www.ncottin.net

Stéganographie Stéganalyse

Nathanaël COTTIN

Version 1.0.4
2011

Dissimuler des informations secrètes au sein de supports d'aspect anodin

Théorie et pratique

Préambule : le jeu des différences

www.ncottin.net

Image originale ?

Image modifiée ?

Plan de la présenation

www.ncottin.net

- Stéganographie
- Stéganalyse

Théorie

- Expérimentations

Pratique

www.ncottin.net

Partie I Stéganographie

« Ce qui n'a pas l'air d'être chiffré n'a aucune raison d'être [décrypté] »

– Didier Godart

Définition

- Du grec « graphie » et « stégano »
- Science de la dissimulation (masquage, camouflage) d'informations secrètes

- Identification des éléments :
  - Medium de couverture (vierge) = hôte textuel ou sémagramme (visuel ou son)
  - Information secrète
  - Stego-medium

Histoire de la stéganographie

- 600 avant J.-C. : messages sur les crânes rasés des esclaves
- Tablettes de cire des Grecs

- Autres techniques ancestrales :
  - Encres invisibles (jus de citron)
  - Modifications typographiques
  - Agencements de mots
  - Espacement des caractères au sein des mots

Histoire de la stéganographie

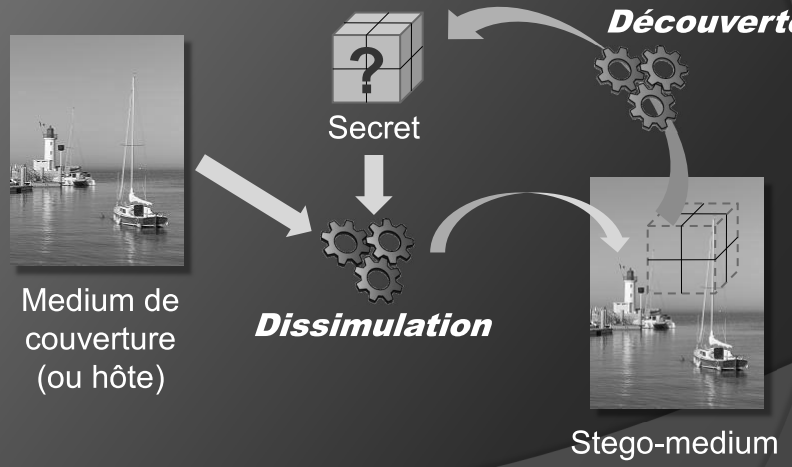
- Seconde guerre mondiale :
  - Allemands :
    - Microfilms cachés sous des timbres postes ou sur des couvertures de magazine
    - Micropoints dans les signes de ponctuation
  - Alliés : textes dont seuls certains caractères sont conservés

De nos jours : utilisations courantes

- Échanger en environnement surveillé
- Sauvegarder / transmettre de l'information sensible sans éveiller les soupçons
- Apporter une preuve de propriété
- Passer outre les restrictions d'usage de la cryptographie
- Compléter les techniques de chiffrement

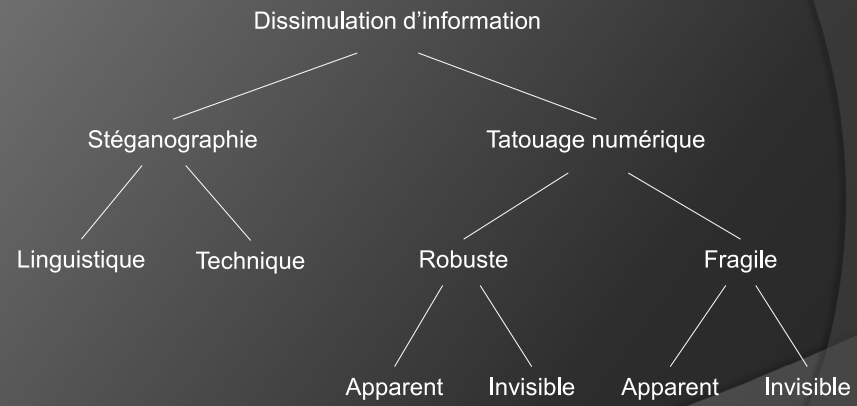
Principes de dissimulation et découverte

