Une image peut en cacher

JEAN-CHRISTOPHE SAHAKIAN [1]

Voici un TP qui permet de faire découvrir aux élèves de terminale S-SI les deux principales approches de la sécurité informatique, par la dissimulation avec la stéganographie et par le chiffrement avec la cryptographie.

mots-clés
information

Selon Gartner, entreprise américaine de conseil et de recherche dans le domaine des techniques avancées, le marché de la sécurité informatique au sens large, protection des transactions bancaires, données personnelles, dossiers médicaux, etc., représentera 86 milliards de dollars en 2016. Il est donc important de sensibiliser les élèves à cette problématique, pour qu'ils prennent conscience aussi bien de l'importance de garantir la sécurité et la confidentialité des échanges de données, notamment dans un contexte de compétition industrielle, que de l'avenir de ce secteur. Voici donc un TP de 3 h 30 qui, après avoir replacé dans un contexte historique et social les enjeux de la sécurisation des échanges d'informations numériques, aborde deux techniques simples et efficaces de protection des données. Précisons qu'il constitue en fait une partie de l'introduction à une séquence sur le thème de la communication d'une durée de six semaines, que vous pouvez retrouver en ligne avec tous ses documents à l'adresse suivante :

http://sahakianjc.free.fr/

Le besoin

La cryptographie est la science du chiffrement, codage des messages à l'aide de codes secrets ou de clés. Le chiffrement des messages vise à en assurer la confidentialité, l'authenticité et l'intégrité. Il permet de rendre l'information secrète. Les gouvernements, les armées et les industriels l'utilisent afin de protéger certaines informations, et considèrent pour cette raison la cryptologie (qui englobe la cryptographie et la cryptanalyse, science du décryptage, c'est-à-dire du déchiffrement sans clé) comme une arme. Les individus l'utilisent maintenant également dans le cadre de leur vie privée. Le développement des outils informatiques permet aujourd'hui une sécurité accrue dans les échanges d'informations, mais facilite aussi le cassage des protections par des personnes indélicates. Les principales techniques des pirates sont l'utilisation soit de la force brute (générer systématiquement les différents codes possibles jusqu'à trouver le bon) soit d'une technique comme l'analyse fréquentielle qui permet de repérer certains motifs répétitifs dans le texte codé et d'en déduire la clé de chiffrement.

[1] Professeur de SI au lycée Jean-Perrin de Marseille (13010). Courriel : sahakianjc13@gmail.com.

La cryptographie permet donc de répondre à une triple problématique, qui induit des contraintes :

- **L'authenticité** : il faut assurer au destinataire d'un message chiffré que son émetteur est bien celui qu'il prétend être.
- **La confidentialité** : il faut assurer à l'émetteur du message chiffré que son destinataire sera le seul à pouvoir le lire.
- **L'intégrité** : il faut que le contenu du message n'ait subi aucune altération entre son envoi et sa réception.

L'entrée du TP se fait à travers plusieurs questions qui relèvent du champ A (analyser) du référentiel, en particulier du domaine A1 (analyser le besoin).

L'objectif de cette partie est d'analyser le besoin à l'origine de la cryptographie (pourquoi chiffrer ?), puis de comparer la solution retenue avec une autre possible.

Pour cette entrée en matière, l'élève prend appui sur un dossier ressource (disponible en ligne). Tout d'abord, il énumère plusieurs usages actuels de la cryptographie. Puis il verbalise le besoin : À qui ? Sur quoi ? Pour quoi faire ? Il est alors amené à compléter le tableau 1 de caractérisation sur les trois critères fondamentaux.

Pour conclure cette introduction, l'élève justifie la pertinence des techniques actuelles de chiffrement par rapport au chiffre de Vigenère ou à la machine Enigma. Cela lui permet de situer ces technologies dans le continuum de l'histoire et de comprendre que finalement les techniques d'aujourd'hui ne sont que des évolutions rendues possibles par les progrès des ordinateurs.

