Rapport pour l’ Etudes
D’Approfondissement

Présenté par Saint-Marcel Frédéric

Stéganographie vs Tatouage

Table des matières

L'information a toujours constitué une denrée prisée. Comment l'acquérir ? Comment la protéger ? Comment en vérifier la provenance et l'intégrité ? Avec l'évolution des technologies et des connaissances, les réponses à chacune de ces questions ont évolué. De nouvelles défenses ont contré de nouvelles attaques, qui s'opposaient elles-mêmes à d'anciennes défenses. La lutte entre l'épée et le bouclier est probablement loin d'être terminée.

Abstract: Information always constituted a data appraisal. How to acquire it? How to protect it? How to check the source and the integrity of it? With the evolution of technologies and knowledge, the answers to each one of these questions evolved. New defences countered new attacks, which were opposed themselves to old defences. The fight between the sword and the shield is probably far from being finished.

Mots-clefs : dissimulation d'informations, stéganographie, protection des droits d'auteur, tatouage.

Chapitre 1

Introduction

1.1 historique

Que signifie le terme information ? Il ne s'agit pas forcément d'une notion rigoureuse, mais plutôt de renseignements, de données qu'une personne ne souhaite pas divulguer.

Deux grandes tendances visent à protéger l'information :

1. le chiffrement dont l'objectif est de rendre l'information incompréhensible à une personne ne possédant pas un secret : une personne surveillant le canal de communication par lequel transite le message sait qu'un échange a lieu, mais est ainsi incapable d'en interpréter le contenu
2. la stéganographie qui cherche plutôt à dissimuler la présence même d'informations pertinentes au sein de plusieurs autres sans réel intérêt pour le

destinataire du message: le message secret est caché dans un support de manière à passer inaperçu lors de la communication.

Les premiers emplois de stéganographie sont relatés par Hérodote. Les cheveux d'un esclave de confiance sont tondus puis le message est tatoué sur son crâne chauve. Une fois que les cheveux ont repoussé, l'esclave est envoyé au destinataire qui lui rase de nouveau la tête afin de lire le message. Une autre approche consiste à graver le message sur une tablette de bois, ensuite recouverte de cire afin d'obtenir une tablette d'écriture normale. Plus tard, l'encre sympathique permet d'écrire de manière invisible sur du papier. Une exposition du papier de couverture devant une flamme révélait le message écrit avec cette encre.

1.2 L'information aujourd'hui

Aujourd'hui, qui contrôle l'information détient énormément de pouvoir. Dans ces situations où l'information représente un tel enjeu stratégique et économique, il est devenu nécessaire de mettre en oeuvre des outils adaptés aux nouvelles technologies: une protection renforcée de la vie privée et des droits d'auteur. Il est devenu extrêmement simple de reproduire parfaitement n'importe quel médium. Dans le cas des média numériques (son, image et vidéo), les recherches se dirigent vers une solution technique: insérer une marque dans le médium afin d'identifier l' ayant-droit légitime.

1.3 Notions de cryptographie

Tout au long de ce document, nous faisons référence à des notions issues de la cryptographie. Dans cette partie, nous rappelons brièvement les domaines principaux de cette discipline, ainsi que leurs objectifs. La cryptographie est peuplée de personnages qui servent à présenter les protocoles:

_Alice et Bob sont les utilisateurs normaux du protocole. En général, Alice utilise la partie publique et Bob la partie privée ;
_Ève est une attaquante passive qui se contente d'écouter sur le canal de communication employé par Alice et Bob ;
_Charlie est un attaquant actif qui non content d'écouter la communication, cherche aussi à intervenir directement sur le contenu du canal.

On distingue deux catégories d'algorithmes de chiffrement : celle à clé secrète et celle à clé publique. Les algorithmes de chiffrement à clé secrète (ou symétriques) nécessitent le partage d'un secret, la clé, pour chiffrer et déchiffrer

un message. L'emploi d'un tel algorithme lors d'une communication demande alors l'échange préalable de cette clé entre les deux protagonistes. Un paramètre essentiel pour la sécurité d'un tel système est la taille de l'espace des clés. En effet, il est toujours possible de mener une attaque dite exhaustive pour retrouver la clé. Cette attaque consiste simplement à essayer toutes les clés possibles du système.