www.ncottin.net



Classification générale

www.ncottin.net



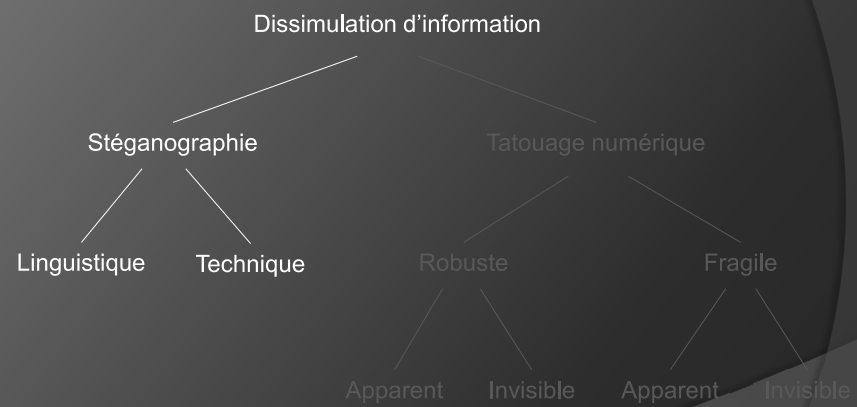
Différences entre stéganographie et tatouage

www.ncottin.net

- Stéganographie :
  - Information à protéger = secret
  - Hôte sans rapport avec le secret
  - Secret conçu pour demeurer invisible
- Tatouage numérique :
  - Information à protéger = hôte
  - Secret en rapport (direct, indirect) avec le medium de couverture
  - Secret conçu pour être révélé

Section 1 : stéganographie

www.ncottin.net



Systèmes de stéganographie

www.ncottin.net

- Stéganographie pure : repose sur le secret de la méthode employée…
- Stéganographie à clé secrète
- Stéganographie à clé publique

Types de dissimulation

www.ncottin.net

- Dissimulation passive
- Dissimulation adaptative
- Dissimulation active

Méthodes de dissimulation

www.ncottin.net

- Substitution
- Insertion
- Génération

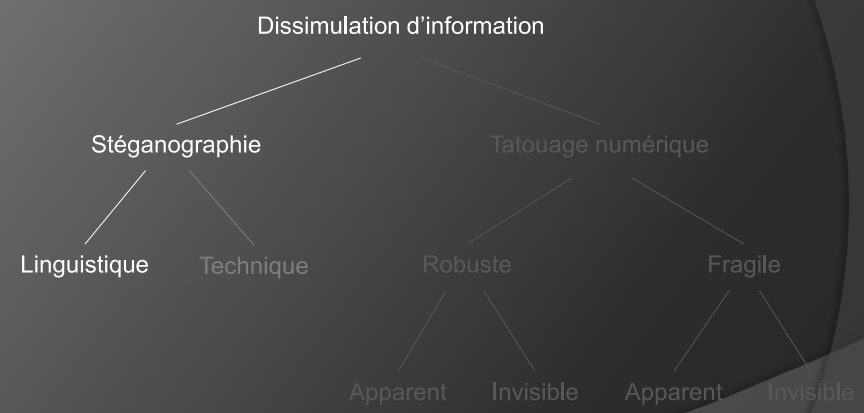
Techniques informatiques actuelles

www.ncottin.net

- Modifications binaires des hôtes
- Medium de couverture prisés :
  - Sémagrammes :
    - Images
    - Sons
    - Vidéos
  - Fichiers texte :
    - Libres
    - Formatés (XML, HTML)