La stéganographie

Cette deuxième partie du TP aborde une technique classique de sécurisation : la stéganographie, ou l'art de la dissimulation. Le principe est simple : il s'agit de faire passer inaperçu un message dans un autre message. On doit le premier usage connu de cette technique, aux alentours de 600 av. J.-C., à Nabuchodonosor, roi de Babylone, qui employait une méthode originale : il écrivait sur le crâne rasé de ses esclaves, attendait que leurs cheveux aient repoussé, et les envoyait à ses généraux. Il suffisait ensuite de raser à nouveau le

Tableau de caractérisation	
Critères	Valeur
Authenticité	Totale
Confidentialité	Totale
Intégrité	Totale

À qui ? aux utilisateurs qui souhaitent échanger de manière confidentielle des données sensibles
Sur quoi ? un message numérique (voix, données, images)
Pour quoi faire ? assurer la confidentialité de la communication

1 L'analyse du besoin : les documents à compléter

u ne autre

messager pour lire le texte. Il s'agit bien de stéganographie à proprement parler et non de cryptographie : l'information est cachée et non codée.

Le questionnement permet d'aborder les points A2, puis B2 et B4 du référentiel :

A. Analyser

A2. Analyser le système
Identifier les éléments transformés et les flux
Identifier l'organisation structurelle

Objectif de cette partie : décrire et analyser le comportement d'un système

B. Modéliser

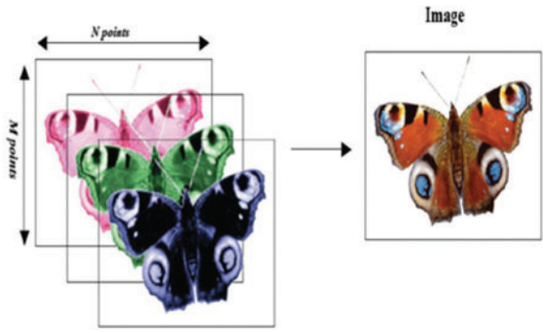
B2. Proposer ou justifier un modèle
Associer un modèle à un système ou à son comportement

B4. Valider un modèle et interpréter les résultats obtenus

Objectifs de cette partie : traduire le comportement d'un système ; vérifier la compatibilité des résultats obtenus ; comparer les résultats obtenus

Dans cette partie de l'expérimentation, l'élève met en œuvre un procédé de stéganographie toujours très utilisé aujourd'hui. Il dispose pour cela d'un PC équipé du logiciel Matlab (ou autre).

Tout d'abord, il est utile de savoir comment Matlab gère les images. Une image couleur est la superposition de trois composantes de base (rouge, vert et bleu, RVB). Sous Matlab, une telle image peut-être codée par un tableau tridimensionnel **2**. Il correspond à la mise en cascade des trois tableaux (2D : N × M) correspondant aux trois composantes primaires. Chacun de ces trois tableaux primaires (aussi appelés plans couleur) contient le niveau de couleur pour chaque point de l'image considérée. En général, chaque niveau de couleur est codé de 0 (canal colorimétrique éteint) à 255 (canal colorimétrique au maximum), ce qui correspond à 255^3 = 16 581 375 combinaisons de couleurs possibles.



2 La composition d'une image couleur

Le codage

La technique mise en œuvre dans cette activité est très simple, le but étant d'incruster l'image secrète dans l'image anodine du conteneur.

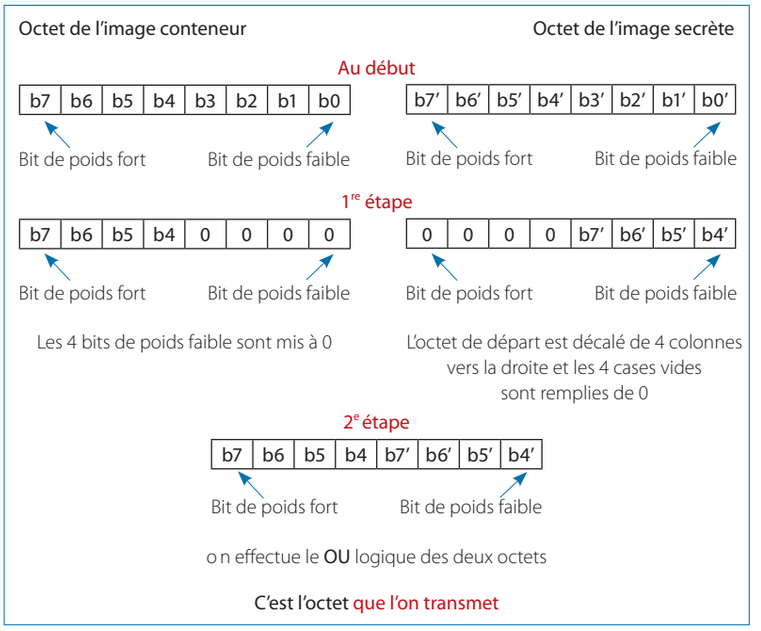
On choisit tout d'abord deux images au format BMP (ou JPEG), la première se nommant « image_conteneur.bmp » (ou « .jpg »), la seconde « image_secrete .bmp » (ou « .jpg »).