La cryptographie à clé publique (ou asymétrique) évite le partage d'un secret entre les deux interlocuteurs. Avec ce système de chiffrement, chaque utilisateur dispose d'un couple de clés. La clé publique est mise à la disposition de tous, dans un annuaire par exemple. L'utilisateur conserve soigneusement la clé secrète pour lui seul. Pour envoyer un message à Bob, Alice le chiffre à l'aide de la clé publique de Bob. Seul ce dernier est en mesure de déchiffrer le message reçu, grâce à sa clé secrète.

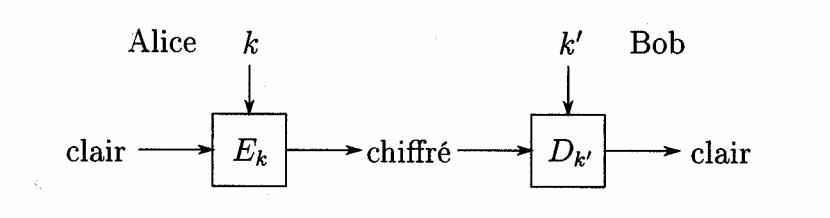

FIGURE 1.1 : Schéma de chiffrement

Chapitre 2

Dissimulation d'information

Le terme *dissimulation d'information* est très général ; il désigne simplement le fait de cacher une information dans un support. Il s'agit d'une libre adaptation de l'expression Anglaise *information hiding* couramment utilisée dans la littérature. Cependant, selon les objectifs, et les contraintes qui en découlent, on distingue différentes variantes.

2.1 Problématiques en dissimulation d'information

2.1.1 Terminologie

Dans la suite de ce mémoire, nous conserverons toujours la terminologie définie ici, afin de distinguer les différents éléments qui interviennent en dissimulation d'information. Tout d'abord, le médium vierge dans lequel des informations sont cachées est le *médium de couverture*, ou plus prosaïquement

le *médium*. Au contraire, une fois que les informations sont insérées, nous utilisons alors l'expression stégo-médium. D'une manière générale, nous appelons *données* l'information dissimulée dans le stégo-médium.

Le processus complet de dissimulation d'information repose sur deux opérations :

1. la dissimulation, qui consiste à insérer l'information dans le médium;
2. l'extraction, qui récupère cette information. Le mot détection est également utilisé lorsqu'il s'agit de vérifier la présence d'une information dans le stégo-médium, sans pour autant vouloir l'extraire.

2.1.2 Objectifs

La stéganographie cherche à cacher un message secret dans un médium de sorte que personne ne puisse distinguer un médium vierge d'un stégo-médium. La nature de l'information dissimulée ne revêt pas d'importance : il peut tout aussi bien s'agir d'un texte en clair que de sa version chiffrée. Ce message n'a a priori aucun lien avec le stégo-médium qui le transporte.

Lorsqu'un attaquant tente uniquement de détecter si un message transite dans un médium sur le canal de communication, on dit de lui qu'il est passif (Eve). La plupart des solutions de stéganographie ne considèrent que ce type d'attaquant, au contraire des deux domaines suivants où il est actif (Charlie): l'attaquant sait alors que le stégo-médium contient une information et il tente de la modifier ou de la retirer.

L'application suivante est parfois considérée comme descendant de la stéganographie. Cependant, la contrainte d'imperceptibilité y est bien moins forte. C'est pourquoi nous préférons utiliser dans ce document la terminologie courante qui consiste à placer ces deux domaines au même niveau, comme des cas particuliers de dissimulation d'information.

Le tatouage cherche à répondre au problème de la protection des droits d'auteur. Il tente de fournir une solution pour prouver qu'une entité est bien le véritable propriétaire d'un médium. Il s'agit bien de dissimulation d'information puisque, pour y parvenir, on insère un tatouage (ou marque, ou filigrane) dans le médium spécifique au propriétaire. Comme celui-ci souhaite protéger son médium et non une version trop déformée, l'insertion doit minimiser les modifications subies par le médium afin d'être imperceptible. Ensuite, chaque copie du stégo-médium contient la même marque, celle du propriétaire légal. Ici, la dissimulation ne signifie pas la même chose qu'en stéganographie : un attaquant sait qu'un tatouage est présent dans le stégo-médium, mais cette connaissance ne doit cependant pas lui permettre de le retirer.

2.1.3 Schéma général

Malgré leurs objectifs distincts, ces deux variantes n'en requièrent pas moins des paramètres communs :

_chaque approche nécessite des données, que ce soit un message, un tatouage ou une empreinte ;
_ces données sont dissimulées dans un support, le médium, qui possède plus ou moins d'importance selon le schéma: aucune pour la stéganographie, capitale pour les deux autres ;
_il est indispensable de pouvoir distinguer des personnes différentes, utilisant des données identiques dans un même médium: chacune doit donc posséder sa propre *stégo-clé* (ou plus simplement clé) afin que l'insertion de ces données identiques permettent quand même de différencier les protagonistes.