Stéganographie linguistique

www.ncottin.net



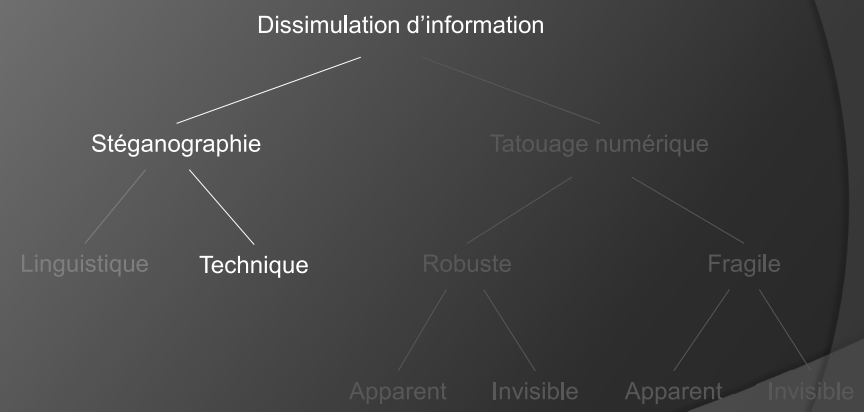
Exemples de stéganographie linguistique

www.ncottin.net

- Codes de Barn
- Modification du schéma grammatical des phrases

Stéganographie technique

www.ncottin.net



Dissimulation dans du texte libre

www.ncottin.net

- Modification des espaces / tabulations entre les mots
- Ajout d'espaces / tabulations en fin de ligne
- Alternances majuscules / minuscules

Dissimulation dans les LSBs d'une image RVB

- Dissimulation dans les bits de poids faible
- Un pixel 32 bits = 4 octets (T, R, V, B)
- Chaque composante de couleur est exprimée par une valeur entre 0 et 255
  ➔ $256^3 = 16777216$ couleurs différentes
- Modifications sensibles non décelées par l'œil humain
  ➔ Dissimulation dans les LSBs

Image servant d'hôte et de clé

- Opération « ou exclusif » entre les bits à masquer et les bits à modifier au sein de l'image hôte
- Besoin de l'image originale pour découvrir l'information dissimulée

➔ Image = hôte + clé de déchiffrement
➔ Extension à 2 images distinctes

Autres conteneurs appropriés

- Fichiers texte : espaces séparateurs de mots, espaces de fin de ligne
- Fichiers WAV : modification des fréquences inaudibles par l'homme (< 20 Hz ou > 20 kHz)
- Fichiers HTML et XML : espaces, commentaires
- Images compressées (JPEG)

Dissimulation dans une image JPEG

- Découpage en blocs carrés de côté 8 pixels
- Application d'une « transformation en cosinus discrète » (DCT) pour chaque couleur de chacun des pixels de chaque bloc
- Cette transformation permet de déterminer les pixels pouvant être modifiés sans impact visuel

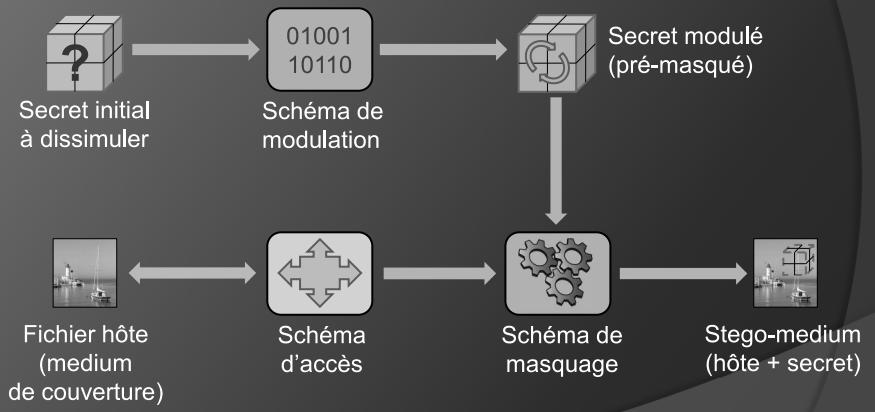
Formats hôtes les plus courants

- Formats textuels :
  - TXT, TEX, HTML, XML
- Sémagrammes :
  - BMP, TIF, PNG, GIF, PCX, PICT, JPEG
  - WAV, PCM, AVI, MIDI, MPEG, MP3, RIFF, VOC
- Mais aussi : codes source, protocoles, …

