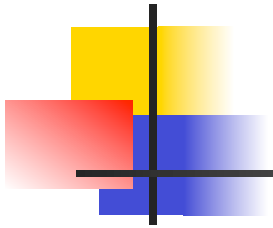
Cours commun TELECOM Nancy Apprentissage 3A et Master ISC – Parcours Bio-santé numérique

« Tatouage d'images médicales : de la protection à l'enrichissement de l'information »

J-M. MOUREAUX

jean-marie.moureaux@univ-lorraine.fr

UMR 7039



Plan du Cours

1. Principes du tatouage
2. Application aux images médicales

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Transmission de documents confidentiels et sécurité

Problème très ancien

transmettre des informations secrètes (militaires), déjouer la censure …

déjà au V^{ème} siècle avant Jésus Christ …

- chefs de guerre qui tatouaient le crâne rasé d'esclaves
- parchemin enroulé autour d'un bâton (la scytale ou bâton de Spartakus)

Aujourd'hui

- **Organisation Mondiale de la Propriété Intellectuelle** (près de 200 états membres)
- **cadre particulier de la protection juridique des documents numériques (premier traité signé le 20/12/96)**

protocole de protection des images numériques qui transitent sur Internet

(enregistrement + tatouage)

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Propriété intellectuelle*

À vie pour le créateur d'une œuvre (le seul autorisé à autoriser la reproduction), puis pendant 5 ans pour ses descendants, ensuite domaine public

Depuis l'avènement du Peer-to-Peer : explosion des copies illégales

Impact des pertes en France en 2007 (Etude de la Sacem – 2008)

- **musique : 369 Millions €, 1600 emplois directs perdus**
- **cinéma : 605 Millions €, 2400 emplois directs perdus**
- **télévision : 234 Millions €, 950 emplois directs perdus**
- **livre : 147 Millions €, 750 emplois directs perdus**

* http://damiencalmes.zapto.org/memoire/memoire_DRM_final.pdf

Gestion des droits d'auteur (DRM : digital right management)*

DRM = ensemble d'outils et de techniques permettant de contrôler l'usage des œuvres numériques :

- **pour des supports physiques (DVD, Blue Ray, logiciel)**
- **pour la transmission (Internet, télédiffusion)**

Exemples de protection des droits d'auteur :

- **restriction ou interdiction de copie**
- **restreindre la lecture d'un DVD à une zone géographique**
- **limiter l'utilisation d'une carte SIM à un type de téléphone autorisé**

* http://damiencalmes.zapto.org/memoire/memoire_DRM_final.pdf

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

DRM : l'exemple de Windows Media Player*

1. Le « créateur » (= fournisseur de contenu) utilise le fichier Windows Media Rights Manager, celui-ci fabrique le fichier chiffré (algorithme AES) et une clé déposée dans un second fichier chiffré et signé par l'algorithme RSA
2. Le fichier chiffré contenant l'œuvre peut alors être mis sur le web, sur un serveur de streaming, envoyé par email ou copié sur un cd
3. Le fournisseur de contenu charge un serveur de transactions de gérer ses droits vis-à-vis des clients
4. Pour pouvoir consulter le fichier chiffré, le client doit d'abord se procurer une licence contenant la clé pour déverrouiller le fichier. Le processus d'acquisition démarre automatiquement quand le client tente de récupérer le média chiffré ou de le lire la première fois : processus transparent ou envoi d'un formulaire d'enregistrement

* http://damiencalmes.zapto.org/memoire/memoire_DRM_final.pdf

DRM : l'exemple de Windows Media Player*

La licence spécifie et contient les différents droits accordés au fichier :

1. **• Combien de fois le fichier peut être lu.**
2. **• Sur quel type d'appareil le fichier peut-il être transféré ; par exemple sur un lecteur MP3 portable. L'appareil doit faire partie du standard« Secure media Music Initiative ».**
3. **• La date de début d'autorisation et de fin de lecture.**
4. **• Le fichier peut-il être gravé.**
5. **• Si l'utilisateur peut enregistrer et réactiver la licence .**
6. **• Quel est le niveau de sécurité nécessaire sur le poste client pour lire le fichier.**

* http://damiencalmes.zapto.org/memoire/memoire_DRM_final.pdf

Transmission sécurisée : les méthodes

Cryptographie

- transformer un message pour qu'il devienne illisible
- clé + moyen de cryptage ⟹ décodage

substitution de lettres d'alphabets décalés (Jules César)…algorithme RSA (Internet)

Stéganographie

- dissimuler un message dans un autre
- connaissance du procédé de dissimulation ⟹ décodage

pochoirs superposés (ère médiévale)…encre invisible (2nde guerre mondiale) …

Tatouage (d'images)

- insérer une signature invisible et indélébile dans une image
- clé secrète + règle ⟹ décodage

schémas substitutifs, additifs … (années 90)

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Cryptographie ancienne

L'exemple du code de César :

substitution monoalphabétique la plus ancienne connue de l'Histoire

Texte clair A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Texte codé D E F G H I J K L M N O P Q R S T U V W X Y Z A B C

Texte à coder : TELECOM Nancy ouvre ses portes pour les cours d'ouverture.

Texte codé : WHOHFRP QDQFB RXYUH VHV SRUWHV SRXU OHV FRXUV G'RXYHUWXUH.

26 décalages possibles seulement :
code très peu sûr mais très longtemps utilisé (simplicité)

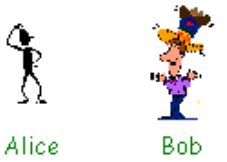
D'après http://www.bibmath.net/crypto/moderne/clepub.php3

Cryptographie moderne à clé publique

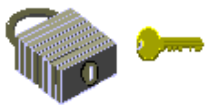
Lorsqu'on ne peut avoir recours à la valise diplomatique …

Algorithme RSA

Cryptographie à clé publique :

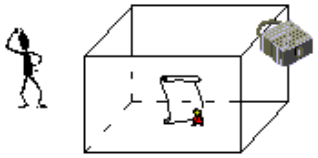


Etape 1 : Fabrication des clés. Bob fabrique une clé publique qui permet de sceller le message codé dans la boite (ici : le cadenas), et une clé privée qui permet d'ouvrir le cadenas.

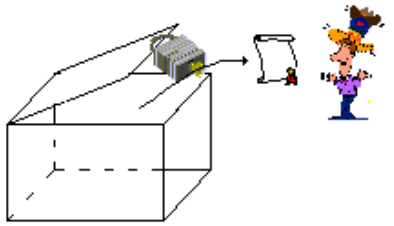


Etape 2 : Distribution des clés. Bob fait parvenir à Alice le cadenas, mais garde la clé pour lui.

Etape 3 : Envoi du message. Alice met son message dans une boite qu'elle ferme à l'aide du cadenas.



Etape 4 : Réception du message. Bob ouvre la boite a l'aide de sa clé, et récupère le message. Personne n'a pu l'intercepter puisque lui seul pouvait ouvrir la boite.



D'après http://www.bibmath.net/crypto/moderne/clepub.php3

Algorithme RSA (Rivest, Shamir, Adleman – 1977)

1- Création des clés :

Bob crée 4 nombres p, q, e et d :

p et q sont deux grands nombres premiers distincts.
Leur génération se fait au hasard, en utilisant un algorithme de *test de primallité* probabiliste.
e est un entier premier avec le produit (p-1)(q-1).
d est tel que ed=1 modulo (p-1)(q-1). Autrement dit, ed-1 est un multiple de (p-1)(q-1).
On peut fabriquer d à partir de e, p et q, en utilisant *l'algorithme d'Euclide*.

2- Distribution des clés :

Le couple (n,e) constitue la clé publique de Bob. Il la rend disponible par exemple en la mettant dans un annuaire.
Le couple (n,d) constitue sa clé privée. Il la garde secrète.

3- Envoi du message codé :

Alice veut envoyer un message codé à Bob.
Elle le représente sous la forme d'un ou plusieurs entiers M compris entre 0 et n-1.
Alice possède la clé publique (n,e) de Bob. Elle calcule C=Me mod n. C'est ce dernier nombre qu'elle envoie à Bob.

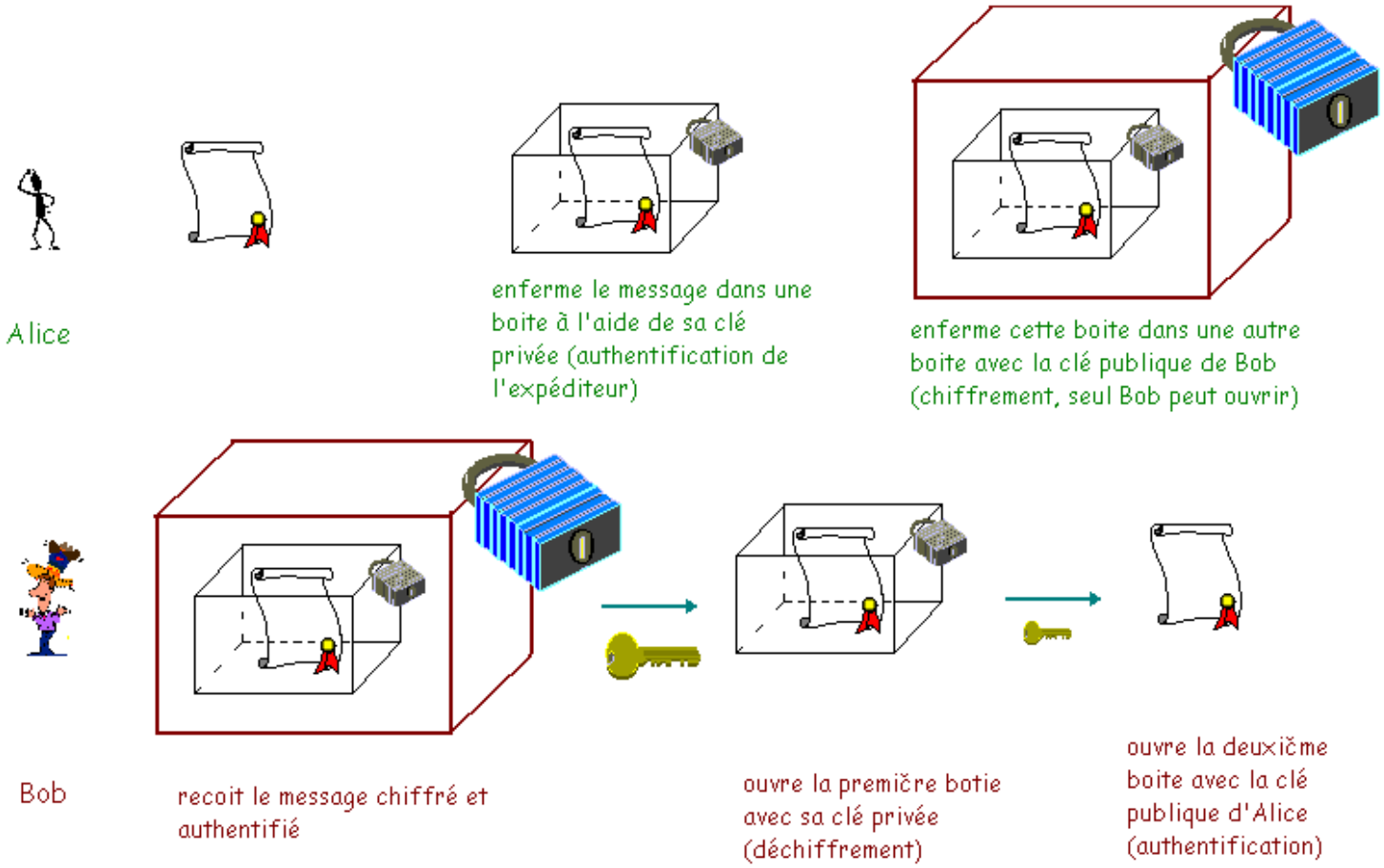
4- Réception du message codé :