	Données	**Médium**	**Clé**
Stéganographie	le message à transmettre	sans importance	utilisée pour insérer/récupérer le message
Tatouage	une marque dépendant du médium et/ou du propriétaire	le médium dont on veut protéger les droits	utilisée pour insérer/détecter la marque dans le stégo-médium et, éventuellement, chiffrer le message
Fingerprinting	une empreinte dépendant du médium **et** de son utilisateur	le médium dont on souhaite prévenir la diffusion de copie illégale	utilisée pour insérer/détecter l'empreinte dans le stégo-médium

TABLEAU 1.1 : Utilisation des données, média et clés en dissimulation d'information

Le tableau montre que le rôle de la clé ne change pas dans ces applications: une clé est nécessaire pour insérer les données dans le médium, et les extraire (ou en détecter la présence) dans le stégo-médium.

Dans le cas du tatouage, les données sont construites, entre autre, à l'aide d'un générateur aléatoire. Celui-ci est initialisé avec une graine, parfois à tort appelée clé, mais qui n'a a priori aucun lien avec celle évoquée précédemment, même si elles se confondent dans certaines méthodes.

Le comportement de la procédure d'extraction/détection dépend des besoins de l'application. Signalons également que la réponse fournie change: alors que la récupération du message est le but même de la stéganographie, une mesure de confiance sur la présence ou non de données dans le stégo-médium peut suffire en tatouage.

2.1.4 Conditions requises

Nous avons vu que les objectifs de la dissimulation d'information peuvent changer de manière subtile. Classiquement, les applications sont triées en fonction de trois critères :

1. *l'imperceptibilité* : les données ne doivent pas être «perceptibles» dans le stégo-médium. Pour le tatouage, l'objectif est de ne pas détériorer le stégo-médium protégé. Cependant, la contrainte est plus forte en stéganographie où il s'agit plutôt d'une indétectabilité statistique afin qu'une personne surveillant le canal ne remarque pas la présence du message;
2. *la capacité* est la quantité de bits significatifs dissimulés dans le stégo-médium par unité d'accès (par exemple, le nombre de bits par seconde en musique) ;
3. *la robustesse* correspond à l'aptitude de préservation des données cachées face aux modifications du stégo-médium.

2.2 Stéganographie

2.2.1 Applications

Il est possible de présenter les techniques de stéganographie selon différentes classifications. On distingue alors six catégories :

_un système par substitutions remplace les parties redondantes du support par le message ;
_les techniques par transformations dissimulent l'information dans une transformée du support, comme par exemple, le domaine des ondelettes ;
_les techniques par étalement de spectre reposent sur le schéma du même nom ;
_les méthodes statistiques modifient plusieurs statistiques du support (fréquences des lettres, distribution des pixels. ..) pour cacher le message, et le récupèrent en testant ces hypothèses ;
_les techniques par distorsions altèrent le support, la différence avec le support initial constituant alors le message ;
_les méthodes par génération de support construisent un support autour du message pour le dissimuler.

2.2.2 Classification des schémas de stéganographie

Le contexte dans lequel se situe un schéma de stéganographie permet de le classer dans une des catégories suivantes :

_stéganographie pure: aucune entente préalable, autre que le choix de l'algorithme, n'est nécessaire, Alice et Bob utilisent le canal pour échanger des informations ;

_stéganographie à clé secrète: Alice et Bob conviennent au préalable d'une clé qui leur sert à insérer puis extraire le message du stégo-médium ;

_stéganographie à clé publique: tout comme en cryptographie, Alice utilise la clé publique de Bob lorsqu'elle souhaite lui envoyer un message. Bob, pour sa part, l'extrait à l'aide de sa clé privée.

2.3 Tatouage

La multiplication des données numériques, et plus encore la facilité avec laquelle il est possible de les reproduire, pose la question de la protection des droits d'auteur.

Deux approches sont possibles pour tenter de résoudre ce problème :

_brider l'outil qui effectue la copie ou la lecture du médium pour empêcher la reproduction: c'est le cas par exemple des lecteurs DVD qui nécessitent une clé pour déchiffrer le disque ;

_protéger le médium lui-même en y insérant des données qui identifient son propriétaire.

