est de mettre en place des solutions logicielles. Notre contribution se compose de plusieurs types de contrôles permettant d'informer l'utilisateur des conditions de collecte du signal. Ces contrôles se décomposent en quatre types : contrôles *a priori*, *a posteriori*, correctifs et externes. Les contrôles *a posteriori* permettent à un utilisateur de disposer d'une méthode qui reste disponible quelles que soient les conditions dans lesquelles il peut se trouver. Les contrôles *a priori* permettent d'alerter un utilisateur des problèmes détectés avant l'acquisition d'une position. Les contrôles *correctifs* sont effectués lors d'une capture de position. Enfin, les contrôlés *externes* utilisent un équipement dédié permettant un contrôle total sur les signaux reçus par les constellations.

Les contrôles ne réduisent pas les possibilités d'attaques et ne les contrecarrent pas non plus. Cependant ils vont permettre de réduire l'impact de ces attaques. En effet, un utilisateur informé pourra prendre la décision de réaliser ou non sa preuve, sachant la qualité de sa localisation. Les solutions proposées sont essentiellement technologiques et s'appuient, à différents degrés, sur des prestations externes, notamment sur une entité capable de fournir une qualité de service de la géolocalisation, voire de calculer la géolocalisation. Cette entité apporte elle-même sa caution et se doit, par conséquent, d'être organi-

sée selon des procédures strictes et contrôlées. En définitive, elle tient lieu de tiers de confiance qui apporte sa garantie dans l'authenticité de la géolocalisation. Or, à ce jour, il n'existe pas réellement, à la manière de ce qui existe pour l'horodatage, de cahier des charges pour être tiers de confiance pour la géolocalisation, décrivant précisément les contraintes techniques et organisationnelles à satisfaire. Une autre perspective consisterait à s'appuyer idéalement sur les services authentifiés des GNSS. Ce niveau de service permettrait de garantir une authentification forte sur l'auteur du signal. Il existerait néanmoins des vulnérabilités liées au *rejeu*. En effet, un signal authentique enregistré puis réémis permet de perturber le receveur, lequel produit une géolocalisation obsolète. Par conséquent, même dans ce contexte pourtant plus sûr, des systèmes complémentaires destinés à accroître la confiance, à l'image de ceux proposés, devront être développés. Enfin, un cadre mathématique comme décrit dans [Martineau 2008], pourrait permettre une formalisation des travaux sur la géolocalisation, notamment en ce qui concerne les problèmes d'intégrité du système de positionnement par satellite déjà évoqués dans [Lacresse 2004].

L'authentification de l'utilisateur et de l'horodatage sont des domaines mieux connus. Ils ont fait l'objet des discussions du chapitre A. Pour l'authentification de l'utilisateur, il existe de multiples méthodes dans le cadre non mobile qui sont aisément transposables à la mobilité. L'important dans notre contexte étant d'obtenir une authentification forte. L'horodatage fait déjà l'objet de protocoles. Ces protocoles sont fiables et réutilisables sur mobiles en passant par les tiers de confiances mettant déjà en œuvre ces services.

Dans notre étude, nous avons décrit différentes méthodes qui permettent, tout au long de la création de la photographie, de garantir l'authenticité des processus d'acquisition et de conservation des méta-données au sein de l'image grâce au tatouage, pour fournir une image probante. Toutes ces méthodes reposent, comme nous venons de le rappeler, sur la collaboration avec différents tiers de confiance (pour l'horodatage et la géolocalisation). Ainsi, l'autorité de certification garantissant la preuve photographique, n'est autre qu'un nouveau tiers de confiance. Celui-ci intègre des tiers de confiance secondaires dans le cadre de ses propres procédures.