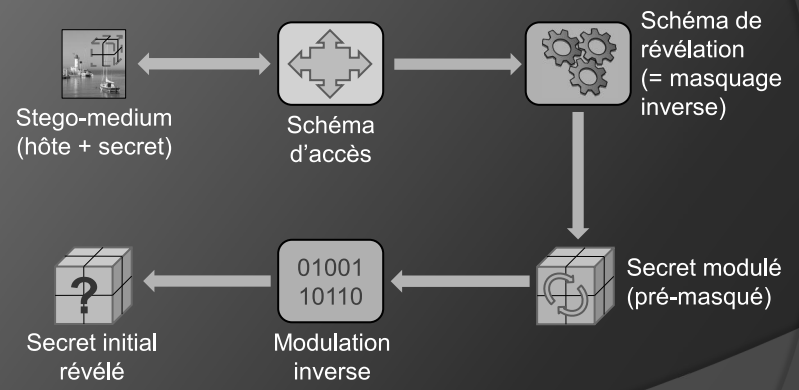
Procédés (schémas) communs

- Accès : méthode de sélection des portions de l'hôte pouvant être modifiées
- Masquage / révélation : méthode de modification des portions sélectionnées (substitution, insertion, génération)
- Validation (optionnel) : vérification que l'hôte est apte à recevoir le secret
- Modulation : modification du secret avant dissimulation

Principe général de dissimulation



Principe général de découverte (révélation)

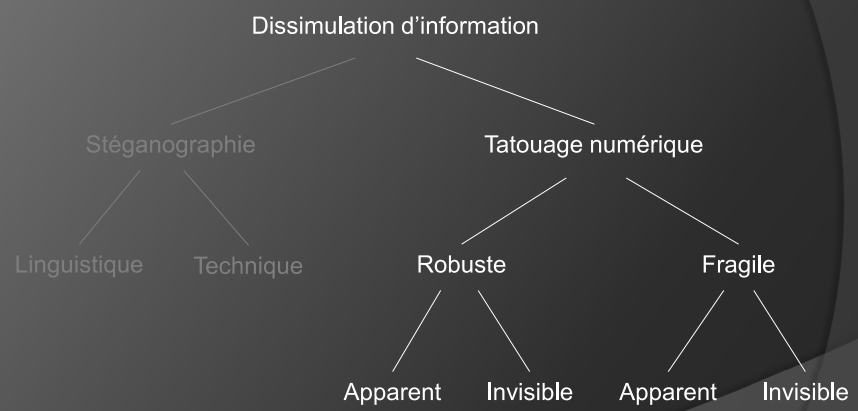


Utilisation d'un type de conteneur comme hôte

- Modifications non décelables par l'homme (visuel, auditif)
- Standards permissifs (HTML, XML, LaTeX, protocoles)
- Possibilité d'employer des macros (OOo, MS Office, …)

➔ Apprécier quels types de changements seront « invisibles » pour l'homme

Section 2 : tatouage numérique



Définition et objectifs du tatouage numérique

- Technique permettant d'ajouter des informations à un document numérique, notamment multimédia (image, audio / vidéo)
- Objectifs :
  - Assurer la propriété intellectuelle (droits d'auteur)
  - Lutter contre le piratage (copie) et la contrefaçon

Watermarking et fingerprinting

- Watermark = tatouage invisible et indélébile
  - ➔ Filigrane numérique
- Fingerprint = watermark personnalisé
  - ➔ Marque unique par exemplaire
  - ➔ Protection des droits d'auteur
  - ➔ ≠ DRMs (restrictions d'utilisation)

Propriétés d'un filigrane numérique

- Garantit sa détection
- Si robuste (non fragile), il doit résister :
  - Au changement de type de support
  - Aux transformations du médium de couverture (rotations, transformations affines, …)
  - À la compression (i.e. changement de format)
  - À sa suppression (numérisation, …)

  → Medium de couverture inutilisable

➔ Indélébile

Partie II
Stéganalyse

Ce medium apparemment inoffensif
renferme-t-il un secret ?
Si oui, comment en prendre connaissance ?

Objectifs de la stéganalyse

- Détecter qu'une information secrète est dissimulée
- Prendre connaissance de cette information (ou éventuellement la détruire)

Types d'attaques

- Passive : identifier la présence d'un secret
- Active : identifier puis détruire le secret

Classification des techniques de stéganalyse

- Stéganographie seule
- Medium (hôte) connu
- Information secrète connue
- Information secrète choisie
- Stéganographie connue
- Stéganographie choisie

Stéganalyse de dissimulation dans les LSBs

- La modification des LSBs d'une image produit des variations entre pixels voisins visibles sur l'histogramme des composantes de couleur de l'image

➔ Un hôte renfermant une information dans ses LSBs a un histogramme non uniforme

Conclusion technique

- Utilisation de 3 schémas :
  - Schéma de modulation du secret
  - Schéma d'accès (ou sélection)
  - Schéma de masquage
- Choix de l'hôte et de la technique de dissimulation les plus appropriés ?
- Rapport avec la compression de données ?