Initialement, chacun des points de l'image conteneur est défini par un octet (de b0 à b7 sur la figure **3**), tout comme chaque point de l'image secrète (de b0' à b7'). La première étape consiste, dans l'image conteneur, à mettre à 0 le quartet de poids faible (de b0 à b3) et, dans l'image secrète, à décaler vers la droite le quartet de poids fort (de b4' à b7'), devenu de poids faible, le quartet de poids fort se retrouvant à 0.

Lors de la seconde étape, on réalise un OU logique entre les deux octets obtenus. Les bits de poids forts de l'image secrète sont transmis dans l'image conteneur et occupent un rang mineur (poids faible).

Le décodage

La première étape consiste à mettre à 0 le quartet de poids fort de l'octet reçu **4**. Ensuite, on décale vers la gauche le quartet de poids faible, devenant fort, et on met à 0 le quartet de poids faible. On remarque que l'image récupérée a perdu en qualité par rapport à l'original, car le quartet de poids faible est perdu. Les élèves reviendront sur ce point lors de l'analyse des écarts.



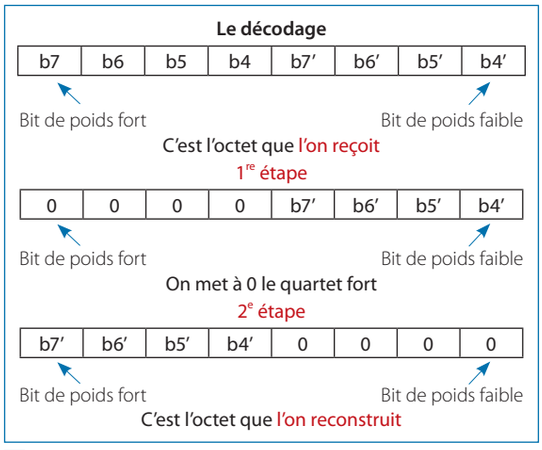
3 Les étapes du codage

Les activités

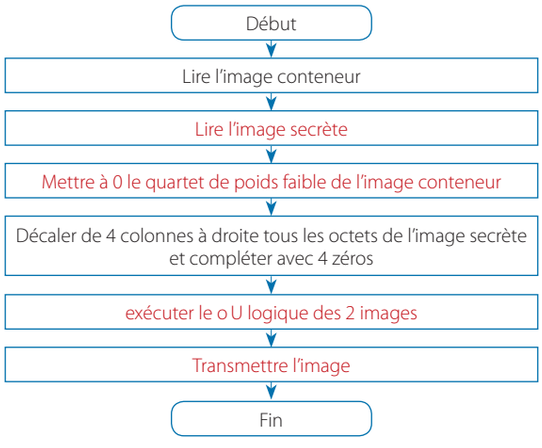
On propose aux élèves, après qu'ils ont pris connaissance des manipulations sur les octets, de compléter deux algorigrammes relatifs au codage et au décodage 5 6.

Les élèves ont à leur disposition, sous Matlab, un programme de stéganographie 7. Ils en complètent plusieurs instructions, puis le lancent 8. Le programme de codage nécessite trois instructions et celui de décodage deux, ce qui est tout à fait convenable en termes de complexité pour des élèves de terminale. Les lignes qui débutent avec « % » sont des commentaires.