Les liens entre les *smartphones* et l'utilisateur se renforcent de plus en plus. Les téléphones proposent davantage de services et les usages deviennent toujours plus variés. Ces services sont parfois très critiques, comme par exemple le paiement en ligne, la consultation des comptes bancaires, mais aussi tout service impliquant l'usage de l'identité de l'utilisateur. Le téléphone devient donc un agrégat de données confidentielles et personnelles de son utilisateur. Il convient donc de s'assurer avec un maximum de précautions de la sécurité de ces informations, notamment en utilisant systématiquement le chiffrement des données et une authentification forte pour l'accès aux services. Par ailleurs, le *smartphone* permettant de produire de l'information, il est important que l'utilisateur puisse choisir de protéger une information dont il est l'auteur. L'image probante est un exemple qui devrait être généralisé à d'autres types d'infor-

mations (enregistrement sonore, texte, *etc.*). Enfin, la sécurité est aussi liée au comportement de l'utilisateur. Lorsque l'automobile s'est démocratisée et a envahi notre vie quotidienne, de très nombreuses procédures sont naturellement venues encadrer son usage (permis de conduire, assurance, *etc.*), notamment pour augmenter la sécurité des usagers (dans la voiture, et en dehors). Désormais, les *smartphones* occupent eux aussi un rôle important dans la société. Ils sont utilisés par tous. Or, personne n'est réellement formé aux bonnes pratiques et ni sensibilisé aux menaces dont ils font l'objet. Il apparaît évident que, comme pour l'automobile, des moyens techniques et des procédures doivent être trouvés pour protéger les utilisateurs en général. La photographie probante est justement une étape dans cette démarche. Bien d'autres restent à franchir pour transformer un smartphone en objet fiable et sûr, sur lequel on puisse reposer en toute confiance une partie de notre vie.

Bibliographie

[Adams 2001] Carlisle Adams, Entrust, Pat Cain, BBN, Denis Pinkas, Integris et Robert Zuccherato. *RFC3161 : Internet X.509 Public Key Infrastructure, Time-Stamp Protocol (TSP)*. Rapport technique, IETF : Network Working Group, August 2001. 49, 156

[Ahmed 1974] N. Ahmed, T. Natarajan et K.R. Rao. *Discrete Cosine Transfom*. IEEE Transactions on Computers, vol. 23, no. 1, pages 90–93, 1974. 32

[Atupelage 2008] Chamidu Atupelage et Koichi Harada. *PKI based Semi-Fragile Watermark for Visual Content Authentication*. World Congress on Engineering and Computer Science, 2008. 39, 40, 41, 42, 44, 56, 59, 60, 62

[Bas 2000] Patrick Bas. *Methodes de tatouage d'images fondées sur le contenu*. PhD thesis, Institut National Polytechnique de Grenoble, 2000. 30, 31, 39

[Benedetto 2005] Francesco Benedetto, Gaetano Giunta et Alessandro Neri. *A new color space domain for digital watermarking in multimedia applications*. In Image Processing, 2005. ICIP 2005. IEEE International Conference on, volume 1, pages I–249. IEEE, 2005. 57

[Bonnecaze 2006] Alexis Bonnecaze, Pierre Liardet, Alban Gabillon et Kaouther Blibech. *Secure time-stamping schemes : A distributed point of view*. In Annales des télécommunications, volume 61, pages 662–681. Springer, 2006. 49

[Bouillon 2004] Emmanuel Bouillon. *Kerberos et la Sécurité*. SSTIC, 2004. 151

[Buchoux 2008] A Buchoux et Nathan L. Clarke. *Deployement of keystroke Analysis on a Smartphone*. Australian Information Security Management Conference, 2008. 153

[Cayre 2005] François Cayre, Caroline Fontaine et Teddy Furon. *Watermarking security part one : Theory and Practice*. Proc. SPIE, vol. 5681, pages 746–757, 2005. 39, 43, 44

[Cox 2008] Ingemar J. Cox, Matthew L. Miller, Jeffrey A. Bloom, Jessica J. Fridrich et Ton Kalker. Digital watermarking and steganography. Morgan Kaufmann, 2008. 37

[Crandall 1998] Ron Crandall. *Notes on steganography*. Steganography Mailing List, page 297, 1998. 71