Conclusion générale

- Complément aux techniques cryptographiques
- Communauté active (procédés de masquage et stéganalyses)
- Détournement possible à des fins illégales…
- Absence de règlementation

Références bibliographiques (extrait)

- *An Overview of Steganography*, Shawn D. Dickman, 2007
- *Principles of Steganography*, Max Weiss
- *Stéganographie, Watermarking*, Dimitri Fossier, Julien Thévenon
- *Information noyée, information cachée*, Jean-Paul Delahaye, Pour la Science, n°229, Novembre 1996
- …

Partie III
Expérimentations

Utilisation de l'outil « Hide & Reveal » et de sa bibliothèque `org.steganography`

Disponibles sur « `www.hidereveal.org` »

Bibliothèque « `org.steganography` »

- Bibliothèque Java sous licence GPL
- Déclare divers hôtes :
  - Image non compressée (BMP, TIF, PNG)
  - Texte (en cours de développement)
- Utilise les schémas de dissimulation :
  - Modulation
  - Accès
  - Masquage

Bibliothèque « `org.steganography` »

- Dissimulation en 2 phases :
  - Taille du secret (en octets)
  - Contenu du secret (octet par octet)
- Intègre la validation automatique des schémas déterministes
- Évolutive (interfaces, classes abstraites)

➔ Conçue pour groupes de recherches

Quelques schémas de modulation disponibles

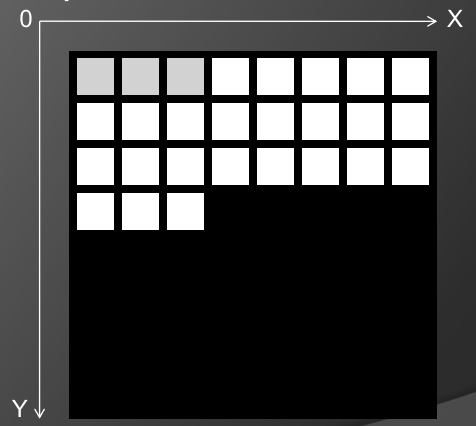
- Tel quel : aucune modification (identité)
- Négations des octets
- Chiffrements par mot de passe
- Chiffrements par décalages

Schémas d'accès à un hôte de type image

- Séquentiel : chaque pixel est pris dans l'ordre, du début à la fin de l'hôte
- Séquentiel inversé : idem séquentiel, à-partir de la fin vers le début de l'hôte
- Uniforme : les pixels sont choisis de manière à répartir uniformément les octets du secret, du début vers la fin
- Uniforme inversé : idem uniforme, à-partir de la fin vers le début de l'hôte

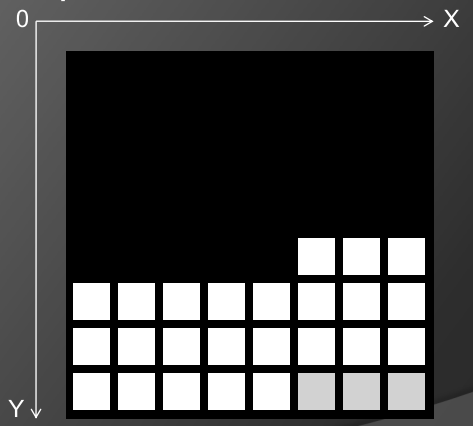
Mise en œuvre des schémas d'accès aux pixels

- Accès séquentiel :



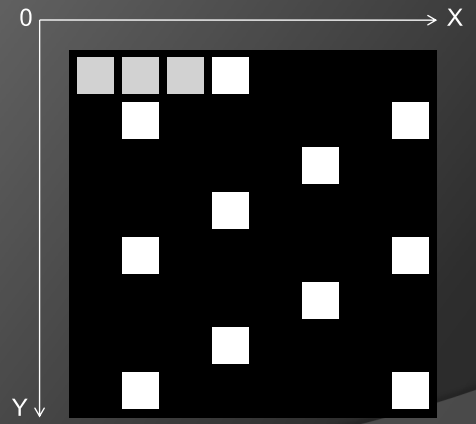
Mise en œuvre des schémas d'accès aux pixels