Bob reçoit C, et il calcule grâce à sa clé privée D=Cd (mod n).
D'après un théorème du mathématicien Euler, D=Mde=M (mod n). Il a donc reconstitué le message initial.

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

La signature électronique

Pour être sûr de l'expéditeur … (être sûr que quelqu'un ne cherche pas à vous envoyer un message en se faisant passer pour quelqu'un d'autre)

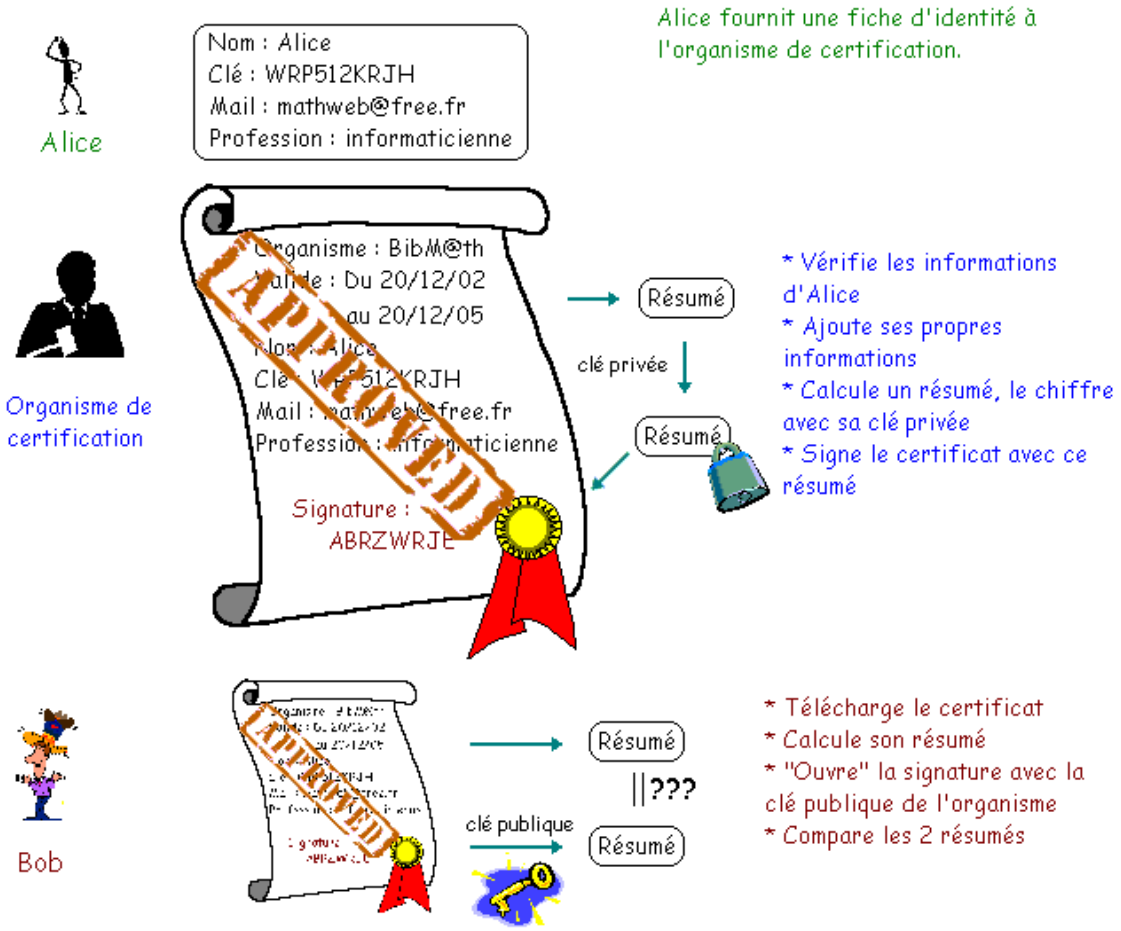


D'après http://www.bibmath.net/crypto/moderne/clepub.php3

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

La certificat électronique

Pour être sûr du destinataire … (être sûr que Bob n'envoie pas son n° de CB à un pirate)



Certification numérique d'une clé publique

D'après http://www.bibmath.net/crypto/moderne/clepub.php3

Stéganographie

Cacher plutôt que chiffrer …

histoire-stéganographie

D'après http://www.bibmath.net/crypto/moderne/clepub.php3

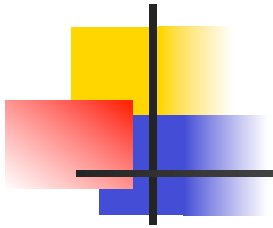
J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Stéganographie

Cacher plutôt que chiffrer …

démo-stéganographie

D'après http://www.bibmath.net/crypto/moderne/clepub.php3



Un peu de vocabulaire

Cryptographie : ensemble d'outils permettant d'assurer la **confidentialité** (chiffrement symétrique ou asymétrique), l'**intégrité** (signature) ou l'**authentification**

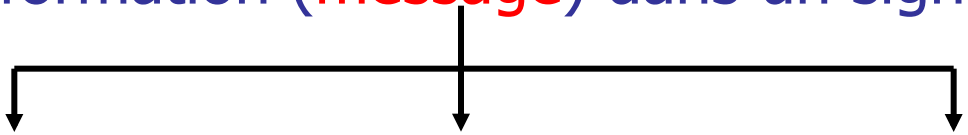
Stéganographie : ensemble d'outils permettant la **dissimulation** d'un message utile dans un message de couverture (hôte) **sans qu'on puisse sans rendre compte**

Tatouage : ensemble d'outils permettant la **dissimulation** d'un message utile dans un message de couverture (hôte) **sans qu'on puisse sans rendre compte** et **de façon à ce que le message utile soit toujours présent si le message de couverture subit des modifications préservant sa sémantique**

Fingerprinting : marquage d'un nouveau message utile dans un document cédé à un nouvel acquéreur

Le tatouage en quelques mots

TATOUER = dissimuler une information (message) dans un signal hôte



Interprétable :
Contrainte d'extraction du message

Imperceptible :
contrainte de dégradation

Indélébile :
contrainte de robustesse

Domaine en croissance exponentielle depuis les années 90 du fait du développement des communications numériques.

Quelques applications majeures :

- protection de la propriété intellectuelle des documents numériques
- authentification
- documents enrichis
- correction automatique d'erreurs de transmission (self-correcting)
- indexation
- …

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Schéma général d'insertion d'un tatouage

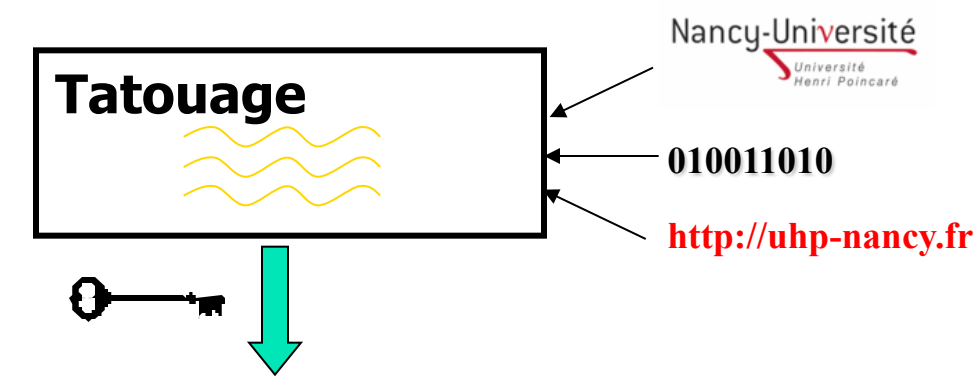


Image originale

Image tatouée

Schéma général d'extraction d'un tatouage

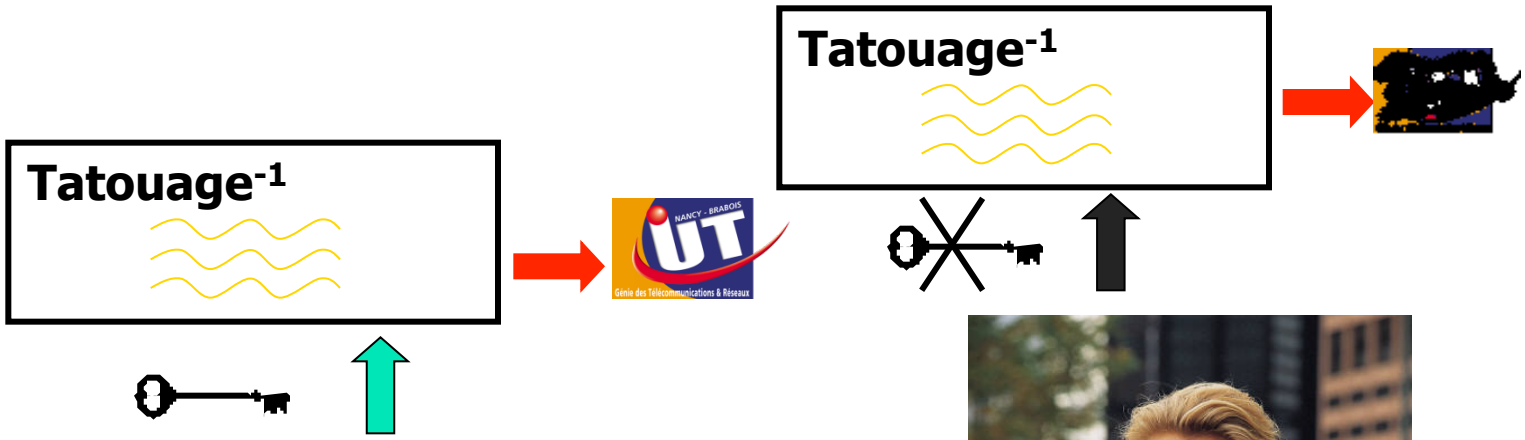


Image tatouée

Image tatouée piratée

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Le tatouage en quelques mots

Tatouage visible : masquage d'un document à l'aide d'une ou plusieurs marques visibles qui sont effaçables correctement que si on possède une clé secrète.