[Eggers 2001] Joachim J. Eggers et Bernd Girod. *Blind watermarking applied to image authentication*. In IEEE International Conference on Acoustics Speech and Signal Processing, volume 3. Citeseer, 2001. 38, 44, 58, 60, 62

[ETSI 2003] ETSI. *Electronic Signatures and Infrastructures (ESI); Policy requirements for time-stamping authorities*. Rapport technique, ETSI, 2003. 156

[ETSI 2006] ETSI. *Time stamping profile*. Rapport technique, ETSI, 2006. 49, 156

[Fei 2006] Chuhon Fei, Deepa Kundur et Raymond H. Kwong. *Analysis and Design of Secure Watermark-Based Authentication Systems*. IEEE Transactions on Information Forensics and Security, vol. 1, pages 43–55, March 2006. 40, 41, 59, 62

[Fridrich 1998a] Jessica Jiri Fridrich. *Image Watermarking for Tamper Detection.* In ICIP (2), pages 404–408, 1998. 40, 42, 56, 59, 62

[Fridrich 1998b] Jessica Jiri Fridrich. *Methods for detecting changes in digital images*. In IEEE Workshop on Intelligent Signal Processing and Communication Systems, 1998. 60

[Fridrich 2002a] Jessica Jiri Fridrich. *Security of fragile authentication watermarks with localization*. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 4675, pages 691–700, Avril 2002. 42, 45, 46, 56, 57, 59, 60, 61, 62

[Fridrich 2002b] Jessica Jiri Fridrich, Miroslav Goljab et Nasir Memon. *Cryptanalysis of the Yeung-Mintzer fragile watermarking technique*. Journal of Electronic Imaging, vol. 11, page 262, 2002. 39

[Fridrich 2006] Jessica Jiri Fridrich et David Soukal. *Matrix embedding for large payloads*. IEEE Transactions on Information Forensics and Security, vol. 1, no. 3, pages 390–395, 2006. 41

[Furon 2002] Teddy Furon. *Use of watermarking techniques for copy protection*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 2002. 37

[Furon 2005] Teddy Furon. *A survey of watermarking security*. Springer, vol. 3710/2005, pages 201–215, 2005. 39, 43, 46

[Group 2004] Safety Regulation Group. *GPS Integrity and Potential Impact on Aviation Safety*. Civial Aviation Authority, 2004. 120

[Group 2009] WPKI Project Group. *WPKI Main specification*. Rapport technique, WPKI Project Group, 2009. 150

[Guillemot 2004] Ludovic Guillemot. *Une approche vectorielle pour exploiter le contenu de l'image en compression et tatouage*. PhD thesis, Université Henri Poincaré, Nancy I, 2004. 30, 31

[Haber 1991] Stuart Haber et W. Scott Stornetta. *How to time-stamp a digital document*. Advances in Cryptology-CRYPT0'90, pages 437–455, 1991. 49, 155

[Hamming 1950] Richard Wesley Hamming. *Error detecting and error correcting codes*. Bell Sys. Tech. Journal, vol. 29, 1950. 71

[Hartung 1999] Frank Hartung et Martin Kutter. *Multimedia watermarking techniques*. Proceedings of the IEEE, vol. 87, no. 7, pages 1079–1107, 1999. 42

[Hassan 2006] M. Hamad Hassan et S.A.M. Gilani. *Embedded Semi-Fragile Signature Based Scheme for Ownership Identification & Color Image Authentication with Recovery*. 2006. 38, 42, 55, 57, 61, 62

[Hassinen 2006] Marko Hassinen, Konstantin Hyppönen et Keijo Haataja. *An Open, PKI-Based Mobile Payment System*. In Günter Müller, editeur, Emerging Trends in Information and Communication Security, volume 3995 of *Lecture Notes in Computer Science*, pages 86–100. Springer Berlin / Heidelberg, 2006. 149

[Hegarty. 2008] Christopher J. Hegarty. et Éric Chatre. *Evolution of the Global Navigation SatelliteSystem (GNSS)*. Proceedings of the IEEE, 2008. 109