- Accès séquentiel inversé :



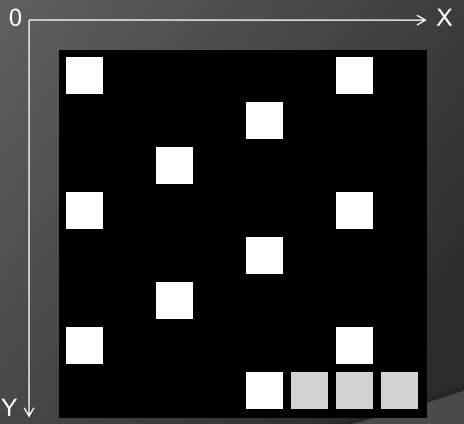
Mise en œuvre des schémas d'accès aux pixels

- Accès uniforme :



Mise en œuvre des schémas d'accès aux pixels

- Accès uniforme inversé :



Schémas de dissimulation dans une image RVB

- Simple LSB :
  - Codage d'un octet sur 3 pixels RVB
  - Pixels 1 et 2 codent 6 bits d'un octet secret, 1 LSB pour chaque couleur
  - Pixel 3 code les 2 bits restants : un LSB rouge et un LSB bleu

$$T_{\min} = 12 + (3 \times t) = 3 \times (t + 4)$$

Schémas de dissimulation dans une image RVB

- Double LSB :
  - Codage d'un octet sur 2 pixels RVB
  - Pixel 1 code 6 bits d'un octet secret, 2 LSBs pour chaque couleur
  - Pixel 2 code les 2 bits restants : un LSB rouge et un LSB bleu

$$T_{\min} = 8 + (2 \times t) = 2 \times (t + 4)$$

Schémas de dissimulation dans une image RVB

- Triple LSB :
  - Codage d'un octet sur 1 pixel RVB : 3 LSBs rouge, 2 LSBs vert et 3 LSBs bleu

$$T_{\min} = t + 4$$

Comparatif des schémas de dissimulation LSB

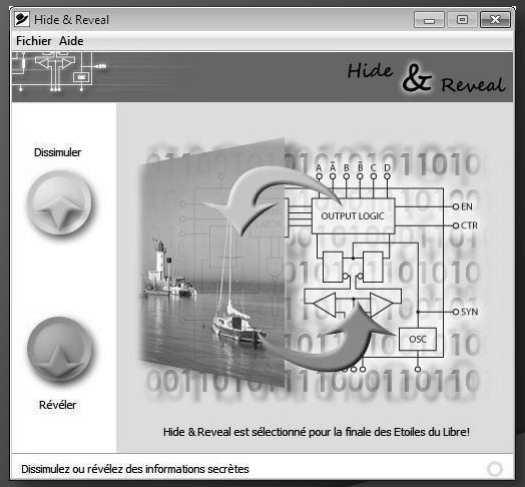
Nom du schéma	Taille hôte minimum*
Simple LSB	$(t+4)\times 3$
Double LSB	$(t+4)\times 2$
Triple LSB	$t+4$

*Rappel : taille exprimée en nombre de pixels nécessaires

Le logiciel « Hide & Reveal »

- Hide & Reveal :
  - 100% Java, sous licence GPL
  - Disponible sur **`www.hidereveal.org`**
  - Sémagramme : image BMP, PNG ou TIF
  - Dissimulation de fichiers quelconques
  - Ajout d'un commentaire secret
  - Chargement dynamique des schémas en fonction du type de conteneur
- Repose sur **`org.steganography`**

Écran d'accueil de « Hide & Reveal »



Dissimulation d'un fichier – étape 1

www.ncottin.net

Dissimulation d'un fichier – étape 2

www.ncottin.net

Découverte du fichier dissimulé – étape 1

www.ncottin.net

Découverte du fichier dissimulé – étape 2

www.ncottin.net

Améliorations à venir

- Support d'autres types de medium :
  - JPEG
  - TXT, HTML, XML, LaTeX
  - WAV, MP3
- Schémas de dissimulation LSB plus compacts et adaptatifs
- Utilisation d'un dossier pour les plugins
- Intégrer une barre de progression
- Installateur graphique

Contact

- nathanael.cottin@yahoo.fr

- Pour en savoir plus sur Hide & Reveal:

`www.hidereveal.org`

Annexes

Quelques informations complémentaires…

Images à double sens

- Jeune fille ou vieille dame ?