Tatouage fragile : permet de prouver qu'un document n'a pas été falsifié (i.e. n'a pas subi de transformation pouvant modifier son interprétation)

Tatouage semi-fragile : permet de détecter localement des manipulations malveillantes tout en étant robuste à certains traitements (comme par exemple la compression)

Tatouage aveugle : la marque est extraite à l'aide du document tatoué (éventuellement attaqué) seulement

Tatouage semi-aveugle : la marque est extraite à l'aide du document tatoué et de la connaissance de la signature (marque)

Information secrète : le fait que l'algorithme d'insertion et d'extraction n'est pas publique n'est pas suffisant. Il faut une information secrète, généralement la clé.

Image déposée (tatouage visible)

Image enregistrée chez Digimarc

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Image déposée

Tatouage : principaux défis

Principaux défis théoriques du tatouage :

- ✓ **Capacité d'insertion :** *de quelques dizaines de bits à plusieurs kilobits selon l'application*
- ✓ **Invisibilité :** *cacher le message sans gêner le confort visuel ni l'interprétation sémantique*
- ✓ **Robustesse :** *face à des traitements bienveillants ou malveillants (attaques) du signal tatoué**
- ✓ **Sécurité :** *liée aux attaques exploitant une faille de l'algorithme lui-même***

* Objectif de l'attaque: faire disparaître le tatouage

** Objectif de l'attaque: accéder à un secret pour ensuite faire disparaître le tatouage de manière « chirurgicale » ou accéder à des informations confidentielles ou encore usurper une identité et s'en servir pour tatouer un document

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Les attaques

2 types d'attaque : malveillantes ou traitements courants

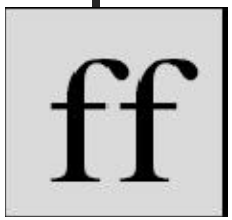
- bruitage de l'image
- transformation géométrique (décalage, rotation, zoom,…)
- filtrage linéaire (passe-bas, passe-haut, passe-bande) ou non linéaire (médian)
- réhaussement de contraste
- compression avec perte
- conversion de format (ex: JPEG vers GIF)
- composition d'images, mosaïque
- …

Logiciels libres pour tester une méthode de tatouage :

- Stirmark (http://www.petitcolas.net/fabien/watermarking/stirmark/)
- Unzign (adresse non disponible)

La quasi-totalité des systèmes de tatouage peut se faire piéger (Stirmark et Unzign)

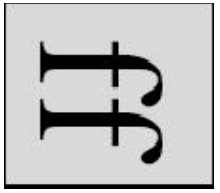
Les attaques : exemples



original

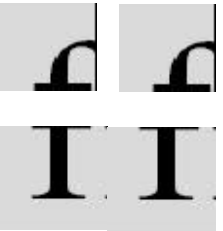
découpage

zoom



rotation

...

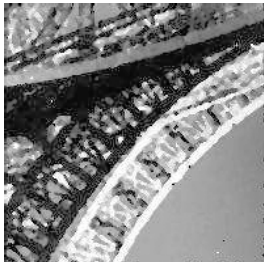


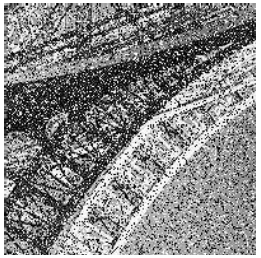
mosaïque

original

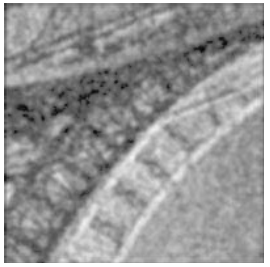
Bruit gaussien

Filtrage non linéaire





Bruit sel et poivre



Filtrage linéaire

original

JPEG (80:1)

JPEG2000 (80:1)

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Principes du tatouage

Tatouer = *insérer une marque contenant ou non de l'information*

marque = *quelques bits à quelques centaines de bits* 10011110101…

2 actions

- ***insertion***
- ***lecture***

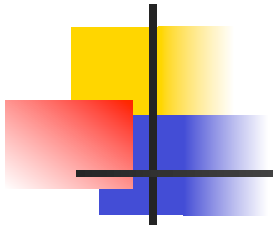
3 propriétés

- ***spécificité***
- ***invisibilité***
- ***robustesse aux attaques***

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Classification des méthodes

- *Domaine initial / domaine transformé*
- *Additive / substitutive*
- *Fondée sur le contenu / de communication*



Tatouage d'images

DOMAINE SPATIAL

L'algorithme du Patchwork (Bender *et al* en 1995)

2 patches A et B de même taille (n pixels) choisis aléatoirement dans l'image (clé)

Règle de tatouage :

paire de pixels $(a_i, b_i) \Rightarrow (a'_i, b'_i)$

$$a'_i = a_i + 1$$
$$b'_i = b_i - 1$$

Extraction :

On calcule : $$S' = \sum_{i=1}^{n} (a'_i - b'_i) = \sum_{i=1}^{n} (a_i + 1 - b_i + 1) = 2n$$

Or on sait que statistiquement sur l'image on a pour n suffisamment grand : $S = \sum_{i=1}^{n} (a_i - b_i) \approx 0$

Donc seul un utilisateur possédant la clé peut retrouver $2n$

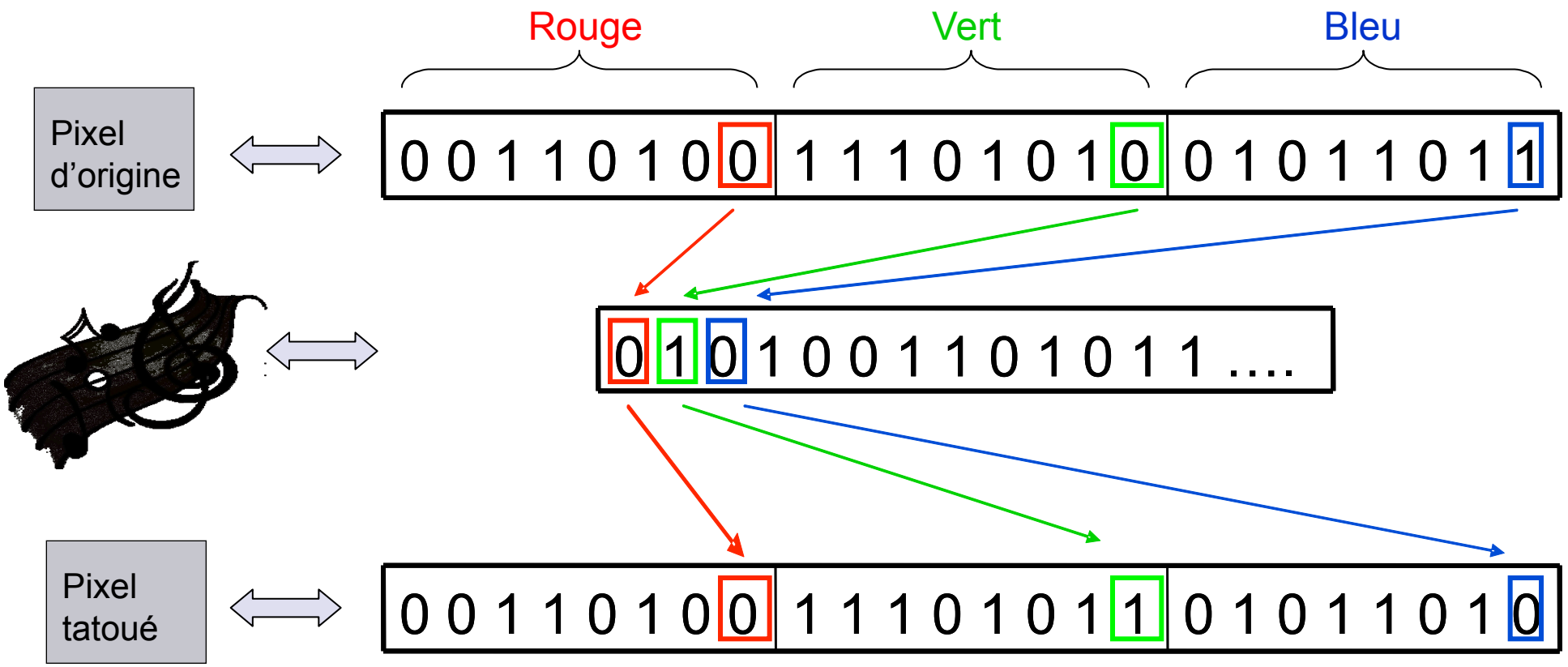
Limites de l'algorithme : Faible robustesse (attaques géométriques, filtrage,…)

Permet juste de répondre à la question : cette personne a-t'elle la clé ?

