Chapitre 18

La Stéganographie

18.1 Définition

La stéganographie est l'art de "cacher" une information privée ou secrète dans un support, apparement anodin. Le support peut être de plusieurs types tels qu'un fichier texte, image, audio, ou vidéo, mais peut également être un système de fichier, ou un code source. La caractéristique principale est que le fichier doit sembler ne contenir aucune information sensible, i.e. ne renfermer aucune information secrète. Le support associé à son message porte le nom de "stégo-médium".

La stéganographie part du principe que la perception humaine n'est pas assez évoluée pour détecter les petites modifications introduites dans les images et sont destinées à renfermer un message. De plus, personne ne sait, à priori, qu'un fichier renferme un message stéganographié (en dehors des personnes visées par le message).

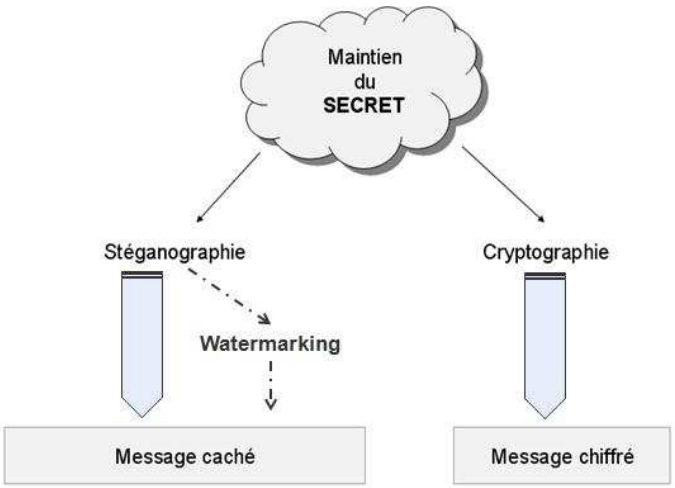

Fig. 18.1 – Stéganographie et concepts liés

On parle également de la stéganographie comme d'une possibilité pour établir un canal de communication secrète (*cover channel*). En effet, par son intermédiaire, il est possible de communiquer avec d'autres individus sans se faire remarquer. Il serait alors possible par exemple, de passer à travers un firewall en se bornant aux protocoles et liens autorisés, mais en transmettant des messages cachés à l'intérieur de ceux-ci.

Trois critères existent pour mesurer l'efficacité d'un système cryptographique :

– L'imperceptibilité : comme mentionné ci-haut, la stéganographie repose sur le fait qu'une tierce personne ne peut percevoir la présence d'un message.
– La capacité : il s'agit du nombre de bits qu'il est possible de dissimulés dans le support.
– La robustesse : elle représente la manière avec laquelle le message dissimulé résiste aux modifications apportées au support.

18.1.1 Cryptographie et Stéganographie

Tout comme la cryptographie, le but principal de la stéganographie est de protéger une information. La différence intervient dans la manière avec laquelle chaque technique procède. La cryptographie va chiffrer l'information pour qu'elle ne soit pas compréhensible par un tiers, alors que la stéganographie va la rendre invisible[1].

On utilisera dès lors cette dernière notamment lorsque les techniques cryptographiques sont inapplicables, ou lorsqu'on souhaite gardé secret le fait même de communiquer.

C'est en effet là le principal avantage de la stéganographie par rapport à la cryptographie : le fichier renfermant le message caché reste anodin. Personne ne peut se douter qu'il cache un message. Au contraire, un message chiffré annonce clairement qu'il renferme une information, puisque, par définition, il est chiffré.

