

# Fingerprint et résistance à la collusion



30/11/2005

Stanislas Francfort

Le présent document contient des informations qui sont la propriété de France Télécom. L'acceptation de ce document par son destinataire implique, de la part de ce dernier, la reconnaissance du caractère confidentiel de son contenu et l'engagement de n'en faire aucune reproduction, aucune transmission à des tiers, aucune divulgation et aucune utilisation commerciale sans l'accord préalable écrit de Recherche & Développement de France Télécom.



# Principes

# Définitions



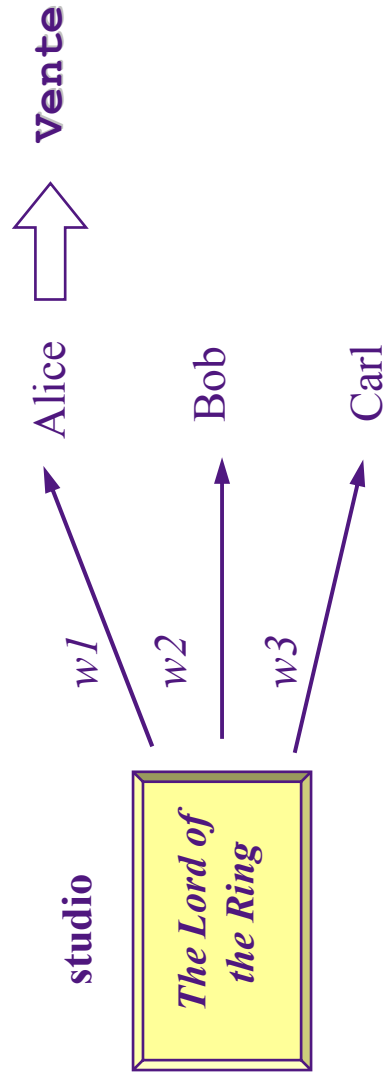
- ▶ **Un *Fingerprint* est la caractéristique d'un objet qui peut être utilisée pour le distinguer parmi des objets similaires**
  - ▶ Empreintes digitales, marques sur une balle tirée par une arme à feu
- ▶ ***Fingerprinter* un objet, c'est lui ajouter un Fingerprint**
  - ▶ Tables de logarithmes, cartes géographiques avec des petites erreurs délibérées, documents de Mme Thatcher
- ▶ **La croissance des échanges de contenus numériques a créé le besoin de nouveaux modèles de Fingerprints**