[Hofmann-Wellenhof 1993] Bernhard Hofmann-Wellenhof, Herbert Lichtenegger et James Collins. Global positioning system. theory and practice., volume 1. 1993. 107

[Holliman 2000] M. Holliman et N. Memon. *Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes*. Image Processing, IEEE Transactions on, vol. 9, no. 3, pages 432–441, 2000. 88

[Hwang 2006a] Jin-Bum Hwang et Jong-Wook Han. *A Security Model for Home Networks with Authority Delegation*. In Marina Gavrilova, Osvaldo Gervasi, Vipin Kumar, C. Tan, David Taniar, Antonio LaganÃi, Youngsong Mun et Hyunseung Choo, editeurs, Computational Science and Its Applications - ICCSA 2006, volume 3983 of *Lecture Notes in Computer Science*, pages 360–369. Springer Berlin / Heidelberg, 2006. 150

[Hwang 2006b] Jin-Bum Hwang, Hyung-Kyu Lee et Jong-Wook Han. *Efficient and User Friendly Inter-domain Device Authentication/Access Control for Home Networks*. In Edwin Sha, Sung-Kook Han, Cheng-Zhong Xu, Moon-Hae Kim, Laurence Yang et Bin Xiao, editeurs, Embedded and Ubiquitous Computing, volume 4096 of *Lecture Notes in Computer Science*, pages 131–140. Springer Berlin / Heidelberg, 2006. 150

[ISO 2004] ISO et IEC. *Information technology - Digital compression and coding of continuous-tone still images requirements and guidelines*. ISO 10918-1, International Organization for Standardization / International Electrotechnical Commission, Geneva, Switzerland, 2004. 32

[ISO 2005] ISO et IEC. *Information technology – Security techniques – Information security management systems – Requirements*. ISO 27001 :2005, Geneva, Switzerland, 2005. 9, 21, 22

[Kalker 2001] Ton Kalker. *Considerations on watermarking security*. In Proceedings of the IEEE Workshop on Multimedia Signal Processing, pages 201–206, 2001. 43

[Kaplan 1995] Elliott D. Kaplan et Christopher J. Hegarty. *Understanding GPS*. Artech house, 1995. 107

[Kerckhoffs 1883] Auguste Kerckhoffs. *La cryptographie militaire*. Journal des sciences militaires, vol. IX, pages 5–83, January 1883. 44, 51

[Kundur 1999] Deepa Kundur et Dimitrios Hatzinakos. *Digital watermarking for telltale tamper proofing and authentication*. Proceedings of IEEE, vol. 87, no. 7, pages 1167–1180, Juillet 1999. 38, 39, 40, 41, 57, 58, 62

[Kutter 2000] Martin Kutter, Sviatoslav Voloshynovskiy et Alexander Herrigel. *Watermark copy attack*. In Proceedings of SPIE, volume 3971, page 371, 2000. 44, 45

[Lacresse 2004] H. Lacresse. *Détection de pannes en présence de paramètres de nuisance non-linéaires bornés*. PhD thesis, 2004. 162

[Lin 2000a] Ching-Yung Lin et Shih-Fu Chang. *Semi-fragile watermarking for authenticating JPEG Visual Content*. SPIE int. soc. opt. eng., 2000. 39, 40, 54, 57, 58, 62

[Lin 2000b] Eugene T. Lin, Christine I. Podilchuk et Edward J. Delp. *Detection of image alterations using semi fragile watermarks*. In PROC SPIE INT SOC OPT ENG, volume 3971, pages 152–163. Citeseer, 2000. 39, 40, 42, 55, 58, 61, 62

[Lorenzo 2009] David De Lorenzo, Sherman Lo et Per Enge. *A secure civil GNSS : Satellite signal authentication and location & time verification using hidden signatures.* SCPNT Symposium, 2009. 104

[Lu 2001] Chun-Shien Lu et Hong-Yuan Mark Liao. *Multipurpose watermarking for image authentication and protection*. IEEE Transactions on Image Processing, vol. 10, no. 10, pages 1579–1592, 2001. 38, 61