La dissimulation d'information correspond exactement à cette seconde catégorie. Par la suite, cette idée a évolué pour proposer des applications variées. Dans cette partie, nous présentons d'abord quelques applications du tatouage, avant de présenter une classification des solutions actuelles.

2.3.1 Applications du tatouage

2.3.1.1 Indexation

Dans ce contexte, l'information dissimulée facilite la recherche de documents multi-média. L'indexation se décompose en deux processus :

_extraction des vecteurs caractéristiques (ou feature vectors) du médium ;

_recherche des média similaires à un médium servant de requête.

Il existe un grand nombre d'attributs représentatifs d'un médium. Par exemple, pour une image, on peut citer la décomposition dans différentes bases (DCT, Fourier, ondelettes ...) , l'histogramme des couleurs, de l'orientation des arêtes. ..

Le but est de parvenir à décrire suffisamment une image pour que la recherche soit pertinente. Le tatouage sert alors à ajouter une information supplémentaire, comme un mot clé par exemple, qui facilite la recherche. Cette

application du tatouage est celle qui demande le moins de robustesse. Cette application ressemble donc plutôt à de la stéganographie, telle que nous l'avons décrite dans la partie précédente. Cependant, il existe une différence majeure qui justifie son appartenance au domaine du tatouage: le médium et les données ont un rapport.

2.3.1.2 Tatouage faible

Dans cette application, le but est de détecter les modifications subies par le médium. On distingue deux approches pour résoudre ce problème :

_le tatouage fragile: si le médium a subi plus de changements qu'autorisés, la détection de la marque échoue ;
_le tatouage fragile évolué: si le médium a été altéré, la marque indique l'endroit où se sont opérées les modifications.

La différence majeure entre les deux utilisations provient des besoins en capacité. En effet, la seconde approche nécessite une capacité plus élevée puisqu'il faut parvenir à décrire le médium dans la marque qui est dissimulée.

2.3.1.3 Protection des droits d'auteur

Il s'agit de la première application pour laquelle la dissimulation d'information est utilisée. Dans la suite de ce document, nous nous concentrons essentiellement sur cet aspect lorsque nous traitons de tatouage.

2.3.2 Classification des méthodes de tatouage

Nous introduisons tout d'abord les différents types de schémas définis selon les paramètres dont ils ont besoin lors de l'extraction. Nous présentons ensuite une classification des algorithmes de tatouage.

2.3.2.1 Types de schémas

Le propriétaire d'un médium possède une clé qui lui est propre, et qui l'associe à ce médium lors de l'insertion de la marque. On distingue deux types de fonction de détection D :

_type I: la fonction D extrait la marque elle-même ;
_type II : la fonction D vérifie uniquement la présence de la marque à l'aide d'une mesure de confiance.

D'autres caractéristiques hormis celles liées à l'extracteur entrent en jeu. C'est ainsi qu'on distingue les algorithmes de tatouage suivants :

_le tatouage privé, où le médium initial, la marque à tester et la clé sont donnés à l'extracteur. Dans ce type de tatouage, on compare l'original au stégo-médium récupéré pour extraire la marque ;
_le tatouage semi-aveugle utilise une fonction de détection qui nécessite la marque à tester et la clé ;
_le tatouage aveugle, où l'extracteur n'a pas connaissance ni du médium original, ni de la marque. Seule la clé secrète lui est nécessaire pour détecter ou extraire la marque du stégo-médium ;
_le tatouage asymétrique, où l'extraction de la marque ne nécessite pas la connaissance d'un secret. Cela implique que tout le monde est capable de lire la ou les marques du stégo-médium sans pouvoir les effacer. Cela pourrait se faire par un tatouage sans clé ou alors par un tatouage avec clé secrète et une extraction avec la clé publique correspondante (dans un schéma analogue à celui de la cryptographie asymétrique).

2.3.2.2 Choix du domaine

Traditionnellement, on distingue les schémas de tatouage selon le domaine sur lequel ils agissent. On dispose alors essentiellement de méthodes spatiales, fréquentielles ou multi-résolutions. Néanmoins, un autre critère de séparation existe en fonction de la manière dont est inséré le tatouage: soit il est ,ajouté au médium, soit il en remplace certains coefficients.Chaque espace de travail utilisé en tatouage possède ses propres avantages et inconvénients.

Les méthodes agissant dans le domaine spatial modifient directement les valeurs des pixels. Comme aucun traitement initial n'est requis, ces algorithmes sont très rapides et permettent de travailler en temps réel. De plus, elles offrent souvent une bonne résistance aux opérations géométriques.