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Méthode de tatouage : exemple

- Bit de poids faible (LSB pour Least Significant Bit) :



Méthode de tatouage : LSB

Démo LSB

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Tatouage fragile

tatouage de son dans une image

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Tatouage fragile

Image tatouée (25 Kbits insérés)

Extraction du tatouage :

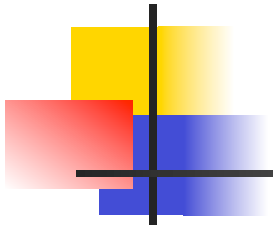
J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Tatouage fragile

Image tatouée <u>piratée</u>

Extraction du tatouage :

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)



Tatouage d'images

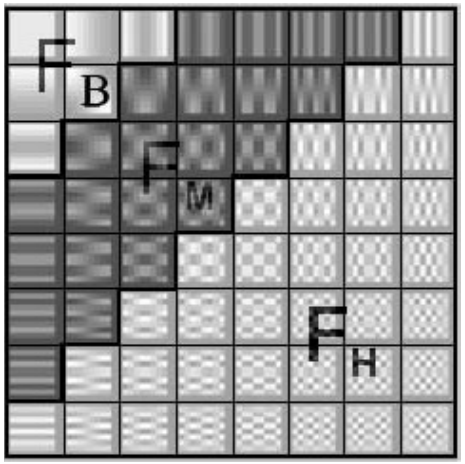
DOMAINE TRANSFORME

Meilleure prise en compte des propriétés psychovisuelles

Robustesse accrue

Algorithme de Koch et Zhao (1994)

Blocs DCT 8x8



Tatouage dans les moyennes fréquences

Fréquences basses (zones homogènes) : robuste mais visible

Fréquences hautes (forts contours) : invisible mais fragile

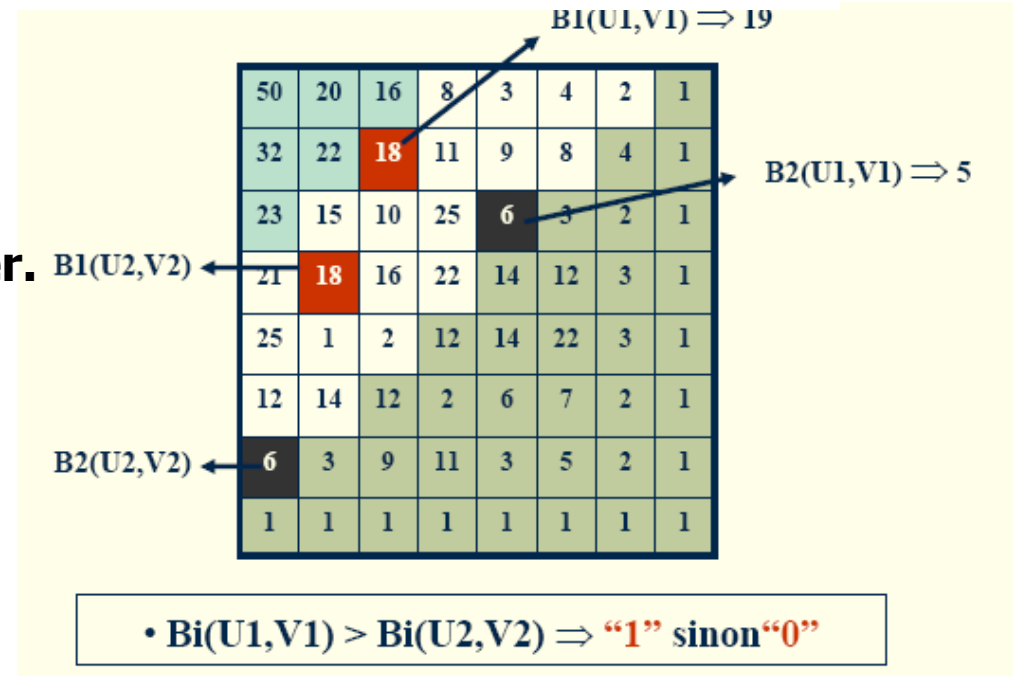
J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Algorithme de Koch et Zhao (1994)

Principes

Blocs DCT 8x8 choisis aléatoirement

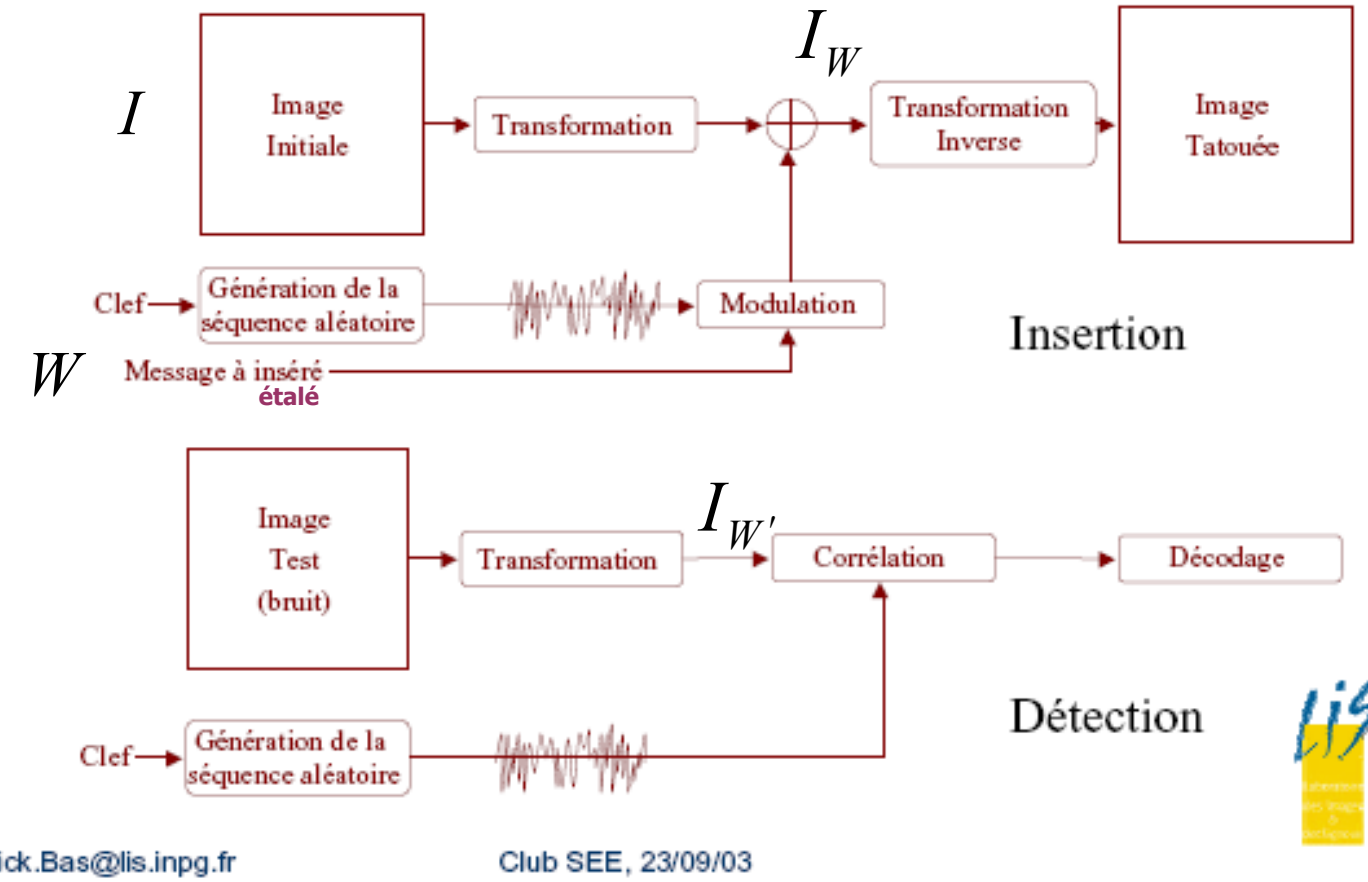
Choisir des zones des blocs fréquentiels avec la **même amplitude de valeur et les modifier.**



Inconvénients :

faible robustesse aux attaques géométriques, faible capacité (1bit/bloc)

Etalement de spectre



Patrick.Bas@lis.inpg.fr

Club SEE, 23/09/03

Etalement de spectre

- Insertion: $I_w = I \pm W,\ W(i,j) = \{-k, +k\}$
- Détection: $\langle I_w, W \rangle = \langle I, W \rangle + \langle W', W \rangle$

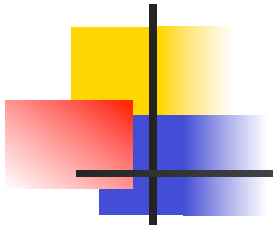
$$\# 0 \pm |W|^2 \text{ si } W' = W$$

$$\# 0 + 0 \text{ si } W' \# W$$

- Le signe de $\langle I_w, W \rangle$ permet de décoder un 0 ou un 1
- La valeur de $\langle I_w, W \rangle$ permet d'attester ou de réfuter la présence du tatouage

Tatouage vs. compression

La compression : une attaque redoutable



Tatouage vs. compression

La compression : une attaque redoutable

Objectif de la compression :

faire disparaître l'information inutile à l'œil (invisible) pour réduire la quantité de données

Objectif du tatouage :

Insérer une information invisible

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

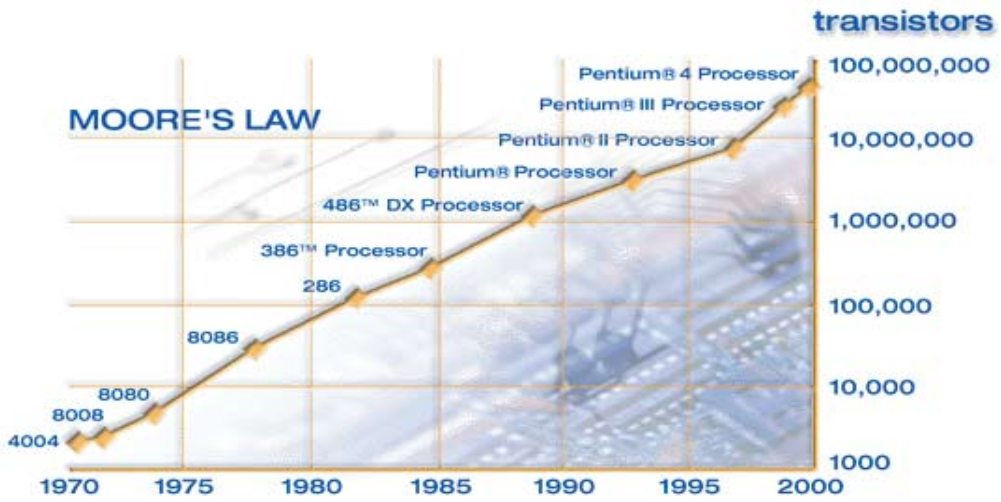
Tatouage vs. compression

La compression : une attaque redoutable

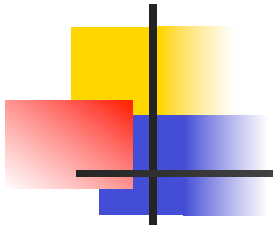
Peut-on éviter la compression ?