[Martineau 2008] A. Martineau. *Etude de la Performance du Contrôle Autonome d'Intégrité pour les Approches à Guidage Vertical*. 2008. 162

[Miller 2000] Matt L. Miller, Ingemar J. Cox et Jeffrey A. Bloom. *Informed embedding : exploiting image and detector information during watermark insertion*. In IEEE International Conference on Image Processing, volume 3, pages 1–4. Citeseer, 2000. 37

[Mills 1992] David L. Mills. *RFC1305 : Network time protocol (version 3) specification*. Rapport technique, IETF : Network Working Group, 1992. 154

[Mintzer 1999] Frederick Cole Mintzer et Minerva Ming-Yee Yeung. *Invisible image watermark for image verification*. no. 5875249, February 1999. 39

[Moon 2009] Jong Sik Moon, Deok Gyu Lee et Im-Yeong Lee. *Device Authentication/Authorization Protocol for Home Network in Next Generation Security*. LNCS, 2009. 150

[Neuman 2005] Clifford Neuman, University of Southern California Information Science Institute, Tom Yu, Sam Hartman, Kenneth Raeburn et Massachusetts Institute of Technology. *RFC4120 : The Kerberos Network Authentication Service (V5)*. Rapport technique, IETF : Network Working Group, 2005. 150

[NIST 1995] NIST. *An Introduction to Computer Security : The NIST Handbook*. National Institute of Standards and Technology, Octobre 1995. 9

[Papadimitratos 2009] Panagiotis Papadimitratos et Aleksandar Jovanovic. *GNSS-based Positioning : Attacks and Countermeasures*. MILCOM, 2009. 119, 121, 123

[Petitcolas 1998] Fabien A. P. Petitcolas, Ross J. Anderson et Markus G. Kuhn. *Attacks on copyright marking systems*. In Information Hiding, pages 218–238. Springer, 1998. 31

[Petitcolas. 1999] Fabien A. P. Petitcolas. et Ross J. Anderson. *Evaluation of copyright marking systems*. In Proceedings of IEEE Multimedia Systems, volume 99, pages 574–579. Citeseer, 1999. 41

[Pinkas 2003] Denis Pinkas, Bull, Nick Pope, John Ross et Security & Standards. *RFC3628 : Policy Requirements for Time-Stamping Authorities (TSAs)*. Rapport technique, Network Working Group, 2003. 49, 156

[Rey 2000] Christian Rey et Jean-Luc Dugelay. *Blind detection of malicious alterations on still images using robust watermarks*. In In IEE Seminar : Secure Images and Image Authentication, pages 7–1, 2000. 39, 40

[Rey 2002] Christian Rey et Jean-Luc Dugelay. *A survey of watermarking algorithms for image authentication*. EURASIP Journal on Applied Signal Processing, pages 613–621, 2002. 40, 45, 46, 54

[Schneier 2003] Bruce Schneier. Beyond fear : thinking sensibly about security in an uncertain world. Springer, 2003. 8

[Shaw 2000] Michael Shaw, Kanwaljit Sandhoo et David Turner. *Modernization of the global positioning system*. Rapport technique, DTIC Document, 2000. 109

[Srivastava 2003] A. Srivastava, A.B. Lee, E.P. Simoncelli et S.C. Zhu. *On advances in statistical modeling of natural images*. Journal of mathematical imaging and vision, vol. 18, no. 1, pages 17–33, 2003. 161

[Stadler 2011a] Yves Stadler, Yann Lanuel, Anass Nagih et Francine Herrmann. *Certification watermarking for digital handheld-captured photography*. In Pervasive and Embedded Computing and Communication Systems (PECCS) - Special session LOCSUE, 2011.

[Stadler 2011b] Yves Stadler, Yann Lanuel, Anass Nagih et Francine Herrmann. *Tatouage de certification pour des images numériques sur téléphone mobile*. In Workshop Interdisciplinaire sur la Sécurité Globale (WISG), 2011.