Si on désire transmettre le message "*Jean arrêté à Paris. Mettre réseau en sommeil. Envoyer un OPR*[2]*.*", on obtiendra

– Par stéganographie (Stéganographie linguistique - Code de Barn) :

Mon cher Pierre,

J'espère que tu voudras bien m'excuser, mais j'ai eu tellement de travail à la maison que je n'ai pas pris le temps d'écrire aux amis. Cependant je t'envoie ce petit mot d'urgence pour te faire savoir que si tu veux des pneus, tu ferais bien de te dépêcher ; en effet :

Hier, Jean est venu nous rendre visite, il descendait du train et s'est arrêté un moment chez nous pour bavarder et donner des nouvelles à mon père de son Paris. En principe, il doit rester quelques jours ici pour mettre en ordre ses affaires avant de repartir pour la capitale. A Paris, c'est calme, mais la veille il avait été dérangé en plein sommeil par les sirènes deux fois dans la nuit ! Ceci mis à part, il doit nous faire envoyer par un ami à lui des pneus neufs pour nos vélos. Il en a pour le moment, profitons-en ! A bientôt de tes nouvelles.

P.-S. Nous irons au mariage de Simone et Henri, dimanche en quinze. Henri est un garçon sympathique qui a connu Simone l'an dernier chez Xavier, notre vieil ami. Il a deux ans de plus qu'elle et nous pensons que Simone va être très heureuse.

– Par chiffrement (Vigénère - Clé = Cryptographie) :

LVYCT FXKPP YQWOV RIKSX JEPBM RUFKB XWRVN KVGIT LLDIF

Bien que le texte chiffré soit beaucoup plus court, il est cependant beaucoup plus explicite quant au fait qu'il renferme un secret.

On utilisera parfois la cryptographie en supplément de la stéganographie. Ainsi, même si le message est découvert, il faudra encore le déchiffrer, ce qui apporte un gain de sécurité non négligeable.

[1] Stéganographie vient du grec "steganos", signifiant "couvrir".
[2] OPérateur Radio.

18.1.2 Watermarking et Stéganographie

Le watermarking consiste à cacher une information dans le but de protéger le support contre la copie, contre toute modification de ce support ou dans un but d'identification. On considère souvent le watermarking comme la forme moderne de la stéganographie, ou en tout cas comme la principale application de sécurité reposant sur la stéganographie.

On trouve aujourd'hui régulièrement ce type de sécurité dans les images protégées par droits d'auteurs, dans certains supports audios et vidéos, ou encore sur des cds ou dvds. La détection du message permettra alors de distinguer un original d'une copie, de valider un accès, ou encore de tracer l'utilisation d'un média.

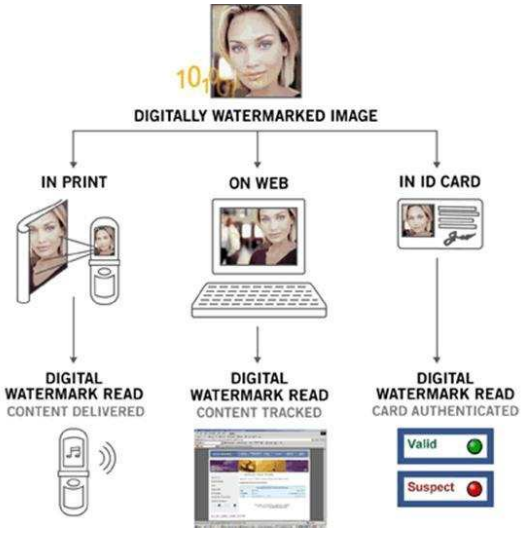

Fig. 18.2 – Quelques applications des techniques de watermarking
Source : http ://www.willamette.edu/wits/idc/mmcamp/watermarking.htm

Hormis l'utilisation qui en sera faite, une autre différence majeure existe entre les deux concepts :

1. Dans le cadre de la stéganographie, le message doit être caché, imperceptible.
2. Dans le cadre du watermarking, le message devra être impossible à enlever, impossible à modifier (mais peut éventuellement être visible).

Il est également fréquent de considérer la stéganographie comme une technique utilisée en deux personnes (one-to-one), le watermarking étant préféré pour les communications one-to-many.