# Fingerprint

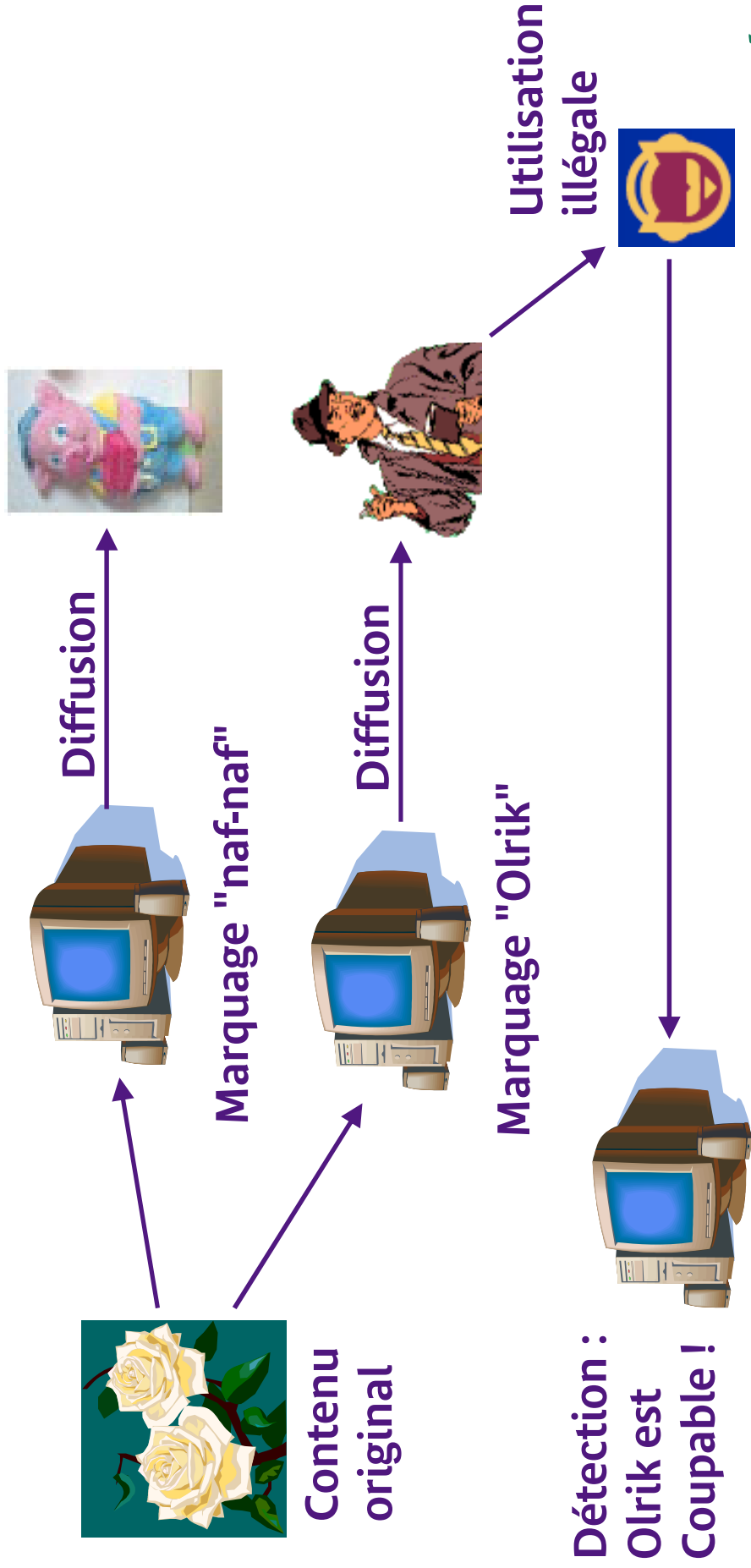


## Contre mesure afin de mettre en échec le piratage multimédia

- Insertion d'une marque ou "Fingerprint" identifiant l'utilisateur légitime (pour chaque utilisateur)
- Contraintes : imperceptibilité, robustesse, capacité d'identification



# Fingerprint



# Fingerprint numérique



- ▶ ***Une marque est la position dans un contenu qui peut être dans l'un des états possibles (dont le nombre est fixé) (Boneh and Shaw)***
  - C'est un mot de code choisi dans un alphabet
- ▶ **Un Fingerprint est une collection de marques**
- ▶ **Fingerprinter :**
  - Comment marquer un contenu ?
  - Comment utiliser les marques pour créer un Fingerprint ?

# Fingerprint numérique



▶ Le Fingerprint n'a pas pour but de prévenir la distribution illégale, mais agit a posteriori en aidant à déterminer qui est la source de la copie illégale

- ▶ **Pirate** : utilisateur légitime qui redistribue illégalement le contenu acquis
- ▶ **Tracer un pirate** : identifier le pirate en se basant sur le contenu redistribué illégalement

# Le Fingerprint numérique existe déjà



- ▶ Une source de redistribution pirate : les VHS et les DVD distribués aux 5803 membres éligibles de la MPAA, votant aux nominations des Oscars
- ▶ Le système de tatouage de Thomson a permis de leur distribuer des supports individuellement tatoués
- ▶ Les films suivants ont été trouvés sur Internet :
  - The Last Samurai
  - Something's Gotta Give
  - Mystic River