[Stadler 2012] Yves Stadler, Yann Lanuel, Anass Nagih et Francine Herrmann. *Contremesures aux attaques des systèmes de positionnement par satellites sur téléphones mobiles*. In Workshop Interdisciplinaire sur la Sécurité Globale (WISG), 2012.

[UCL 2002] UCL, Philips et MST. *Deliverable D 5.5*. In Certimark, volume IST - 1999 - 10987, 2002. 43

[Walton 1995] Steve Walton. *Information authentication for a slippery new age*. Dr. Dobbs Journal, vol. 20, no. 4, pages 18–26, 1995. 44

[Wang 2004] Z. Wang, A.C. Bovik, H.R. Sheikh et E.P. Simoncelli. *Image quality assessment : From error visibility to structural similarity*. Image Processing, IEEE Transactions on, vol. 13, no. 4, pages 600–612, 2004. 160

[Warner 2003] Jon S. Warner et Roger G. Johnston. *GPS spoofing countermeasures*. Homeland Security Journal, 2003. 121, 123

[Wen 2004] Hengqing Wen, Peter Yih-Ru Huang Huang, John Dyer, Andy Archinal et John Fagan. *Countermeasures for GPS signal spoofing*. Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2005), 2004. 121, 123

[Weng 2007] Li Weng et Bart Preneel. *On encryption and authentication of the DC DCT coefficient*. In International Conference on Signal Processing and Multimedia Applications, 2007. 56

[Westfeld 2001] Andreas Westfeld. *F5-a steganographic algorithm : High capacity despite better steganalysis*. Lecture Notes in Computer Science, vol. Volume 2137/2001, pages 289–302, 2001. 70, 71

[Wong 1998] Ping Wah Wong. *A public key watermark for image verification and authentication*. In Proceedings of the IEEE International Conference on Image Processing, volume 1, pages 455–459. Citeseer, 1998. 40

[Wong 2000] Ping Wah Wong et Nasir Memon. *Secret and public key authentication watermarking schemes that resist vector quantization attack*. 2000. 60

[Özgür Ekici 2004] Özgür Ekici, Bülent Sankur, Baris Coskun, Umut Naci et Mahmut Akcay. *Comparative evaluation of semifragile watermarking algorithms*. In Journal of Electronic Imaging, volume 13, pages 206–216, 2004. 39

[Zirari 2010] Soumaya Zirari, Philippe Canalda et François Spies. *WiFi GPS based combined positioning algorithm*. In Wireless Communications, Networking and Information Security (WCNIS), 2010 IEEE International Conference on, pages 684–688. IEEE, 2010. 123

Liste des acronymes

3G *Third Generation*. 1, 11

4G *Fourth Generation*. 1

AES Advanced Encryption Standard. 12

C/A *Coarse Acquisition*. 106

CA Autorité de certification - *Certification Authority*. 147, 148, 156

CDMA Code Division Multiple Access. 108

CRC Cyclic Redundancy Check. 12

DCT *Discrete Cosine Transform* - Transformée en Cosinus Discrète. i, 32–34, 38, 54, 55, 57, 59, 60, 65–68, 71–73, 89, 158

DES Data Encryption Standard. 12

DOP Dilution of Precision. 125, 131, 134

EDAS *EGNOS Data Service*. 111

EDGE Enhanced Data Rates for GSM Evolution. 19

FDMA Frequency Division Multiple Access. 108

GBAS Ground Based Augmentation Systems. 113, 124, 126, 142, 143

GISMO *GNSS Integrity and Signal Monitoring Observatory*. 125

GNSS Global Navigation Satellite System - Système de navigation globale par satellites. 101

GPRS General Packet Radio Service. 19

GPS *Global Positioning System*. ii, 3, 99, 101–103, 105–112, 114, 116–127, 129, 130, 134, 136, 137, 141, 153, 170

GSM *Global System for Mobile Communications*. ii, 3, 19, 121, 124, 126, 128, 130, 141–143, 146, 149

JPO Joint Program Office. 105

LSB *Least Significant Bit* - Bit de poids faible. i, 34, 50, 51, 55, 58–60, 66, 67, 69, 70, 97