La compression et les lois

Cyril Northcote Parkinson a établi que **les volumes de données augmenteraient toujours jusqu'à remplir l'espace de stockage disponible**.
Or la loi de Moore nous permet de savoir que l'espace de stockage et la capacité de traitement des données stockées doublent tous les 18 mois. Les experts de l'industrie prévoient donc que, **d'ici à la fin du 21e siècle, chaque personne sur terre disposera d'un téraoctet de données stockées**. Parkinson est également connu pour sa loi sur l'absorption de la bande passante : « **Le trafic réseau augmente jusqu'à occuper la largeur de bande passante disponible** »



Source : http://www.siemon.com/fr/white_papers/10GAssurance.asp



Tatouage vs. compression

La compression : une attaque redoutable

On peut donc difficilement éviter la compression.

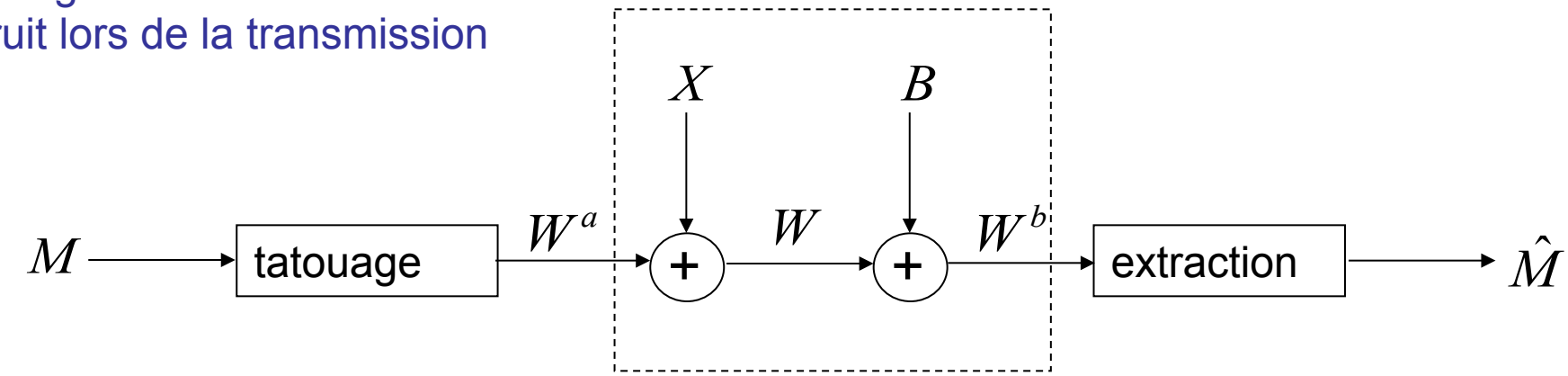
J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Tatouage *ET* compression (1)

Différence majeure avec le tatouage

Le signal hôte est totalement connu lors de l'insertion

Le signal hôte vu comme un bruit lors de la transmission



Premières méthodes issues des techniques d'étalement de spectre

Tatouage *ET* compression (2)

Méthodes actuelles : prise en compte du signal hôte

- Exploiter les caractéristiques (ex: propriétés de masquage dans les images)
- Codage canal avec information adjacente au codeur

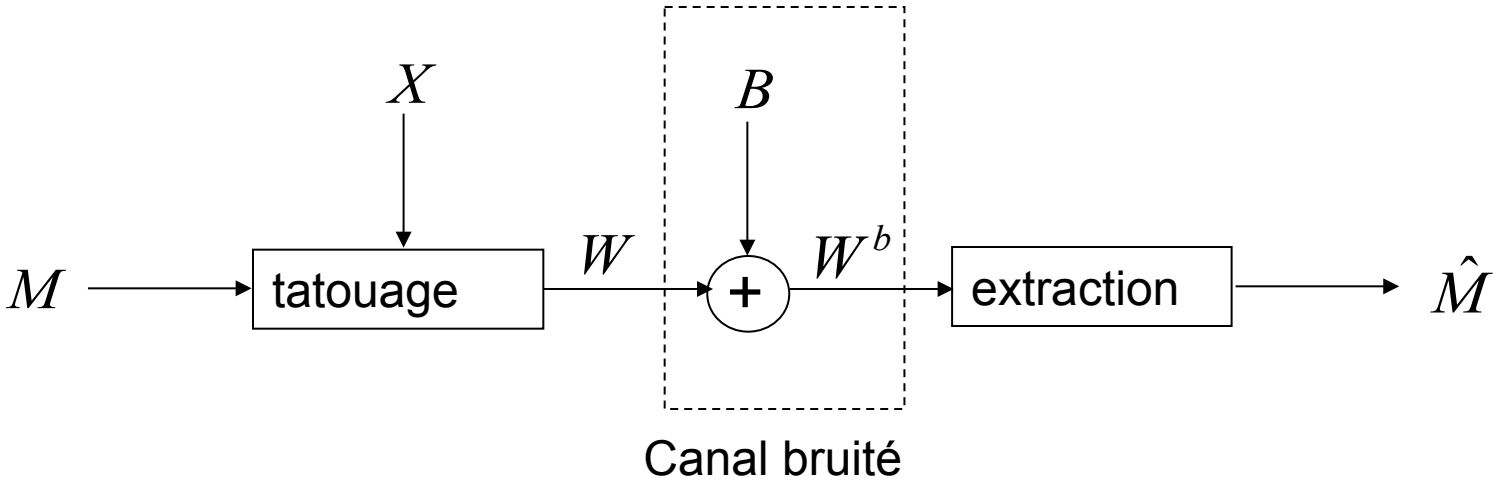


Schéma de transmission avec information adjacente au codeur

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Tatouage/compression conjoints

Image originale

Transformation → Quantification/ tatouage → Codage entropique → Image tatouée & compressée

Message : 0100110…

2 méthodes proposées :

- ✓ Méthode « fondée sur le contenu »
- ✓ QVA modulée (QVAM)

Image reconstruite

Image tatouée & compressée → décodage → Reconstruction → Transformation inverse

Reconstruction → Extraction → Message : 0100110…

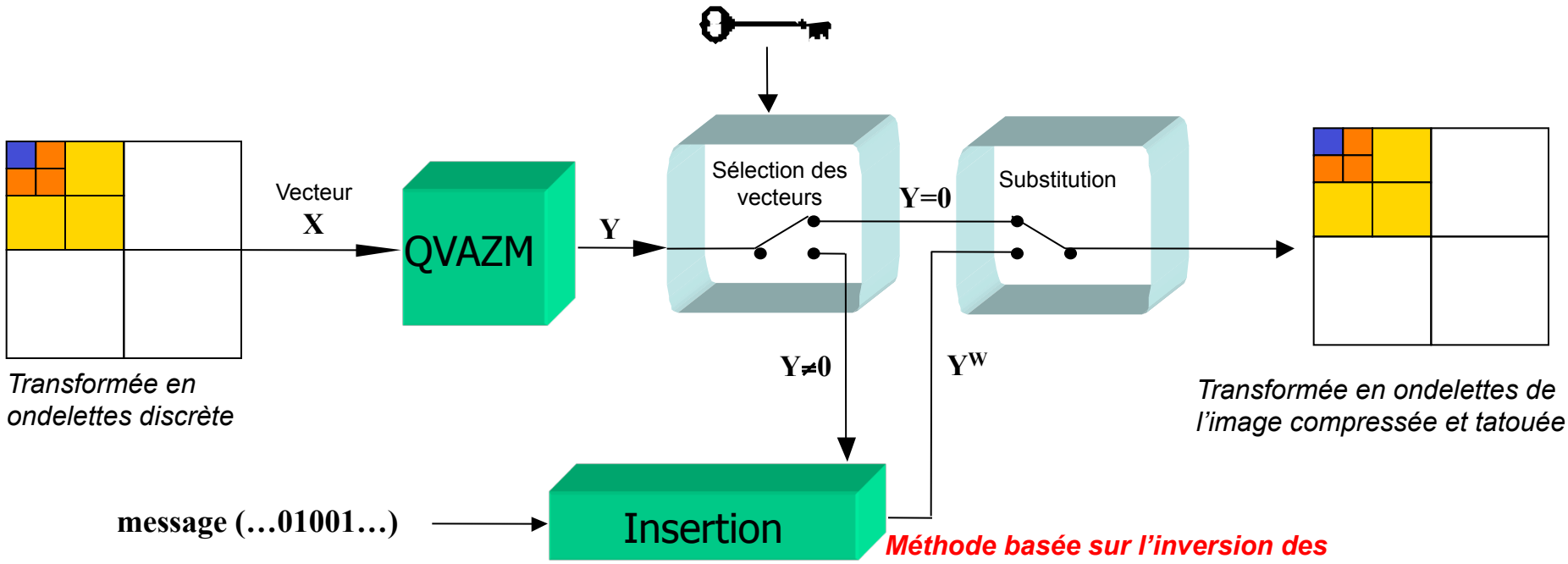
J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Tatouage/compression conjoints

Compression tatouage conjoints : 1ère approche (fondée sur le contenu)

Méthode d'insertion ne modifiant pas l'énergie des vecteurs quantifiés

Utilisation d'un seul dictionnaire



Méthode basée sur l'inversion des 2 composantes les plus proches

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Où se cache le tatouage ?