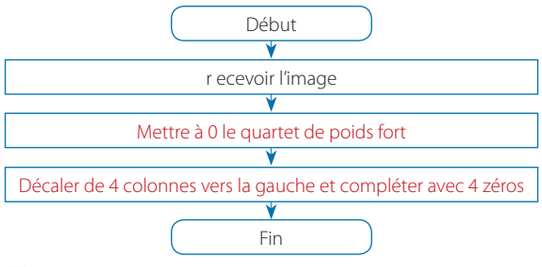
Les résultats relatifs aux différentes images sont automatiquement générés 9.



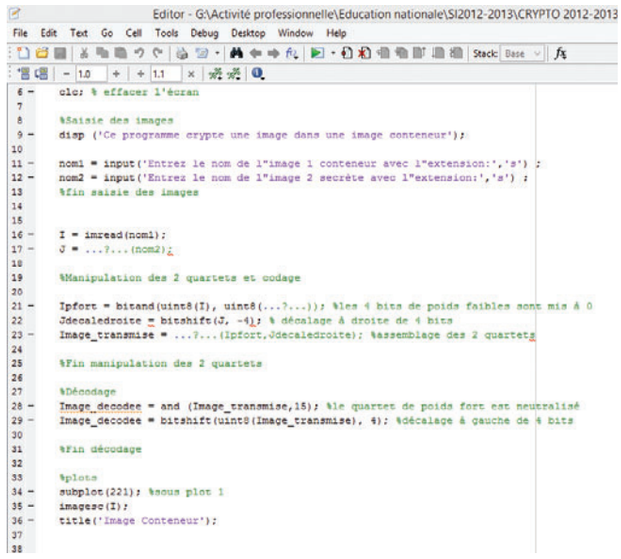
4 Les étapes du décodage



5 Les algorigrammes relatifs au codage



6 Les algorigrammes relatifs au décodage



7 Le programme élève à compléter

L'élève est ensuite amené, guidé par différentes questions, à observer la valeur de quelques octets dans les différentes images (sous Matlab, un clic droit suffit). L'objectif pédagogique est alors, en conformité avec le référentiel (A3), d'analyser les écarts entre le souhaité, le simulé et le réalisé.

La figure 10 correspond à un zoom sur le quadrant nord-est de l'image secrète source et sur celui de l'image secrète reconstruite. Des défauts sont visibles. Les élèves s'aperçoivent que l'image secrète reconstruite n'est pas égale à l'image secrète source (les 4 bits de poids faibles sont à 0). Ils sont alors questionnés sur la satisfaction des trois critères, authenticité, confidentialité, intégrité.

Il apparaît que l'intégrité du message n'est pas totale : l'image secrète est dégradée, car seul le quartet de poids fort est conservé. Quant à l'authenticité et la confidentialité, elles ne peuvent être garanties : rien ne prouve que l'image secrète n'a pas été interceptée, voire transformée, durant la communication par un tiers indélicat connaissant la technique.

L'information secrète n'est pas chiffrée, elle n'est que cachée ! Pour sécuriser l'échange, il faut chiffrer l'image avec une clé informatique, ou bien insérer les bits de l'image secrète dans l'image conteneur 1 à 1 et non plus par quartet – l'inconvénient étant que le format de l'image conteneur est alors 8 fois plus grand, le temps de transmission évoluant en conséquence.

La technique employée dans cette partie du TP est assez rustique, et conduit à la non-satisfaction des trois critères essentiels. C'est le prix à payer pour avoir un algorithme simple… et une mise en œuvre motivante pour les élèves.

La cryptographie

La troisième partie du TP aborde la technique de chiffrement suivant l'algorithme de Naor et Shamir (le S du