Cependant, un tel schéma n'offre qu'une protection minime contre une compression: il suffit par exemple de compresser l'image en JPEG avec un faible taux de compression pour détruire la marque.

L'intérêt des schémas de tatouage s'est ensuite naturellement porté vers les domaines fréquentiels (DCT - Discrete Cosine Transform -et DFT -Discrete Fourier Transform). Ces espaces de travail sont fréquemment utilisés, par exemple dans les normes JPEG et MPEG2. Les schémas agissant dans ces domaines gagnent alors une certaine robustesse contre la compression.

Les algorithmes fonctionnant avec la DCT ne sont pas très résistants aux transformations géométriques comme les translations ou les rotations car celles-ci affectent grandement les coefficients de la DCT. Au contraire, l'espace de Fourier possède des propriétés d'invariance aux translations et rotations qui augmentent la fiabilité de la méthode.

Enfin, on voit maintenant apparaître de plus en plus de méthodes dans le domaine multi-résolutions. Cette évolution semble naturelle dans la mesure où elle suit celle des standards récents comme JPEG2000 et MPEG4.

Mais au-delà du domaine dans lesquels ils agissent, les schémas de tatouage se distinguent essentiellement par leur manière d'insérer la marque dans le médium.

2.3.2.3 Schémas additifs

Lors de l'insertion, le signal représentant la marque est ajouté à certaines composantes du médium. Pour y parvenir, il s'agit d'adapter la marque au médium, afin que le signal qu'elle représente ne soit ni trop faible (risques de non détectabilité et problèmes de robustesse), ni trop fort (effacement du signal initial, et donc trop grande dégradation de celui-ci).

Le tableau présente le principe d'insertion par addition. La génération de la marque se fait généralement par étalement de spectre.

Insertion par addition de la marque au signal
1. extraire des coefficients du médium initial m
2. réordonner ces coefficients selon une permutation paramétrée par une clé k pour obtenir un vecteur $c_k(m)$
3. générer une marque w_k, dépendante ou non du médium initial
4. tatouer le médium en ajoutant la marque aux coefficients : $c_k(m_w) = c_k(m) + w_k$
5. réordonner les coefficients puis reconstruire le médium tatoué m_w

TABLEAU 1.2 : Insertion dans un schéma de tatouage par addition

2.3.2.4 Schémas substitutifs

La différence majeure entre les schémas additifs et substitutifs provient de la phase d'insertion. Dans ces derniers, au lieu d'ajouter un signal au médium, certaines composantes sont remplacées afin que le stégo-médium exhibe une propriété caractéristique. Lors de la phase de détection, si cette particularité est présente, le stégo-médium est considéré comme tatoué avec une marque donnée.

Insertion par substitution de la marque au signal

1. extraire des coefficients du médium initial m
2. réordonner ces coefficients selon une permutation paramétrée par une clé k pour obtenir un vecteur $c_k(m)$
3. générer une marque w_k, dépendante ou non du médium initial
4. tatouer le médium en remplaçant certains coefficients $c_k(m)$ par ceux nouvellement générés par la marque :
$$c_k(m_w) = (\text{condition} ? \; w_k : c_k(m))$$
5. réordonner les coefficients puis reconstruire le médium tatoué m_w

TABLEAU 1.3 : Insertion dans un schéma de tatouage par substitution

Conclusion

Dans ce document, nous avons présenté les objectifs de la dissimulation d'information. Selon les attentes, les contraintes d'imperceptibilité, de capacité et de robustesse varient. Cependant, comme ces besoins sont à l'encontre les uns des autres, un compromis est toujours nécessaire.

Par ailleurs, ce chapitre montre la diversité des approches proposées. La dissimulation d'information du moins dans sa version informatique, est une discipline récente. Elle connaît en effet un grand essor depuis la seconde moitié des années 1990.

Les recherches portent actuellement autant sur la mise au point d'algorithme d'insertion et de détection que de protocoles susceptibles de les encadrer afin d'en accroître la sécurité.

Bibliographie

- Proceeding of the IEEE Protection of Multimedia Content , 1999 vol 67
- http://www.securite.org/db/crypto/steganographie
  - Neil Johnson, http://www.jjtc.com/Steganography/
  - Fabien a. p. petitcolas http://www.cl.cam.ac.uk/~fapp2/steganography/
- http://tsi.enst.fr/~maitre/tatouage/
  - International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
  - International Conference on Image Processing (ICIP)