MEO Medium Earth Orbit. 107

MSB *Most Significant Bit* - Bit de poids fort. 58, 67

NOTAM *Notices to Airmen*. 118

NSL *Nottingham Scientific Limited*. 25, 105, 108, 109, 125, 136, 138, 142, 169, 170

NTM *Notice to Mariners*. 118

OTP *One Time Password*. 149

P/Y *Precise code*. 106

PDCA Plan-Do-Check-Act. 21

PIN Personnal Identification Number. 19, 149, 150

PKI *Public Key Infrastructure*. 17, 43, 58, 64, 143, 146–148, 170

PRN *Pseudo Random Noise*. 106, 117

PSNR *Peak Signal-to-Noise Ratio*. 40, 80, 81, 83, 98, 171

PUK *PIN Unlock Key*. 150

RA Autorité d'enregistrement - *Registration Authority*. 147, 148

RFID Radio-frequency identification. 10

RIM Research In Motion. 11

RSA Rivest Shamir Adleman algorithm. 12

RSSI Responsable des Systèmes de Sécurité de l'Information. 22

RTK Real Time Kinematic. 106

SBAS Satellite-Based Augmentation System. 110–112, 129, 143

SHA Secure Hash Algorithm. 12

SIM *Subscriber Identification Module*. 148–150

SMSI Système de Management de la Sécurité de l'Information. 21

TSA *Time-Stamping Authority* - Autorité d'horodatage. 154

TSR *Time-Stamp Query* - Demande d'horodatage. 154

TSR *Time-Stamp Response* - Réponse (de demande) d'horodatage. 154

TTP *Trusted third party* - Tiers de confiance. 154

UMTS Universal Mobile Telecommunications System. 19

Wi-Fi *Wireless Fidelity*. ii, 1, 3, 11, 121, 124, 130, 141–143

Table des figures

Liste des tableaux

Résumé

Les smartphones équipent désormais toutes les couches de la population, indépendamment de l'âge ou de la profession. Ces équipements sont de plus en plus utilisés pour créer, manipuler et diffuser des informations contraintes par la sécurité (confidentialité, intégrité, authenticité). Plusieurs moyens de protection existent selon le type d'informations et les contraintes de sécurité (droit d'auteur, traçabilité, etc.).

Cette thèse propose une solution de tatouage spécifique à l'image capturée depuis un smartphone afin de la rendre valable en justice. Son objectif est de lier la preuve à l'image et d'empêcher toute modification du contenu de l'image ainsi que des éléments de preuve. La confidentialité de l'image n'étant pas recherchée, le tatouage présente l'intérêt de conserver la lisibilité de l'image probante et donc autorise sa diffusion. La preuve est aussi constituée des données contextuelles de l'image : son auteur, la date de prise de vue ou encore sa position géographique. Peu de moyens sont disponibles pour vérifier cette géolocalisation. La seconde partie de la thèse propose ainsi des méthodes logicielles permettant la mise en œuvre de contrôles destinés à améliorer l'authenticité de la géolocalisation. Enfin, le dernier chapitre propose une analyse critique des travaux de sécurité concernant les autres besoins d'authentification nécessaire à la réalisation d'une bonne preuve.

Abstract

Smartphones are nowadays ubiquitous, they can be found in anybody's hands with no consideration of one's age or work. They are used to create, manipulate and broadcast security constrained pieces of information (in term of confidentiality, integrity or authenticity). Different protection types can be found, with regard to the security constraints (copyrights, document tracking, etc.).

This thesis presents a watermarking scheme tailored for smartphone-captured images, which scheme allows the use of the image as an evidence. The goal is linking the evidence to the image and forbidding any content modification. Confidentiality not being a constraint, the scheme keeps the image visible and allow its broadcast. Contextual pieces of data are part of the evidence: author name, date of shot and geographic location (geolocation). Few means are available to assess this geolocation. The second part of the thesis aims to fill this gap by proposing software based countermeasures to enhanced geolocation authenticity. Finally, the last chapter presents a critic of security works on the other authentication methods required to forge a strong proof.