Editor - G:\Activité professionnelle\Education nationale\SI2012-2013\CRYPTO 2012-2013\

```
clc; % effacer l'écran

%Saisie des images
disp ('Ce programme crypte une image dans une image conteneur');

nom1 = input('Entrez le nom de l"image 1 conteneur avec l"extension:','s') ;
nom2 = input('Entrez le nom de l"image 2 secrète avec l"extension:','s') ;
%fin saisie des images


I = imread(nom1); %lecture de l'image containeur
%J = ...?...(nom2); %lecture de l'image secrète
J = imread(nom2);

%Manipulation des 2 quartets et codage
%Ipfort = bitand(uint8(I), uint8(...?...)); %les 4 bits de poids faibles sont mis à 0
Ipfort = bitand(uint8(I), uint8(240)); %les 4 bits de poids faibles sont mis à 0
Jdecaledroite = bitshift(J, -4); % décalage à droite de 4 bits
%Image_transmise = ...?...(Ipfort,Jdecaledroite); %assemblage des 2 quartets
Image_transmise = bitor(Ipfort,Jdecaledroite); %assemblage des 2 quartets

%Fin manipulation des 2 quartets

%Décodage
%Image_decodee = and (Image_transmise,...?...); %le quartet de poids fort est neutralisé
Image_decodee = and (Image_transmise,15); %le quartet de poids fort est neutralisé
%Image_decodee = bitshift(uint8(Image_transmise), ...?...); %décalage à gauche de 4 bits
Image_decodee = bitshift(uint8(Image_transmise), 4); %décalage à gauche de 4 bits

%Fin décodage

%plots
```

steganographie_eleves.m | steganographie.m

8 Le programme élève complété

9 Les résultats du codage et du décodage : les images automatiquement générées

chiffrement RSA). Le questionnement permet d'aborder les points B2 et B4 du référentiel :

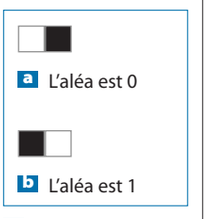
B. Modéliser

B2. Proposer ou justifier un modèle

Associer un modèle à un système ou à son comportement

B4. Valider un modèle et interpréter les résultats obtenus

Objectif de cette partie : traduire le comportement d'un système ; vérifier la compatibilité des résultats obtenus, comparer les résultats obtenus



13 La création du bipixel clé

10 Zoom pour comparaison des deux images

La cryptographie visuelle permet de mettre en place des systèmes où c'est l'œil humain qui va être seul capable de décoder totalement et d'authentifier le message (les Anglo-Saxons parlent de *visual secret sharing*, VSS).

Le codage

L'image est tout d'abord binarisée (deux niveaux sont conservés, noir et blanc) **11**. Après binarisation, l'image source est constituée de 0 (pixels noirs) et de 1 (pixels blancs). Puis le chiffrement est effectué **12** :

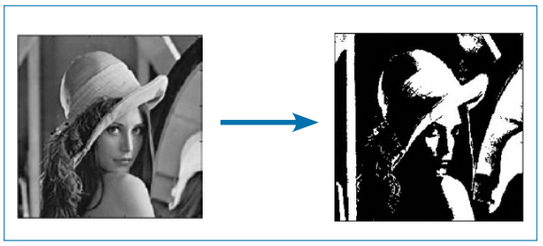
❶ **Création d'un tableau aléatoire de 0 et de 1.** Le tableau a la même taille que l'image à chiffrer. Donc à chaque pixel de l'image source correspond un aléa, soit 0 soit 1.

❷ **Création du bipixel clé** : si l'aléa est 0, alors la clé est le bipixel horizontal donné en **13 a** ; si l'aléa est 1, alors la clé est le bipixel horizontal donné en **13 b**. C'est une règle qui doit être connue des deux parties (expéditeur et destinataire). Elle est secrète, et doit être acheminée par un canal différent.

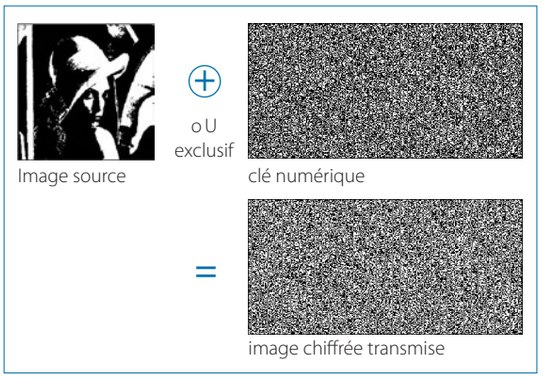
❸ **OU exclusif** entre chaque composante du bipixel et le pixel de l'image binarisée. Le résultat est une image chiffrée dont la taille fait le double de celle de l'image source.