# Le Fingerprint numérique existe déjà

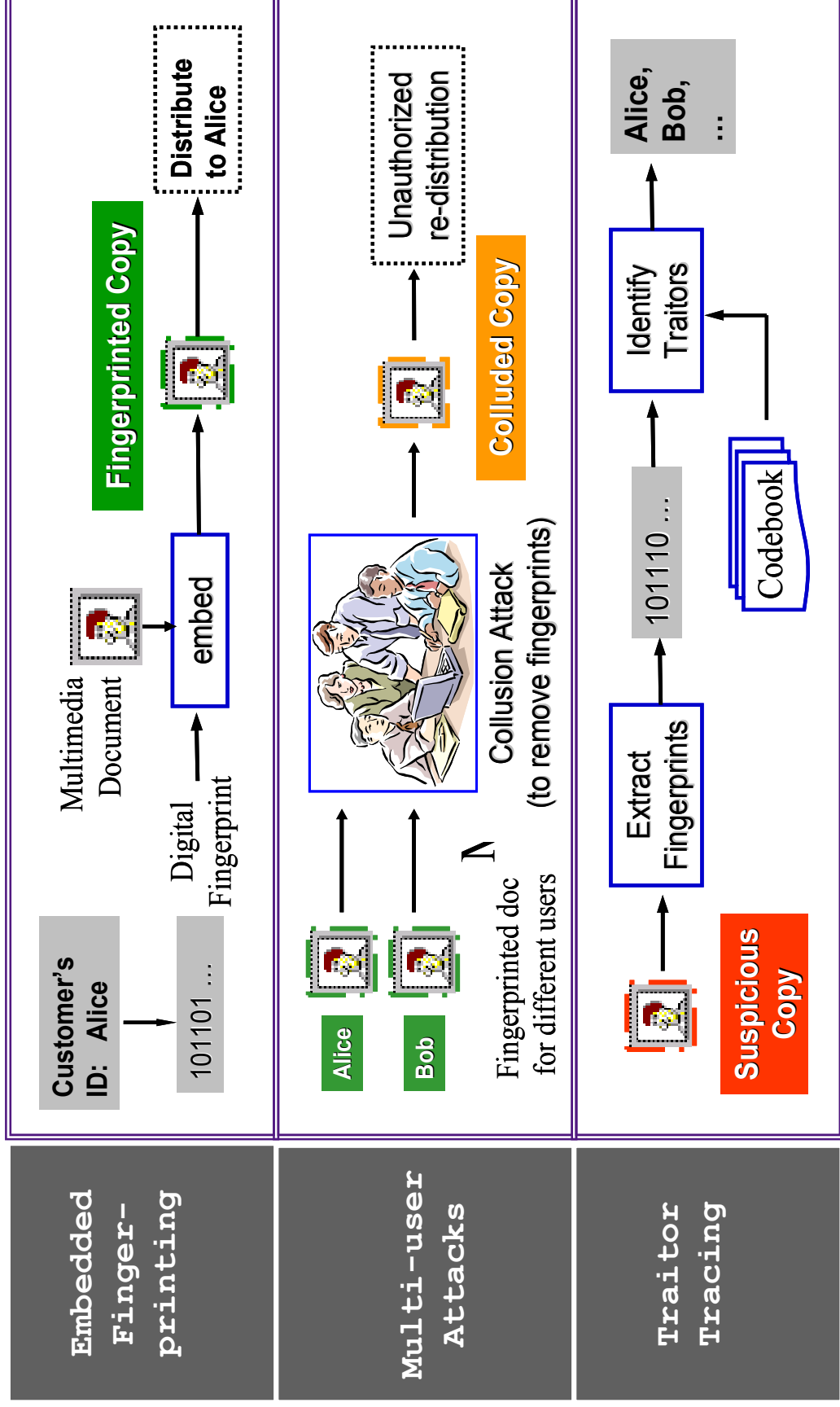
- ▶ L'acteur Carmine Caridi a été identifié comme la source de diffusion illégale
- ▶ Il avait donné la copie à Russell Sprague qui l'a diffusé sur Internet
- ▶ Exclusion de la MPAA
- ▶ 300.000\$ de dommages et intérêts en Novembre 2004