Message de **60 bits** inséré dans l'une des deux images

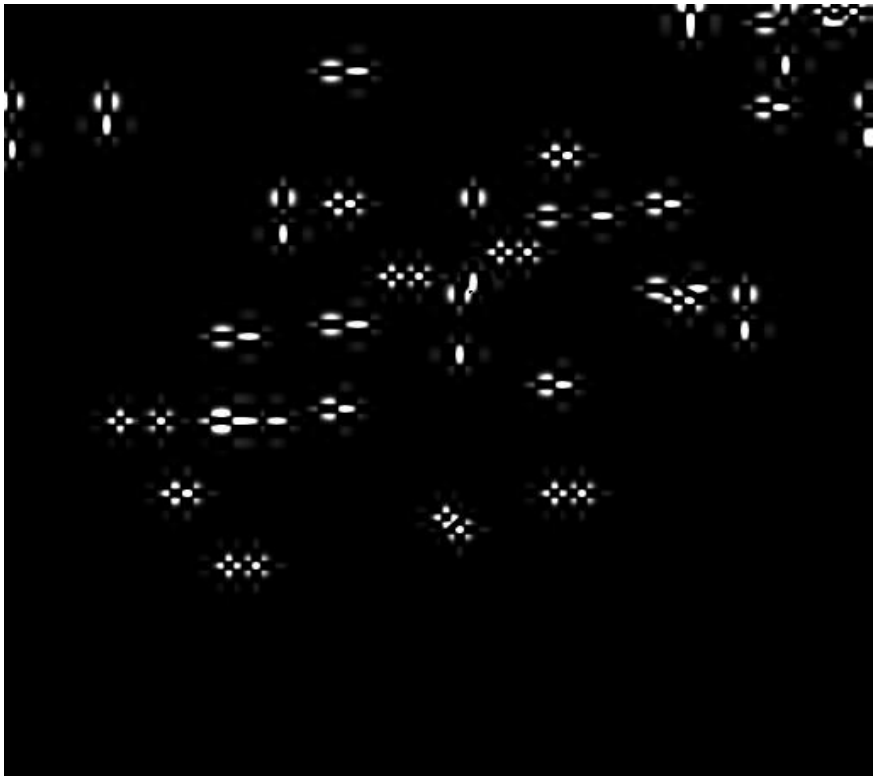
A

B

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Où se cache le tatouage ?

Réponse A !



B - A = localisation du tatouage

A : image tatouée

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Où se cache le tatouage ?

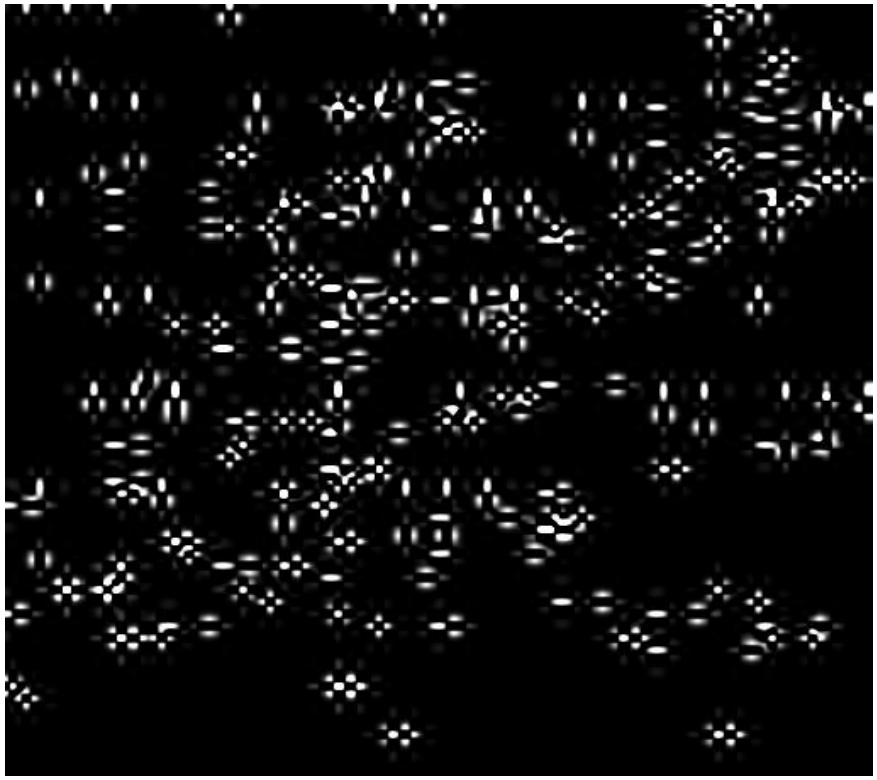
Message de **300 bits** inséré dans l'une des deux images

A

B

Où se cache le tatouage ?

Réponse B !



A - B = localisation du tatouage

B : image tatouée

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Tatouage/compression conjoints

1ère approche : Utilisation d'un seul dictionnaire

Avantages : méthode aveugle, robustesse à la compression, au filtrage et au bruit

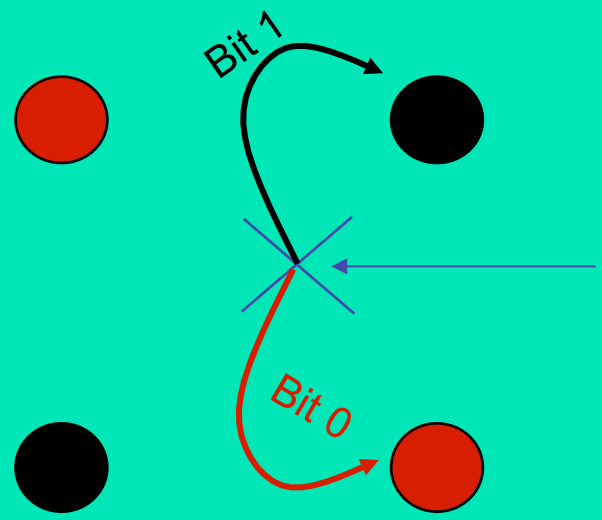
Inconvénients : sensibilité aux attaques géométriques, capacité limitée

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

2ème approche : QVA Modulée

Compression tatouage conjoints : 2ème approche

- Vecteurs du dictionnaire associés au bit 0
- Vecteurs du dictionnaire associés au bit 1



Insertion d'un message binaire par QIM*

QIM = partition du dictionnaire en **m sous-dictionnaires**

⇒ insertion **d'un message m-aire**

*QIM = quantification par modulation d'index (*Chen et Wornel*)

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

2ème approche : QVA Modulée

Insertion/quantification :

$$Q_i(X) \triangleq mQ\left(\frac{X - i\frac{\gamma}{m}}{\gamma}\right) + i$$

Sous-ensemble correspondant au i^{ème} mot du code :

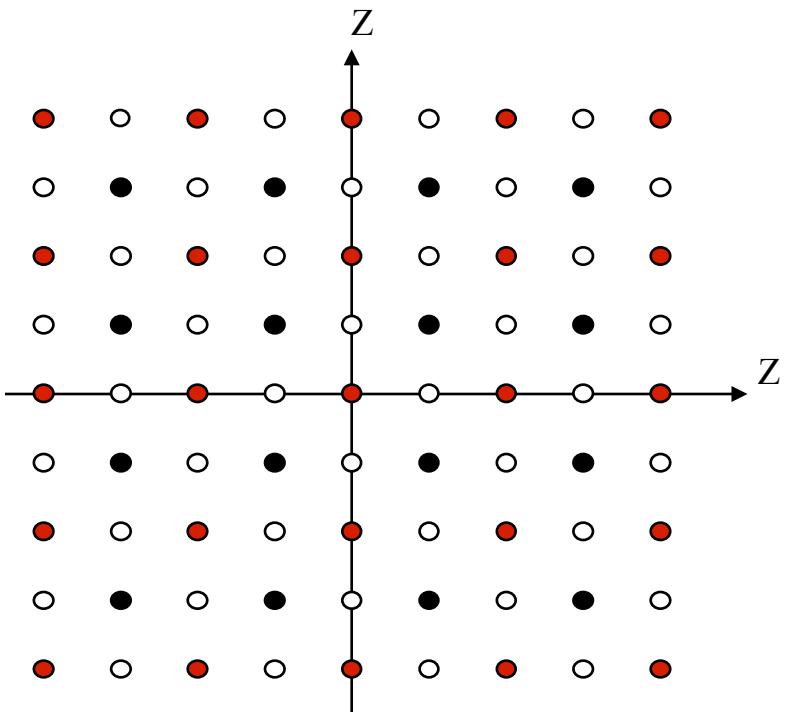
$$S_i \triangleq \{m\mathbb{Z}^n + [i]\}$$

Réseau modulé :

$$\mathbb{Z}_m^n \triangleq \bigcup_{i=0}^{m-1} S_i$$

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

2ème approche : QVA Modulée



QIM à partir de deux quantificateurs décalés : point de vue QVA

- Sous-réseau 0
- Sous-réseau 1
- Vecteurs non utilisés

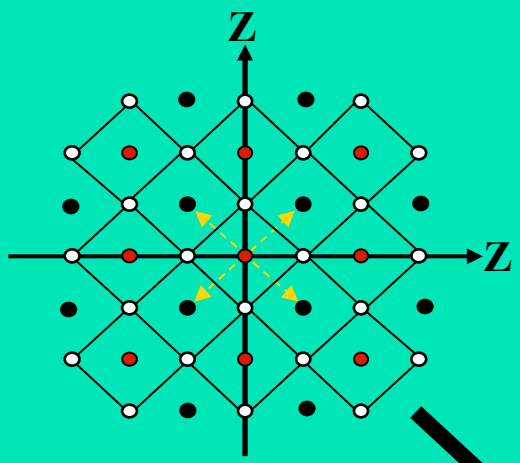
2 problèmes :

- Exploiter la concentration de vecteurs nuls?
- Présence de vecteurs inutiles

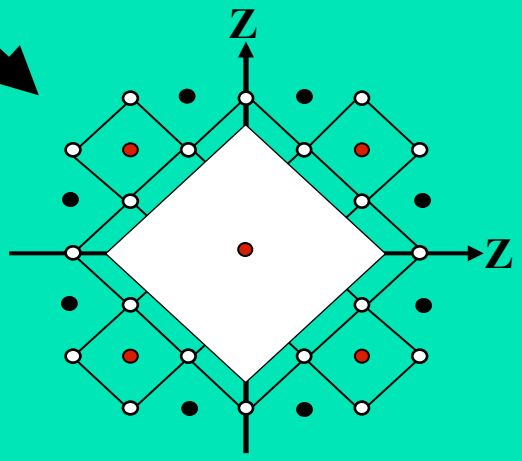
J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

2ème approche : QVA Modulée

Quantification des vecteurs de faible énergie



Approche QIM



Approche QVAM*

Problème N°1 :

Exploitation de la concentration de vecteurs de faible énergie entravée par la modulation