❹ **Transmission** de l'image chiffrée.

Prenons l'exemple du codage d'un pixel quelconque, de coordonnées $x = 123$ et $y = 228$, de l'image à coder préalablement binarisée de format 250×250. Dans l'hypothèse retenue, ce pixel est noir (0 binaire) **14**. À cette image, on associe une matrice de tirage aléatoire (0 ou 1) de même format. Aux coordonnées



11 La binarisation



12 Le chiffrement

(123,228), le tirage a donné 0. Le codage consiste à associer aux valeurs aléatoires une clé de deux bits produisant une nouvelle matrice de format 250 × 500. Dans notre exemple, au 0 sera associé la clé de deux bits 1,0. L'image codée sera obtenue en réalisant l'opération OU exclusif entre la valeur correspondant à la couleur du pixel et la clé. Dans notre exemple, 0 ⊕ 1,0 = 1,0. Cette opération répétée pour chaque pixel de l'image binarisée donnera l'image chiffrée de format 250 × 500.

Le décodage

On réalise successivement les opérations suivantes 15 :

❶ Réception de l'image chiffrée.

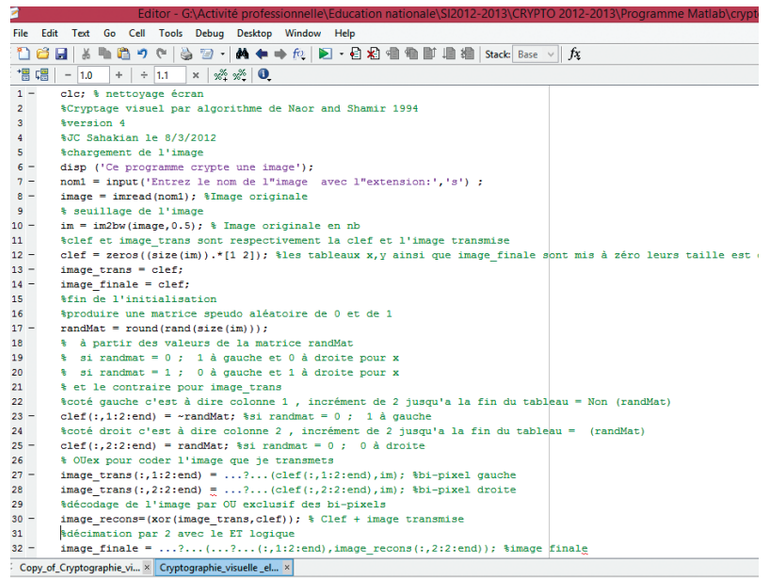
❷ OU exclusif entre l'image reçue et la clé.

❸ Décimation par 2 (on conserve un sous-pixel sur deux).

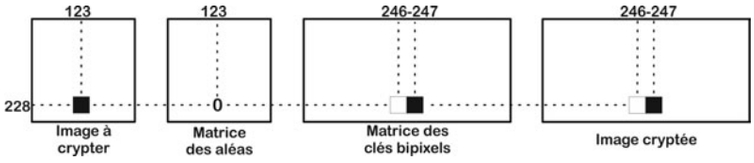
Le processus complet est présenté dans le tableau 16.

Pour garantir la sécurité, trois fortes contraintes existent sur la clé (secrète). Celle-ci doit être totalement aléatoire, de même longueur que le message à chiffrer et, surtout, à usage unique.

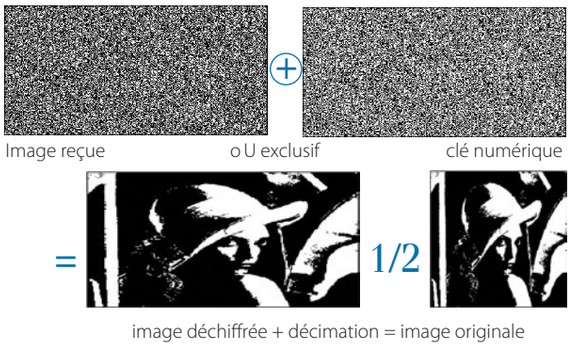
Dans notre cas, nous appliquerons ce procédé à une image en niveaux de gris, mais le concept est déclinable pour des images en couleurs. Il peut paraître décevant de ne pas reconstituer l'image source originale, mais seulement l'image binarisée. C'est en fait un avantage, car si une machine arrive à casser le code en un temps acceptable, elle ne peut pas pour autant l'identifier avec certitude. Il est nécessaire de procéder alors à une authentification humaine du ou des éléments de l'image. Cela augmente encore le temps de décryptage.