18.2 La stéganographie dans l'Histoire

La première utilisation de la stéganographie remonte à la Grèce Antique où on utilisait certains esclaves pour transmettre les messages. Ceux-ci étaient écrits sur les crânes des messagers, et passaient donc inaperçu lorsque les cheveux repoussaient. Une fois suffisamment longs, le messager pouvait être envoyé, avec l'ordre de se faire raser le crâne une fois arrivé à destination. Le principal désavantage de cette méthode était l'attente pour l'envoi d'un message.

Une autre technique était d'utiliser des tablettes de cire. Une fois la cire raclée, on gravait le message dans le bois de la tablette. Il suffisait ensuite d'y remettre de la cire, et le message était parfaitement caché.

En Chine ancienne, les messages étaient écrits sur de la soie, qui était ensuite roulée en boule, elle-même recouverte de cire. Un messager devait enfin avaler cette boule.

Vers 1500, l'abbé Trithème utilisa des louanges afin de faire passer des messages de manière anodine. Chaque lettre du message était associée à un groupe de mots, comme indiqué à la figure 18.3. Sa technique se rapproche d'une technique cryptographique, mais le concept de "recouvrement du message" en fait un exemple de stéganographie linguistique.

A	dans les cieux	N	en paradis
B	à tout jamais	O	toujours
C	un monde sans fin	P	dans la divinité
D	en une infinité	Q	dans la déité
E	à perpétuité	R	dans la félicité
F	sempiternel	S	dans son règne
G	durable	T	dans son royaume
H	sans cesse	U, V, W	dans la béatitude
I, J	irrévocablement	X	dans la magnificence
K	éternellement	Y	au trône
L	dans la gloire	Z	en toute éternité
M	dans la lumière		

FIG. 18.3 – Les Ave Maria de l'abbé Trithème

Au XVI^e siècle, Giovanni Porta, un scientifique italien, découvrit le moyen de cacher un message dans un oeuf dur : en écrivant sur la coquille avec une encre contenant une once d'alun par pinte de vinaigre. L'encre pénètre alors la coquille et le message se voit écrit sur le blanc de l'oeuf sans apparaitre sur la coquille. Pour lire le message, il suffit d'éplucher l'oeuf.

Enée le Tacticien, un historien de la Grèce Antique, imagina de percer de minuscules trous sous certaines lettres d'un texte anodin. Une fois toutes les lettres ainsi marquées relévées, on pouvait lire le message caché. Ces trous étaient invisibles pour une personne n'étant pas au courant de la méthode employée.

Durant la seconde guerre mondiale, on utilisa l'encre invisible. Cette encre, pouvant être créée notamment à partir de lait, de vinaigre, d'urine, de jus de citron ou encore de chlorure d'ammoniac, permettait d'écrire sur du papier sans pour autant afficher les caractères ainsi notés.

Différentes méthodes linguistiques existent également. Pour cacher un message dans un texte, on peut jouer sur l'espace entre les mots, la ponctuation, ou encore l'orthographe. A l'origine, c'est l'acrostiche qui permettait de cacher ces messages. L'acrostiche est un poème dont la première lettre de chaque vers compose un mot ou une phrase. Dès qu'un enfant commence à écrire, il pratique donc souvent l'acrostiche pour son premier cadeau de fête des mère ! Un des exemples les plus imposant fut sans doute le livre *Hypnorotomachia Poliphili* publié en 1499 par un anonyme. On découvre en effet dans cet ouvrage que si on regroupe la première lettre de chacun des chapitres, au nombre de 38, on peut recomposer la phrase *Poliam frater Franciscus Columna peramavit* ce qui signifie *Frère Francesco Colonna aime Polia passionnément.*

Mais il existe également des techniques plus évoluées où, pour pouvoir comprendre le message caché, il faut savoir quels mots ou quelles lettres lire. Par exemple, ci-dessous, un texte à première vue innocent envoyé par un espion allemand pendant la seconde guerre mondiale renfermait un message très important :

Apparently neutral's protest is thoroughly discounted and ignored. Isman hard it. Blockade issue affects pretext for embargo on byproducts, ejecting suets and vegetable oils.