# Collusions

# Collusion

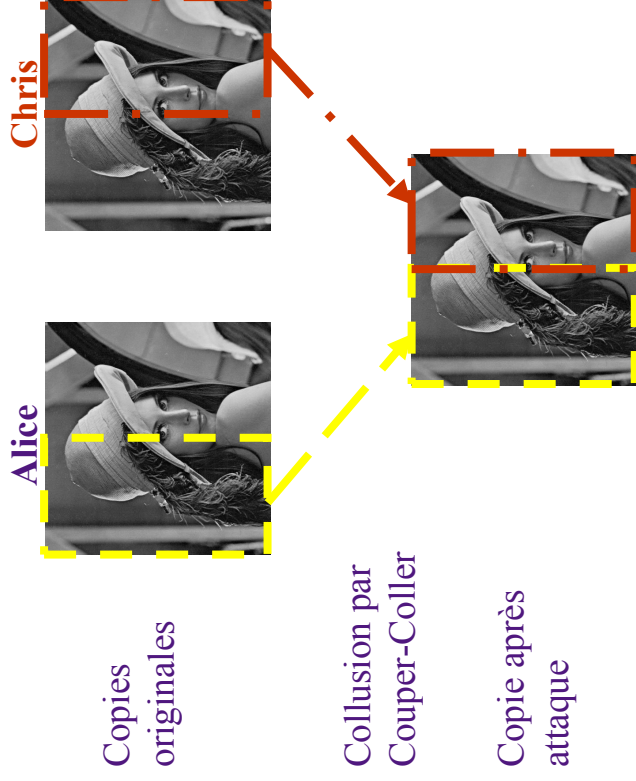
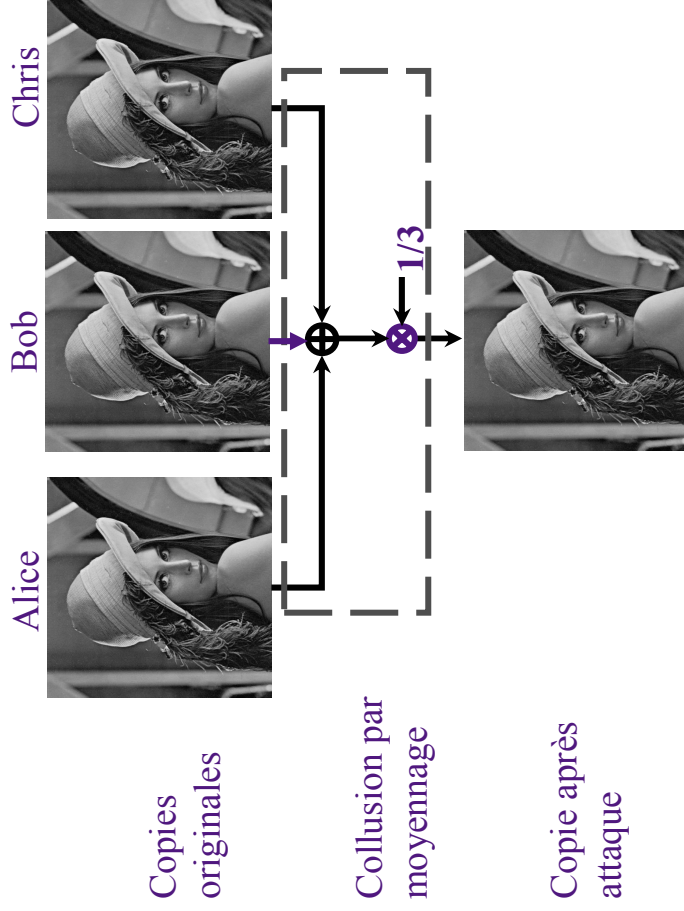


# Définitions

- ▶ Une *collusion* est un groupe d'utilisateur réunissant leurs contenus respectifs afin d'effacer leurs marques
- ▶ Par extension, la transformation que la collusion effectue à partir de leurs contenus s'appelle également une collusion



# Attaques par collusion



# Tatouage – résistance à la collusion



## ▶ Schémas combinatoires

- ▶ Identifie les pirates ayant participé à la collusion à partir des marques (partie de Fingerprint) leur étant propres
- ▶ Similarité avec les codes correcteurs d'erreurs

## ▶ Schémas orthogonaux

- ▶ Les Fingerprints des différents utilisateurs sont des modulations orthogonales du contenu
- ▶ Le Fingerprint après collusion est dans l'espace engendré par les vecteurs des utilisateurs pirates