17 **Le programme élève à compléter**



14 **Le codage d'un pixel**



15 **Le déchiffrement**

Les activités

Les élèves prennent connaissance des manipulations sur les octets. Une fiche support sur le langage Matlab est à leur disposition. Ils complètent le programme de codage et de décodage fourni par le professeur 17 18. La programmation permet aux élèves de mettre en œuvre les fonctions logiques élémentaires ET, OU exclusif (sous forme logicielle) 19.

Le résultat est alors automatiquement généré par le logiciel Matlab, de la façon que l'on peut voir sur la figure 20, qui reprend notre exemple (en 21, les résultats pour un message constitué d'un document texte en noir et blanc, et en 22 pour une image couleur). On remarquera que le doublement du format horizontal de l'image chiffrée n'est indiqué que par le changement d'échelle.

Image source binarisée	Aléa	Clé	Pixel image binarisée OU exclusif clé	Transmission	Bipixel reçu OU exclusif clé	Image reconstruite (décimation par 2)
Pixel noir ■	0	□■	□■	□■	■■	■
	1	■□	■□	■□	■■	■
Pixel blanc □	0	□■	■□	■□	□□	□
	1	■□	□■	□■	□□	□

16 **Le processus complet de déchiffrement**

E1	E2	S
0	0	0
0	1	1
1	0	1
1	1	0

OU exclusif

E1	E2	S
0	0	0
0	1	0
1	0	0
1	1	1

ET

19 Les tables de vérité des fonctions logiques élémentaires

Pour finir, l'élève, à travers plusieurs questions, est amené à observer la valeur de quelques octets dans les différentes images. L'objectif pédagogique est alors d'analyser les écarts entre le souhaité, le simulé et le réalisé (A3). On « reboucle » ainsi le questionnaire sur l'estimation de l'erreur et la satisfaction des trois critères définis lors de l'analyse fonctionnelle. Les élèves s'aperçoivent que le texte reconstruit est égal au texte source, et que les images reconstruites sont identiques aux images binarisées.

Le bilan des activités

La première partie du TP nécessite de 1 h 30 à 2 h 00. Pour que les élèves les plus rapides ne se dispersent pas, on peut prévoir une activité du type « pour aller plus loin », où ils devront imaginer des techniques simples qui satisfassent entièrement aux trois critères susvisés. Différentes solutions sont envisageables, telles que l'insertion de l'image secrète bit à bit, le chiffrement à clé symétrique avec un OU exclusif, l'entrelacement dans les différents plans couleur, etc.

La seconde partie est légèrement plus courte. Les élèves manipulent le logiciel Matlab plutôt avec aisance, et sont vraiment captivés par les notions abordées.

Au final, ce TP aborde de manière ludique deux techniques efficaces en termes de sécurité numérique, et suscite un vif intérêt. ■

En ligne

Stéganographie

r APh A (W.), *Traitement parole image : Stéganographie par remplacement des bits de poids faible : principe et implémentation*, École centrale de Paris, 2005 :