Dans la cas présent, en prenant la deuxième lettre de chaque mot, on voit apparaitre le message suivant :

Pershing sails from NY June 1.

Parmi les exemples les plus célèbres illustrant le principe de stéganographie linguistique (ou littéraire), on trouvera encore l'échange de courier entre George Sand et Alfred de Musset.

Cher ami, Je suis toute émue de vous dire que j'ai
bien compris l'autre jour que vous aviez
toujours une envie folle de me faire
danser. Je garde le souvenir de votre
baiser et je voudrais bien que ce soit
une preuve que je puisse être aimée
par vous. Je suis prête à montrer mon
affection toute désintéressée et sans cal-
cul, et si vous voulez me voir ainsi
vous dévoiler, sans artifice, mon âme
toute nue, daignez me faire visite,
nous causerons et en amis franchement
je vous prouverai que je suis la femme
sincère, capable de vous offrir l'affection
la plus profonde, comme la plus étroite
amitié, en un mot : la meilleure épouse
dont vous puissiez rêver. Puisque votre
âme est libre, pensez que l'abandon où je
vis est bien long, bien dur et souvent bien
insupportable. Mon chagrin est trop
gros. Accourez bien vite et venez me le
faire oublier. À vous je veux me sou-
mettre entièrement.

Votre poupée

Quand je mets à vos pieds un éternel hommage,
Voulez-vous qu'un instant je change de visage ?
Vous avez capturé les sentiments d'un coeur
Que pour vous adorer forma le créateur.
Je vous chéris, amour, et ma plume en délire
Couche sur le papier ce que je n'ose dire.
Avec soin de mes vers lisez les premiers mots,
Vous saurez quel remède apporter à mes maux.

Alfred de Musset

Cette insigne faveur que votre coeur réclame
Nuit à ma renommée et répugne à mon âme.

George Sand

Une autre méthode ingénieuse fut inventée par Gaspar Schott (1608 - 1666). Le principe, illustré à la figure 18.4 était de coder le message selon des notes de musique. L'avantage du procédé était que le message apparaissait comme une partition musicale, et donc passait totalement inaperçu. Cependant, si la partition était jouée, il y avait très peu de chance pour que la mélodie soit agréable.

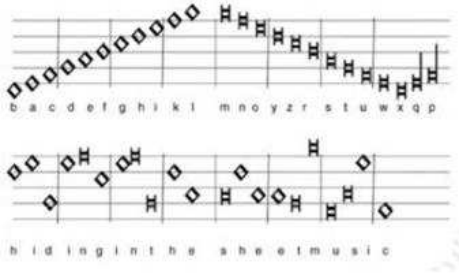

FIG. 18.4 – La stéganographie par partition musicale.

La ponctuation permit aussi d'identifier certains messages cachés. Au $XVII^e$ siècle, Sir John Trevanion fut arreté et emprisonné dans un chateau. Il reçut alors une lettre, que les gardiens avaient jugés sans danger. Lorsque John lut celle-ci, il détecta la présence suspecte de certaines virgules étrangement placées. Il repéra également qu'en prenant la troisième lettre de chaque mot suivant ces virgules, il pouvait former la phrase *Panel at east of chapel slides* ce qui signifiait *Le panneau à l'extrémité Est de la chapelle peut glisser.* C'est ainsi qu'il demanda un instant de recueillement dans la chapelle et s'évada.

Dans les années 1940, la puissance des techniques stéganographiques était si crainte que les Etats-Unis mirent en place un service de censure constitué d'environ 10.000 personnes, chargées d'étudier et de détecter les messages cachés. On interdit les parties d'échecs internationnales, les dessins d'enfant accompagnant les lettres pour les grands-parents, les diffusions de disques à la radio, ou encore les annonces pour chiens perdus.