- Jeune couple désire enfant

Correspondance masquée

www.ncottin.net

George Sand :

Je suis très émue de vous dire que j'ai bien compris, l'autre jour, que vous avez toujours une envie folle de me faire danser. Je garde un souvenir de votre baiser et je voudrais que ce soit là une preuve que je puisse être aimée par vous. Je suis prête à vous montrer mon affection toute désintéressée et sans cal-cul. Si vous voulez me voir ainsi dévoiler, sans aucun artifice mon âme toute nue, daignez donc me faire une visite

(…)

Alfred de Musset :

Quand je vous jure, hélas, un éternel hommage
Voulez-vous qu'un instant je change de langage
Que ne puis-je, avec vous, goûter le vrai bonheur
Je vous aime, ô ma belle, et ma plume en délire
Couche sur le papier ce que je n'ose dire
Avec soin, de mes vers, lisez le premier mot
Vous saurez quel remède apporter à mes maux

George Sand :

Cette grande faveur que votre ardeur réclame
Nuit peut-être à l'honneur mais répond à ma flamme

Correspondance révélée

www.ncottin.net

George Sand :

Je suis très émue de vous dire que j'ai bien compris, l'autre jour, que vous avez toujours une envie folle de me faire danser. Je garde un souvenir de votre baiser et je voudrais que ce soit là une preuve que je puisse être aimée par vous. Je suis prête à vous montrer mon affection toute désintéressée et sans cal-cul. Si vous voulez me voir ainsi dévoiler, sans aucun artifice mon âme toute nue, daignez donc me faire une visite

(…)

Alfred de Musset :

Quand je vous jure, hélas, un éternel hommage
Voulez-vous qu'un instant je change de langage
Que ne puis-je, avec vous, goûter le vrai bonheur
Je vous aime, ô ma belle, et ma plume en délire
Couche sur le papier ce que je n'ose dire
Avec soin, de mes vers, lisez le premier mot
Vous saurez quel remède apporter à mes maux

George Sand :

Cette grande faveur que votre ardeur réclame
Nuit peut-être à l'honneur mais répond à ma flamme

Dissimulation linguistique : codes de Barn

www.ncottin.net

- Message :

Enfin	voici	le	printemps
un	oiseau	de	passage
chante	un	message	emprunt
d’	un	doux	secret

Dissimulation linguistique : codes de Barn

www.ncottin.net

- Message :

Enfin	voici	le	printemps
un	oiseau	de	passage
chante	un	message	emprunt
d’	un	doux	secret

- Clé : BACD

Chiffre de Trithème : procédé

www.ncottin.net

- Texte secret dissimulé dans une litanie religieuse
- Utilise un code permettant de chiffrer un caractère en portion de phrase :
  - A = dans les cieux
  - B = à tout jamais
  - C = un monde sans fin
  - D = en une infinité
  - E = à perpétuité
  - F = sempiternel
- Les portions de phrases peuvent être complétées avec des ponctuations

Chiffre de Trithème : exemple

www.ncottin.net

- Le mot secret “FADE” est ainsi dissimulé dans la prière suivante :

  “Sempiternellement dans les cieux,
  En une infinité et à perpétuité”

Chiffre bilitère : procédé

www.ncottin.net

- Représentation des caractères en séquences binaires de ‘A’ et de ‘B’ :
  - A = AAAAA
  - B = AAAAB
  - C = AAABA
  - D = AAABB
  - E = AABAA
  - F = AABAB
- Chaque caractère en clair est remplacé par sa séquence binaire
- Dans un texte hôte quelconque, les ‘B’ sont représentés en italique

Chiffre bilitère : exemple

www.ncottin.net

- Le texte en clair “MESSAGE” est chiffré en :

  ABABB AABAA BAAAB BAAAB AAAAA ...