Exclusion des vecteurs de faible énergie de l'insertion

Zone morte vectorielle

*QVAM = quantification vectorielle algébrique modulée

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

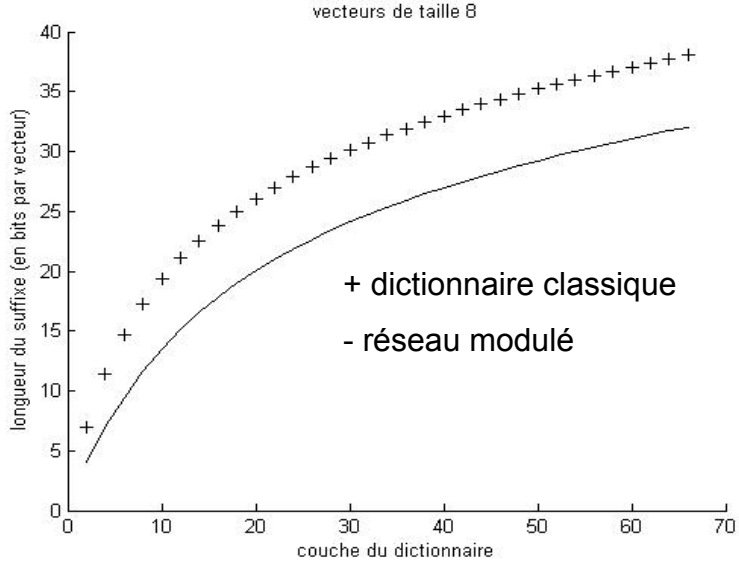
2ème approche : QVA Modulée

Problème N°2 :

Proportion des vecteurs n'appartenant pas au réseau modulé :

99,6% des vecteurs de Z^8 n'appartiennent pas au réseau modulé**!**

Indexage sur le réseau modulé ⇒ diminution de la longueur moyenne du code du suffixe



Longueur des mots du code du suffixe

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Résultat QVAM

Taux de compression de 22:1

Approche directe

PSNR = 16,8 dB; 8,6 kbits insérés

QVAM + zone morte et indexage modulé

PSNR = 33,1 dB; 3,8 kbits insérés

Tatouage/compression conjoints

2ème approche : Utilisation de 2 dictionnaires (QVAM)

Avantages : méthode aveugle, robustesse à la compression, au filtrage et au bruit, forte capacité

Inconvénients : sensibilité aux attaques géométriques

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Tatouage vidéo

Caractéristiques essentielles

• bande passante de dissimulation plus grande que pour les images fixes
(Attention : elle n'est pas égale au nombre d'images multiplié par la bande passante de chacune d'elles)

• contrainte de temps réel cruciale (algorithme peu complexe)

• attaques beaucoup plus difficiles que pour les images fixes

• les méthodes de tatouage s'appliquent souvent à des flux compressés

Quelques applications

• **protection des droits d'auteur (anti-copie DVD, cinéma numérique,…)**

• **transport de métadonnées** 

• **authentification et intégrité des vidéos (vidéosurveillance, …)**

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Tatouage vidéo

Quelques méthodes sur flux non compressé

- **spatiales** : méthode dérivée de celle de Koch et Zhao (coefficients de la DCT), …
- **spatio-temporelles** : méthode de Swanson basée sur les ondelettes (temporel) et la DCT (spatial), …
- **temporelles** : utilisées pour la protection du cinéma numérique (tatouage = modification des très basses fréquences spatiales de chaque image qui entraîne une grande dégradation de la vidéo récupérée)

Quelques méthodes sur flux compressé

- modification des vecteurs mouvement (composantes vx et vy paires si bit 0 impaires sinon)
- modification de la structure du GOP (images P = bit 0, images B = bit 1)

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Tatouage vidéo : démonstration

démo

Domaine de la santé : masse de données multi-sources et multi-formats

- Quelques chiffres :
  - Service d'imagerie du CHU de Nancy : 49 To de données générées en 2009 (croissance annuelle : 10%)
  - Mammographies aux USA en 2009 : 2,5 Po !
  - Archivage d'images médicales en 2010 : 30% de la capacité mondiale !

Problèmes : transport, archivage, manipulation, **protection**

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Domaine de la santé : masse de données multi-sources et multi-formats

- Quelques exemples :
  - Données de type texte (historique patients, compte-rendus, résultats d'analyse, …)
  - Données de type signal (électrocardiogrammes, parole, …)
  - Données de type image 2D ou 3D (radiographies, scanners, IRMs, petscans, puces ADN, …)
  - Données de type vidéo 2D+T, 3D+t (caméras endoscopiques, robots de chirurgie, …)

Problèmes : performances de compression, codages multiples, fusion, **protection**

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Domaine de la santé : masse de données multi-sources et multi-formats

- Multitude de formats:
  - Données de type texte : word, pdf, …
  - Données de type signal : ecg, xml, wav, mp3,…
  - Données de type image 2D ou 3D : gif, tiff, png, jpg, jp2, dicom, …
  - Données de type vidéo 2D+T, 3D+t : mp2, mp4,…
  - HL7 : spécifications techniques pour les échanges informatisés de données cliniques, financières et administratives entre systèmes d'information hospitaliers (SIH)

Problèmes : compatibilité avec les lecteurs, intégration

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Quelques défis du « big data » en santé

- **Caractériser** les multitudes de données en informations sémantiques :

  - réduire la quantité de données pour les rendre accessibles (parcours d'image volumique, …)

  - obtenir une information pertinente pour l'aide à la décision (diagnostic, thérapie, chirurgie, …)

- **Archiver** ces données
- **Transmettre** ces données
- **Protéger**, **contrôler** et/ou **enrichir** ces données

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

L'imagerie radiologique

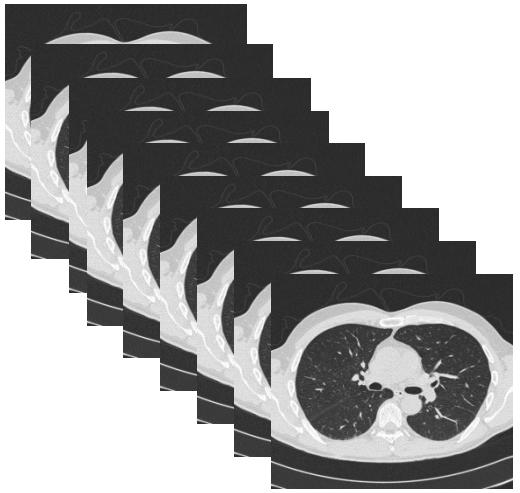
Une imagerie radiologique de plus en plus précise …

… donc très lourde à manipuler, stocker, transmettre.

Hier (et encore aujourd'hui)

Radiographie des poumons
(source : http://stsp.creteil.iufm.fr/article29.html)

Aujourd'hui

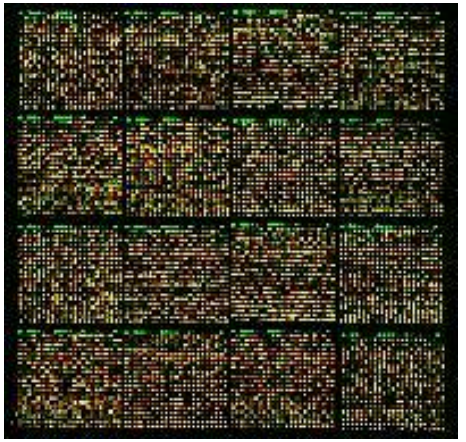


Scanner de poumons

Un examen, c'est environ 200 Moctets à stocker (ou à transmettre) !

L'imagerie biologique

Cas des puces ADN



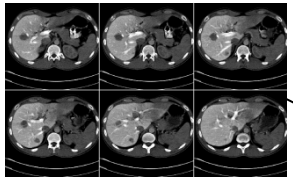
- Données de puces à ADN : 33000 gènes
- données extrêmement volumineuses pour les systèmes
- indispensable pour améliorer la recherche sur le traitement des maladies (des maladies orphelines aux maladies les plus courantes)
- Taille : taille puces adn
- Nécessité de comparer différentes images entre elles

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

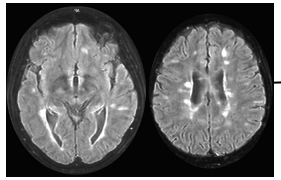
Contexte de l'imagerie radiologique

Images de plus en plus précises

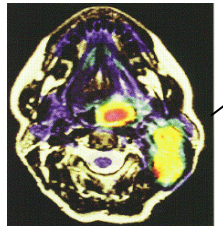
Scanner



IRM



PET



⋮

Images DICOM

PACS

Diagnostic

Traitements

Consultations postérieures

PACS : Picture archiving communication system.

→ : Echanges de données numériques

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Estimation de la qualité

- Problème ouvert et complexe pour la qualité diagnostique
- 2 façon d'évaluer la qualité après un traitement (compression, tatouage, etc) :
  - ✓ Subjective (tests avec un panel d'experts)
  - ✓ Objective (ensemble de métriques)

Evaluation subjective de la qualité

- Un groupe d'experts (au minimum 3 « séniors ») ou 16 (incluant juniors et séniors)
- Au minimum 30 images
- Protocole de tests rigoureux à construire (type de pathologie, type d'images, question posée, nombre de patients, diversité des patients, …)
- 2 « types » de tests :
  - ✓ Détection (binaire : présence ou absence d'une pathologie)
  - ✓ Estimation (qualité globale) qui conduit à une note (généralement entre 1 et 5)

Test type détection (1)

		Patients	
		Réellement pathologique	Réellement normal
Réponses observateur	pathologique	VP	FP
	normal	FN	VN