http://people.via.ecp.fr/~bilou/tpi/RapportTPI.pdf

Cryptographie visuelle : algorithme de Naor et Shamir

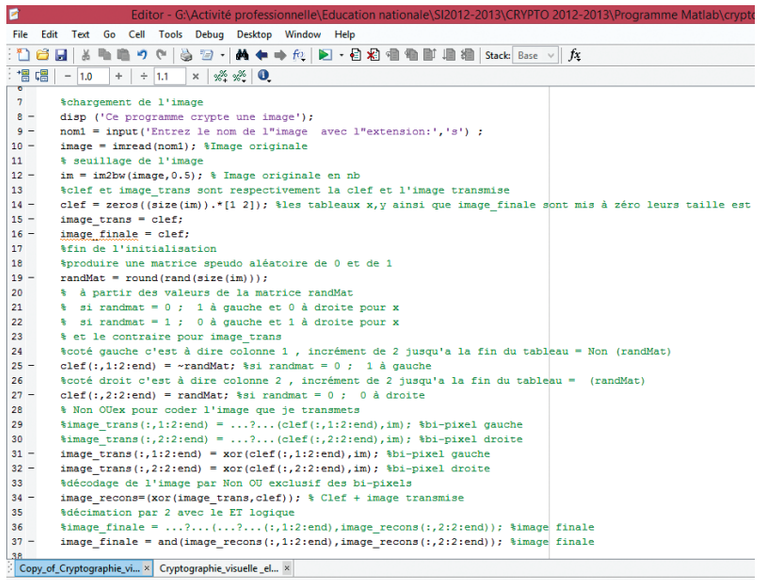
« Secret Sharing and Visual Cryptography » (en anglais) :

www.cs.jhu.edu/~fabian/courses/CS600.624/slides/VisualCrypto.pdf

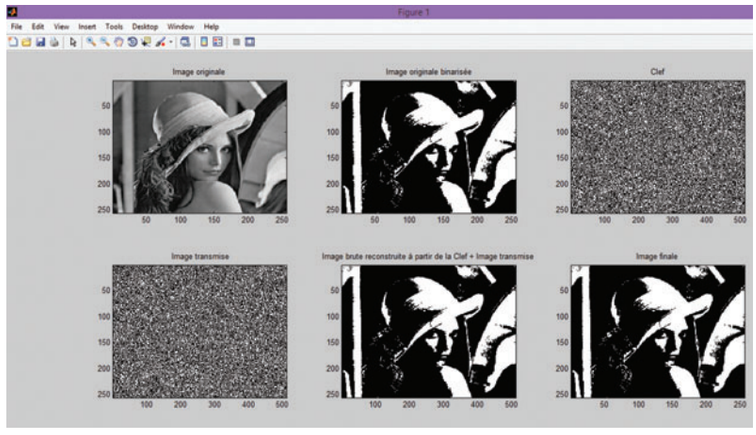
« Une aventure de James Bond » :

https://wiki.inria.fr/sciencinfolycee/Cryptographie_visuelle_(une_aventure_de_James_Bond)

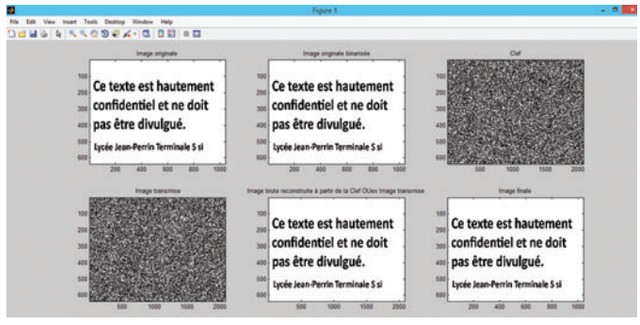
Retrouvez tous les liens sur
http://eduscol.education.fr/sti/revue-technologie



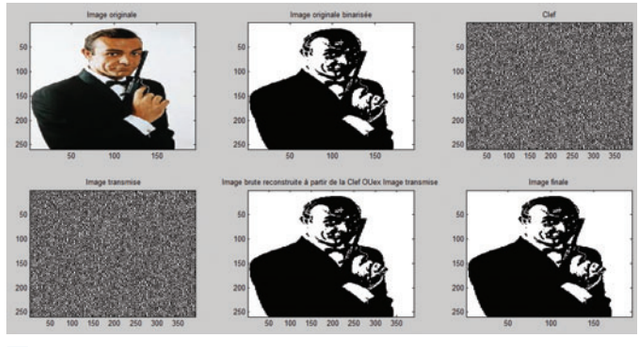
18 Le programme complété



20 Le traitement d'une image en niveaux de gris



21 Le traitement d'un texte en noir et blanc



22 Le traitement d'une image en couleurs