## ▶ Schémas de communication synchrone multi-utilisateurs

- ▶ Travaux récents
- ▶ Recherche à minimiser l'interférence des marques



# Marking assumption

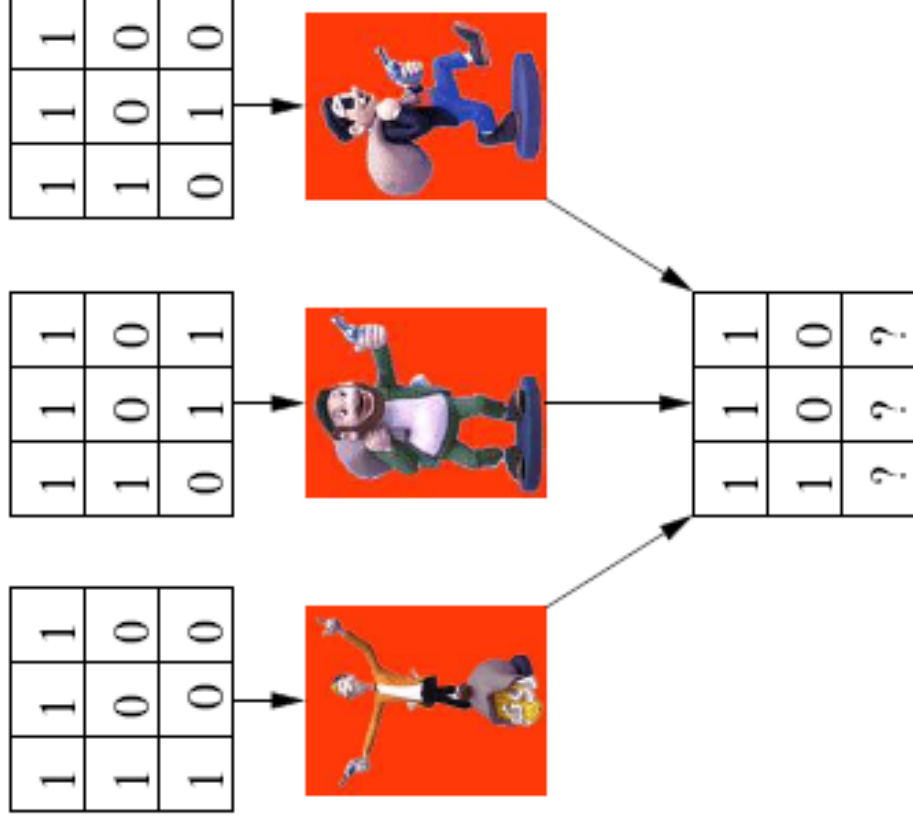
# Marking Assumption



- ▶ **Hypothèse qu'il existe un schéma d'insertion de marque conçu pour résister à la collusion avec les propriétés suivantes :**
  - La collusion peut détecter une marque spécifique si et seulement si leurs marques sont distincts entre elles (sinon, la marque ne peut pas être détectée)
    - Si il n'y a pas de collusion, le Fingerprint se réduit à un numéro de série
  - Les utilisateurs ne peuvent pas changer l'état de leurs marques indétectables sans rendre le contenu inutilisable
- ▶ **Limite l'action des utilisateurs participants à la collusion**



# Marking Assumption



**Les marques identiques sont indétectables**



# La construction de Boneh-Shaw

## ▶ Contenu arbitraire et “Marking assumptions” (1998)

- Une abstraction du modèle de collusion
- Soit un contenu de 6 Bits marqué aux positions 2, 4, et 5 et soit  $m_1$ ,  $m_2$  et  $m_3$  les contenus marqués

$$\begin{array}{l} m_1 = \\ m_2 = \\ m_3 = \end{array} \begin{array}{|c|c|c|c|c|c|} \hline 0 & 1 & 1 & 1 & 0 & 1 \\ \hline 0 & 1 & 1 & 0 & 1 & 1 \\ \hline 0 & 0 & 1 & 1 & 1 & 1 \\ \hline \end{array}$$

- Si  $m_1$ ,  $m_2$  et  $m_3$  effectuent une collusion, alors les positions des marques sont déterminées
- Si  $m_1$  et  $m_2$  effectuent une collusion, alors seulement les marques 4 et 5 peuvent être identifiées



# La construction de Boneh-Shaw

- ▶ **Trace un pirate parmi les membres de la collusion**
  - Un code *totalemment c-secure* : étant donnée une collusion d'au plus  $c$  pirates, une copie illégale peut être tracée vers au moins un pirate de la collusion.
    - Il est prouvé que pour  $c > 1$  il n'existe pas de tels codes
  - Utilisation de techniques probabilistes pour construire des codes  $\epsilon$ -*error c-secure* qui ont la propriété suivante : au moins un pirate de la collusion est tracé, avec une probabilité de  $1 - \epsilon$  pour un petit  $\epsilon$  donné.