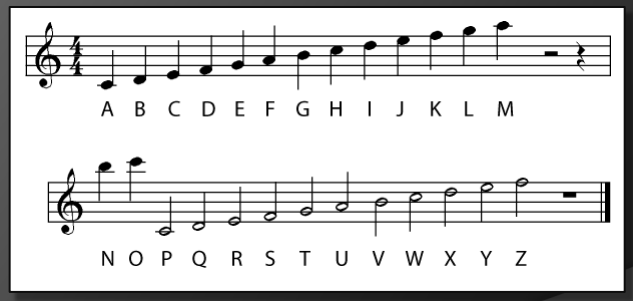
- Un texte renfermant ce code sera par exemple :

  C*e*c*i* *e*st *u*n t*e*xte *an*odi*n* …

Partition de musique : procédé

www.ncottin.net

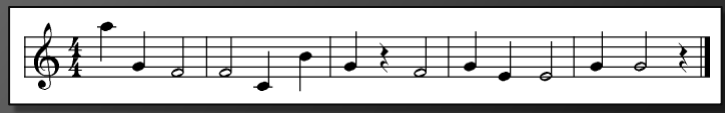
- Définition de correspondances entre caractères et notes de musique



Partition de musique : exemple

www.ncottin.net

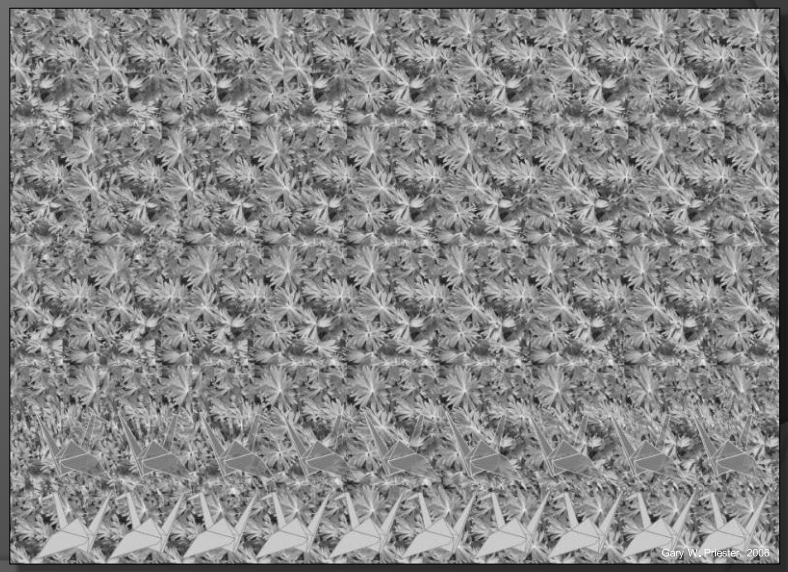
- Le texte “MESSAGE SECRET” est masqué dans la partition suivante :



- Les figures de silences servent de séparateurs de mots

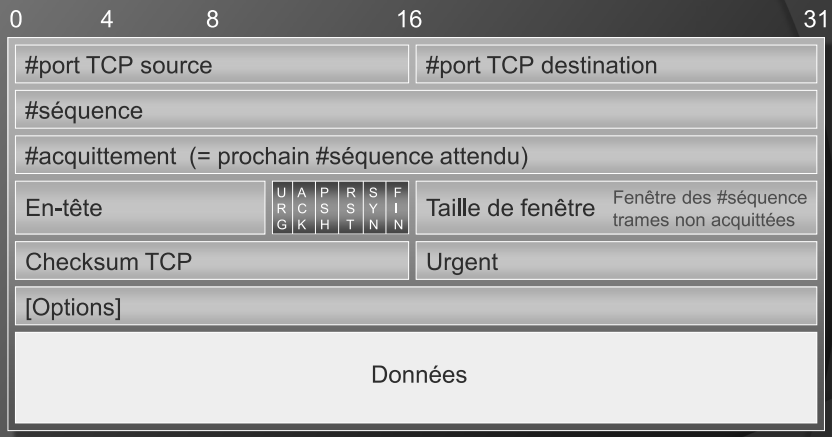
Stéréogrammes

www.ncottin.net



Dissimulation dans les en-têtes TCP/IP

www.ncottin.net



URG	interpréter message urgent	RST	réinitialiser la connexion
ACK	interpréter #acquittement	SYN	synchroniser les # séquence
PSH	transmission immédiate	FIN	terminer la connexion