- VP: vrais positifs, VN : vrais négatifs, FP : faux positifs, FN : faux négatifs
- Nécessité d'un gold standard (test qui indique de manière formelle le diagnostic du patient inclus dans l'expérimentation)
- Sensibilité, Spécificité, Valeur Prédictive Positive, Valeur Prédictive Négative :

$$Se = \frac{VP}{VP + FN}, \qquad Sp = \frac{VN}{VN + FP}$$

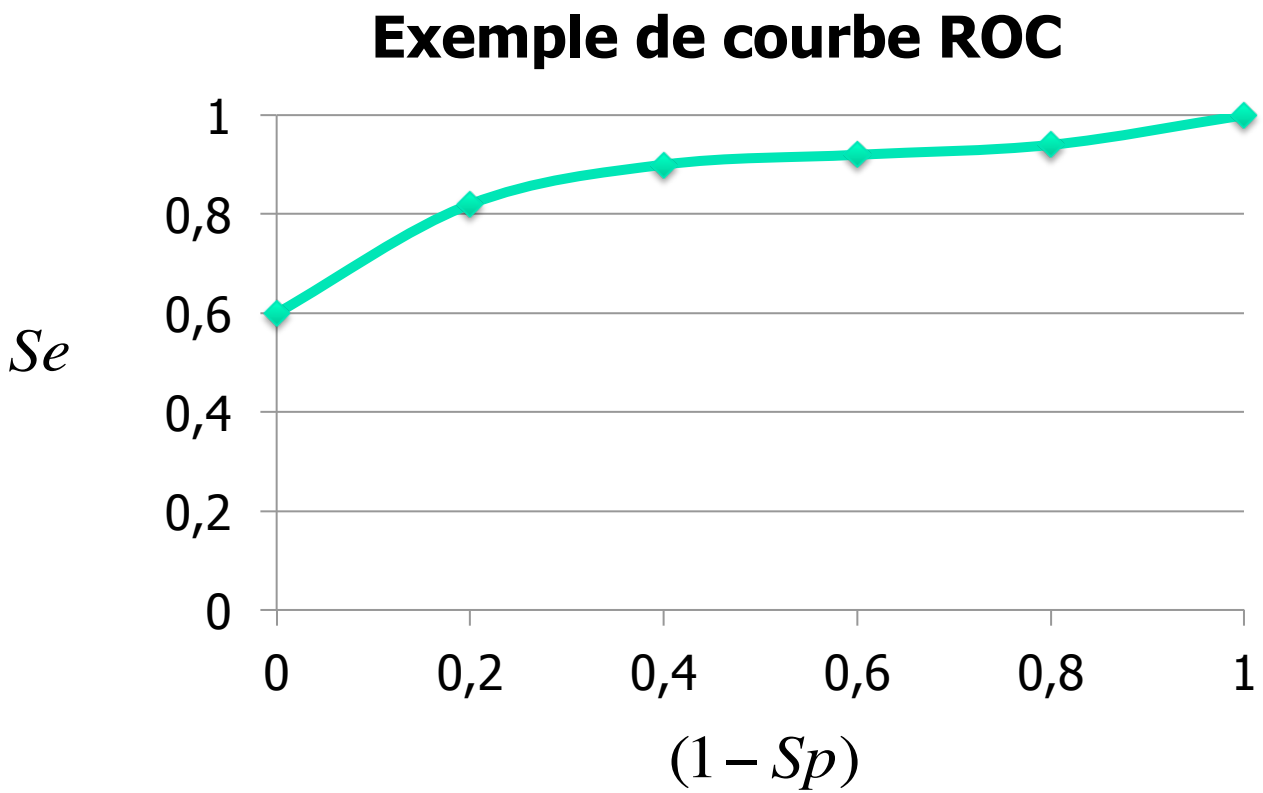
$$VPP = \frac{VP}{VP + FP}, \qquad VPN = \frac{VN}{VN + FN}$$

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Test type détection (2)

- Difficulté : sensibilité dépend du seuil de décision que se fixe le médecin et de la « subtilité » de la pathologie
- La méthodologie ROC (Receiver Operating Characteristics) permet de prendre en compte cette difficulté :

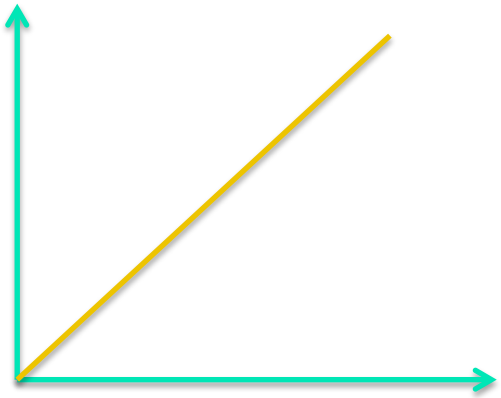
$$ROC : \ Se = f(1 - Sp)$$



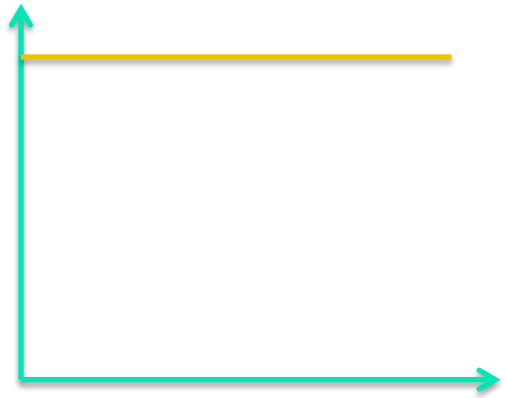
J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Test type détection (3)

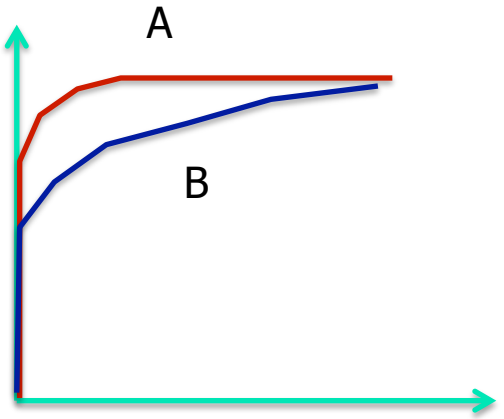
- Méthode ROC : méthode de référence dans la communauté radiologique
- Interprétation :



non discriminant



discriminant (il existe un seuil de décision qui conditionne la détection)



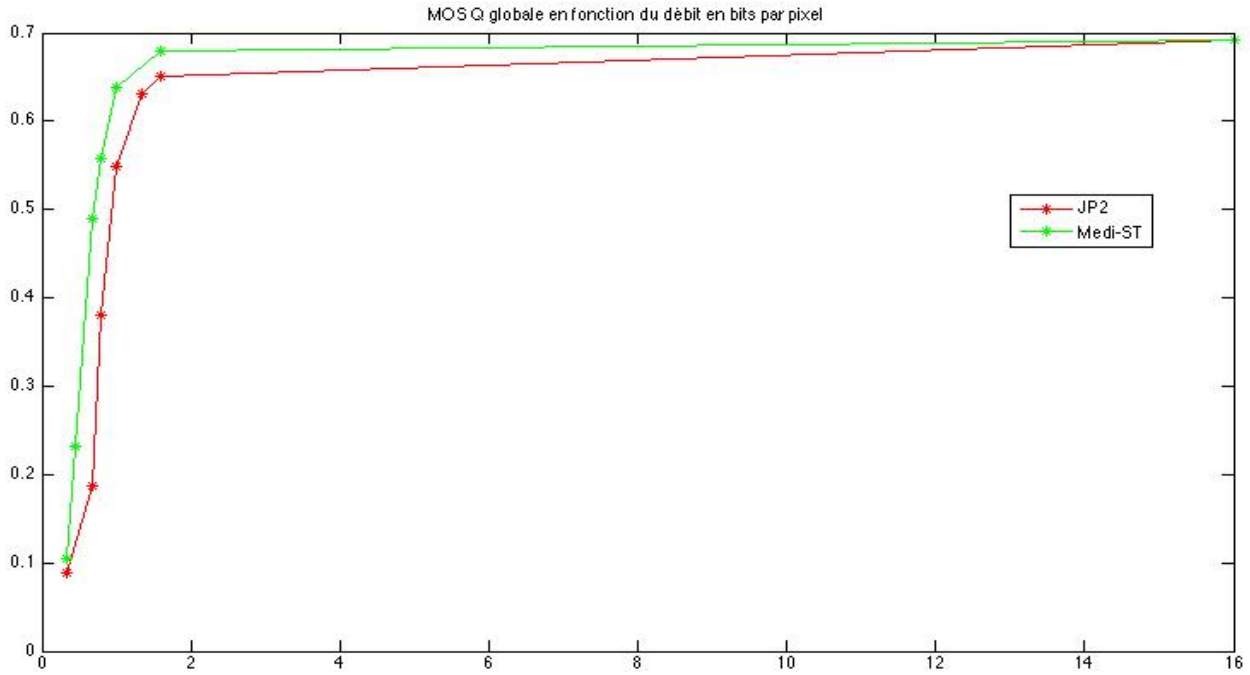
« classique » (A > B en termes de détection)

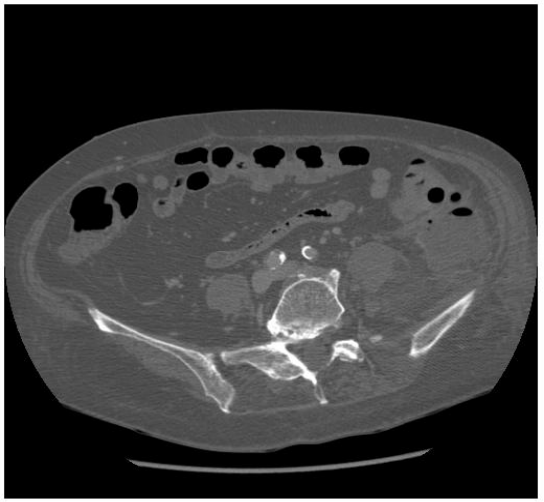
J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Test type estimation (1)

- Analyse de nature globale
- Définition de critères diagnostiques à évaluer et d'une échelle de notation
- Echelle continue vs. Échelle discrète ?
- Protocole d'évaluation strict
- Simple stimulus (images fixes) vs double stimulus (vidéo)
- MOS : Mean Opinion Score

Test type estimation (2)





J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Conclusion tatouage

- Tatouage **fragile** : authentification
- Tatouage **robuste** : protection des droits d'auteur, …

Quelques enjeux essentiels

Protection de la propriété intellectuelle des données numériques

Méta-documents (images « intelligentes », commerce électronique, …)

Authentification de documents

JPEG2000, MPEG4 et DVD font apparaître le « watermarking »

Une multitude de nouvelles applications : web spider, …

J-M. Moureaux – TELECOM Nancy – Université de Lorraine- (jean-marie.moureaux@univ-lorraine.fr)

Bibliographie

- F. Davoine, S. Pateux, « Tatouage de documents audiovisuels numériques », Traité IC2, Editions Hermès Lavoisier, 2004.
- J-L Dugeley et S. Roche, « Introduction au tatouage d'images », *http://www.eurecom.fr/~image*
- P. Bas, « Compression d'Images Fixes et de Séquences Vidéo», cours ENSERG/INPG, LIS Grenoble, Patrick.Bas@inpg.fr
- La cryptographie expliquée : http://www.bibmath.net/crypto/plan.php3
- Stirmark : http://www.petitcolas.net/fabien/kerckhoffs/la_cryptographie_militaire_i.htm#desiderata
- http://www.i3s.unice.fr/~crescenz/publications/watermarking-linfo-diaporama-2004-06.pdf
- http://www.yuvsoft.com/