# Pas de codes totalement c-secure



- ▶ Soit  $w^{(1)}, w^{(2)}, w^{(3)}$  trois marques distincts
- ▶ Soit la fonction "majorité"  $M = \text{MAJ}(w^{(1)}, w^{(2)}, w^{(3)})$

$$M_i = \begin{cases} w^{(1)}_i & \text{si } w^{(1)}_i = w^{(2)}_i \text{ ou } w^{(1)}_i = w^{(3)}_i \\ w^{(2)}_i & \text{si } w^{(2)}_i = w^{(3)}_i \\ ? & \text{sinon} \end{cases}$$

- ▶ Le mot  $M$  ainsi formé peut être créé par les trois coalitions (1,2), (2,3) et (3,1)
- ▶ Or, leur intersection est vide...



# Codes "Collusion Secure"

- ▶ Génération d'une matrice dont les lignes sont les Fingerprints
- ▶ Matrice "triangulaire", que des 1 au dessus de la diagonale, que de 0 en dessous
  - Ressemble à un escalier, la largeur des marches définit le  $\epsilon$ 
    - $m_1$ : 111111111111
    - $m_2$ : 000111111111
    - $m_3$ : 000000111111
    - $m_4$ : 000000000111
  - Avant d'être utilisé, tous les Fingerprints sont permutés selon une permutation donnée
- ▶ Une collusion ne pourra pas, avec une forte probabilité, générer un mot de code différent de  $m_1$ ,  $m_2$ ,  $m_3$  ou  $m_4$

# Codes "Collusion Secure"



- ▶ Initialement, les Fingerprints sont éloignés les uns des autres (selon la distance de Hamming)
- ▶ Le détecteur décode le Fingerprint ayant subit une collusion comme le Fingerprint le plus proche (dans la matrice initiale)
- ▶ Un  $\varepsilon$  arbitrairement petit implique un code extrêmement long
  - La résistance à la collusion est proportionnelle au cube de la taille de la collusion max. acceptée (cad. Pour capturer un moins un membre de la collusion, la longueur du code doit être de l'ordre de  $O(c^3 \log n)$  où n est le nombre de Fingerprints distribués)
- ▶ Le schéma de Boneh-Shaw est intensivement étudié et la littérature cryptographique propose un grand nombre

**d'améliorations**



# Fingerprinter un contenu multimédia



# Fingerprint et multimédia

- ▶ **Contraintes issues du multimédia**
  - ▶ L'hypothèse “Marking assumptions” est réductrice...
  - ▶ Quelques Bits des mots du code peuvent être mal lues par le détecteur (dépend du type d'insertion dans le contenu)
  - ▶ Il est nécessaire de choisir une insertion qui empêche la collusion de changer arbitrairement les Bits du code

- ▶ **Volonté de tracer des collusions de tailles arbitrairement grandes**

- ▶ **Problèmes :**

- ▶ Comment insérer/détecter le Fingerprint ?
  - Techniques de watermarking
- ▶ Comment générer le Fingerprint ?
  - Techniques de la théorie des codes
- ▶ Le type d'attaques qu'un Fingerprint est susceptible de subir



# Marking Assumption et Fingerprint de multimédia



## ▶ La "Marking assumption" suppose :

### ▶ La fidélité (facile)

- Les marques sont perceptuellement invisibles et peuvent être découvertes uniquement par comparaison
- Le contenu non marqué n'est pas disponible

### ▶ la robustesse (Difficile)

- Les marques non détectables ne peuvent être changées ou supprimées

# Attaques sur un schéma de Fingerprint



- ▶ **Attaques sur le système de marquage**
  - Exploitation de la robustesse de l'insertion du fingerprinting et de la détection
  - **Attaque par collusion.** Collusion:  $Y=g(X_1, X_2, \dots, X_k)$  où  $g(.)$  est une fonction conçue pour modifier des contenus ayant un Fingerprint.
    - L'attaque la plus efficace
    - Peut même créer un contenu de meilleur qualité
- ▶ **Le pirate peut avoir deux buts**
  1. Supprimer les Fingerprints des contenus
  2. Mener à l'accusation d'un innocent
- ▶ ***But : améliorer la résistance à la collusion pour les attaques de type 1, et augmenter la robustesse aux attaques de type 2***