La stéganographie reste également un sujet d'actualité. Après les incidents du 11 septembre 2001, beaucoup ont avancé l'hypothèse que des échanges entre terroristes avaient eu lieu afin de préparer les attaques.

Remarque : Un autre "phénomène stéganographique" est apparu avec le livre "The Bible Code" de Michael Drosdin. Ce dernier prétendait que la Bible renfermait des messages cachés, tels que l'assassinat d'Yitzhak Rabin. Après plusieurs expériences, il s'avéra que ce "code" n'était rien d'autre que le fruit du hasard. Ainsi, on découvrit notamment l'assasinat de JFK dans Moby Dick (fig. **??**), ou encore le décès accidentel de la princesse Diana.

L'explication prend sa source dans un théorème, portant le nom de Théorème de Borel. Celui-ci indique que si on prend un nombre suffisant de combinaisons de lettres, il est possible de retrouver à peu près n'importe quel sujet ayant trait au passé ou au futur. Ainsi, sur un texte d'environ 1000 lettres, il est possible de trouver plusieurs millions de mots, selon leur agencement par la méthode appelée ELS (Equidistant Letter Sequence).

18.3 Principes

Aujourd'hui, les messages cachés se transmettent de manière digitale et non plus par des techniques manuelles (bien que le système du message écrit sur le crâne d'individus semblent encore avoir été utilisé au début du XX^e siècle par des espions allemands). Le schéma du processus stéganographique est illustré à la figure 18.5.

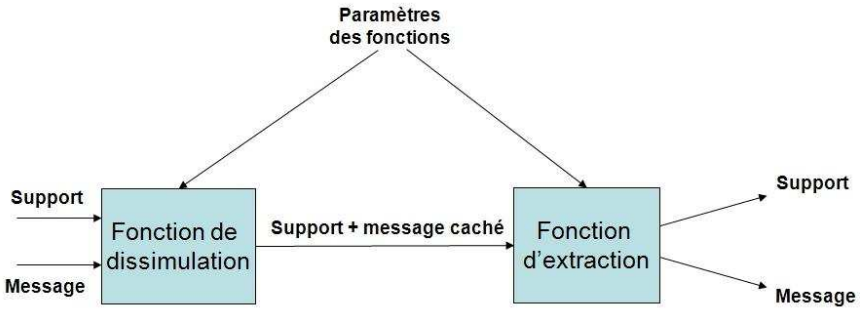

FIG. 18.5 – Principe d'un système stéganographique.

Il est à remarquer que plus le message à dissimuler est long, plus la modification du fichier support sera importante, et donc plus la détection sera facile.

On distingue 3 utilisations de la stéganographie :
– la stéganographie simple : seul l'algorithme à utiliser doit être connu des deux parties communicantes.
– la stéganographie à clé secrète : les deux parties utilisent une clé commune (ce sont les "paramètres" de la figure 18.5) afin d'introduire et d'extraire le message du support.
– la stéganographie à clé publique : il est possible d'utiliser un système à clé publique/privée (utilisable en paramètre), comme dans le cadre cryptographique.

18.3.1 Première Technique : la substitution

La substitution consiste à utiliser des zones inutilisées ou de faibles importances dans un fichier. Une méthode employée dans ce contexte porte le nom de LSB (Least Significant Bit).

Le LSB utilise comme son nom l'indique les bit de poids le plus faible pour transmettre un message. Si on regarde le codage d'une image RGB, on remarque que chaque pixel est codé selon 3 octets (un octet par couleur). En modifiant le bit de poids le plus faible, l'image sera effectivement modifiée, mais les nuances résultantes seront tellement faibles que l'oeil humain ne pourra que très rarement s'en rendre compte[3].

Le message caché pourra être d'autant plus important que l'on utilisera un grand nombre de bits pour le cacher. En contrepartie, utiliser plusieurs bits modifiera plus fortement l'image de base, et donc rendra la présence d'un message beaucoup plus détectable.
L'avantage du LSB est que la taille du fichier n'est pas modifiée, puisque le message est encodé dans les parties peu ou pas utilisées du fichier. C'est également une méthode rapide et facile à mettre en oeuvre. Cependant, le LSB possède un désavantage de taille, à savoir la perte du message lorsque des changements importants ont lieu sur le support, comme par exemple une rotation, ou un redimmensionnement de l'image.

Une amélioration du LSB consiste à introduire un paramètre aléatoire permettant de distribuer les bits de faible utilisés. Ainsi, les modifications n'auront pas lieu "uniquement" dans les premiers octets de l'image, mais seront au contraire répartis aléatoirement dans l'entièreté de l'image.

[3]Si la modification avait eu lieu sur le bit de poids le plus fort, les différences auraient été très marquées, et on aurait pu en déduire la présence d'un message caché.

18.3.2 Deuxième Technique : l'injection

L'injection repose sur le fait d'inclure le message directement dans le fichier support. L'inconvénient de cette approche est que la taille du fichier support est modifiée. La présence d'un message est alors plus facilement détectable, bien que ce problème ne se pose que lorsque des regards indiscrets possède une copie du fichier support original.

18.3.3 Troisième Technique : la création d'un nouveau fichier

Contrairement aux deux autres méthodes qui utilisent un fichier support comme base, c'est le message qui servira ici de base. A partir de celui-ci, on construira une "enveloppe" qui le renfermera. La lettre de George Sand à Alfred de Musset peut être vu comme un exemple de cette technique. En effet, pour transmettre son message à Alfred de Musset, elle a créé une lettre anodine *autour* de son message.

18.4 Les types de support

Les techniques de stéganographie peuvent s'utiliser sur pratiquement tous les types de support. Les plus communs sont explicités ci-dessous.

18.4.1 Les images

La technique du LSB a déjà été explicitée dans une section précédente. On trouvera principalement 2 types d'images : les images 24 bits et 8 bits. On remarquera cependant qu'il est plus facile de cacher un message dans une image 24 bits en raison du nombre plus important d'octets.

18.4.2 L'audio

Il existe 4 techniques principales permettant d'intégrer des données dans un flux audio :

- **LBE (Low Bit Encoding)** : les données sont stockées dans le flux audio d'une manière semblable à celle utilisée dans le cas du LSB pour les images.
- **SSE (Spread Spectrum Encoding)** : cette méthode ajoute un bruit aléatoire au signal sonore. Le message est ensuite dissimulé dans ce bruit additif.
- **EDH (Echo Data Hiding)** : dans la plupart des cas, un echo dans une mélodie permet de grandement améliorer la qualité sonore. Plus le temps séparant le son original et son echo est réduit, moins l'oreille humaine peut le détecter. La durée de ce délai peut alors être utilisée pour dissimuler des bits d'information. Ces délais une fois mesurés, il est alors possible de recréer le message initial.

18.4.3 La vidéo

Plusieurs techniques existent, mais sont pour la plupart des améliorations ou modifications de la technique de la transformée en cosinus discret (DCT). Celle-ci utilise la quantification des parties les moins importantes des images (arrondies aux valeurs supérieures par exemple). L'oeil étant relativement peu sensible aux hautes fréquences, la DCT pourra appliquer des modifications plus importantes dans cette plage, sans pour autant créer de modifications visibles dans l'image. Le message est alors injecté dans l'image en jouant sur les facteurs d'arrondi. Une compression sans perte est ensuite appliquée.

18.4.4 Autres supports

La plupart des fichiers existants peuvent être utilisés pour transmettre un message de manière cachée. Cela peut aller de la simple page HTML à des formes plus évoluées telles que l'insertion d'octets dans une portion inutilisée de code assembleur. La stéganographie ne se cantonne donc pas aux simples images