# Modélisation de la collusion

# Collusion linéaires et non linéaires



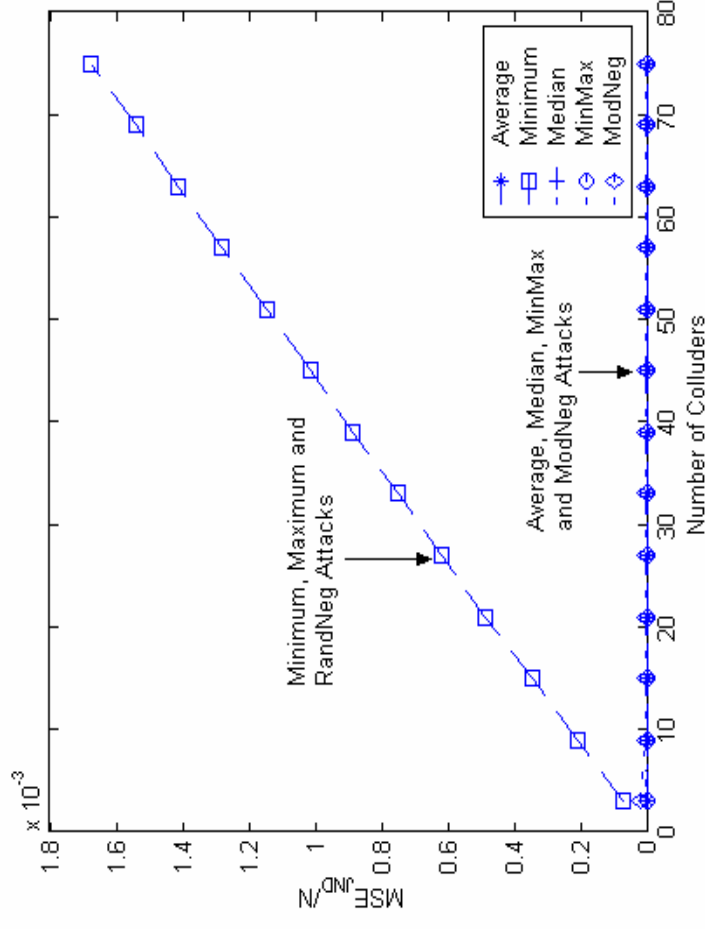
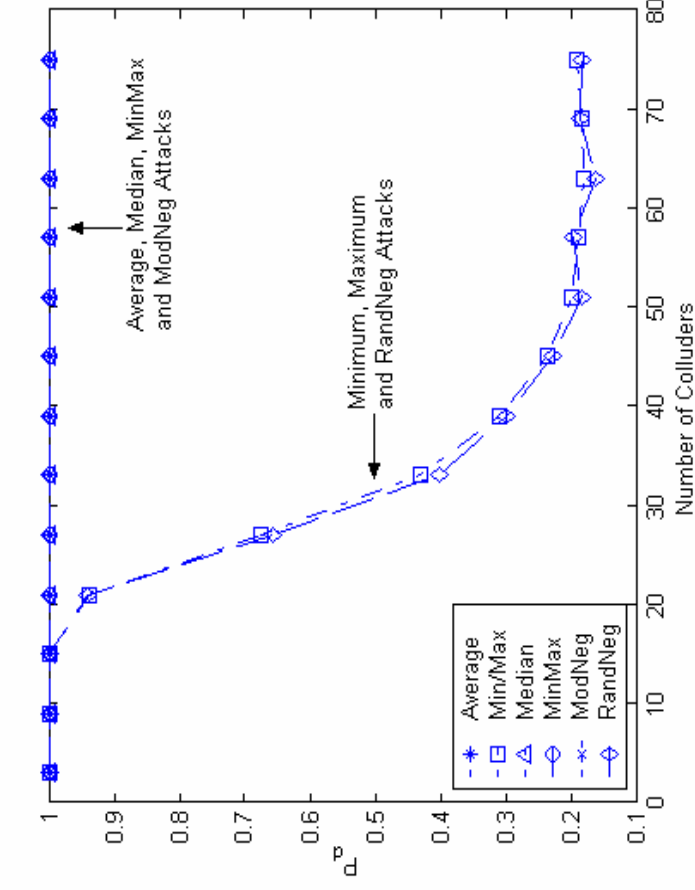
- ▶ La collusion linéaire par moyennage est simple et efficace
- ▶ La collusion peut aussi construire n'importe quelle valeur entre le minimum et le maximum des valeurs observées,  
→ ***Il est important de considérer les collusions non linéaires***
- ▶ Collusions non linéaires basés sur des statistiques du n<sup>ème</sup>

ordre

$$V_j = g(y_j^{(k)})_{k \in S_C} = x_j + JND_j \cdot g(w_j^{(k)})_{k \in S_C}$$

$$\begin{aligned} V_j^{ave} ; V_j^{\min} ; V_j^{\max} ; V_j^{\text{median}} \\ V_j^{\min \max} &= \text{average}(V_j^{\min}, V_j^{\max}) \\ V_j^{\text{modneg}} &= V_j^{\min} + V_j^{\max} - V_j^{\text{med}} \\ V_j^{\text{randneg}} &= \begin{cases} V_j^{\min} & w \cdot p \\ V_j^{\max} & w \cdot p \cdot 1 - p \end{cases} \end{aligned}$$

# Détection de collusion non linéaire



- Si le contenu a  $N = 10000$  coefficients modifiables, et en présence de  $n = 100$  utilisateurs.  $P_{fp} = 10^{-3}$  fixé et les Fingerprints  $\sim N(0, 1/9)$ .
- **"Randomized negative attack"** est la plus efficace (sans normalisation de la distorsion générée par les différentes attaques).
- **Minimum, maximum et "randomized negative attacks"** génèrent une distorsion bien plus importante sur le contenu généré

# Moyenne et collusion non-linéaire



Un détecteur fonctionnant par seuil est  
résistant à différents types d'attaques :  
*moyenne ; statistique du n<sup>ième</sup> ordre*  
*(min, max, ...)*

## Statistique du détecteur

la statistique de détection est une  
gaussienne de même moyenne si la  
collusion est : une moyenne ou une  
autre collusion non linéaire

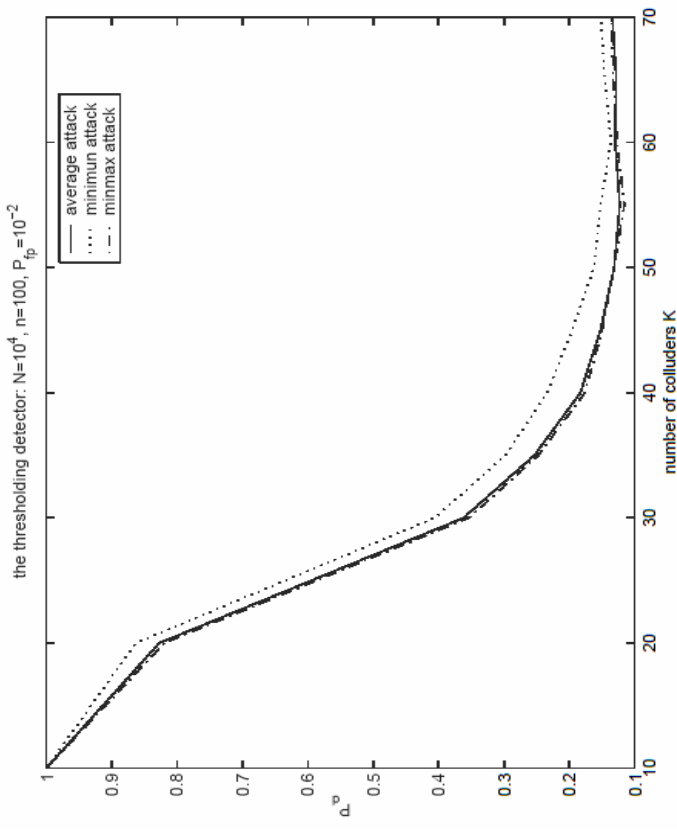
=> Les performances sont les mêmes si  
les fonctions sont similaires

$$g(\mathbf{s}_j, j \in S_c) + \mathbf{d}_1$$

$$\frac{1}{K} \sum_{j \in S_c} \mathbf{s}_j + \mathbf{d}_2$$

Attaque non linéaire

Moyenne





# Fingerprint orthogonal

# Génération de Fingerprint orthogonal



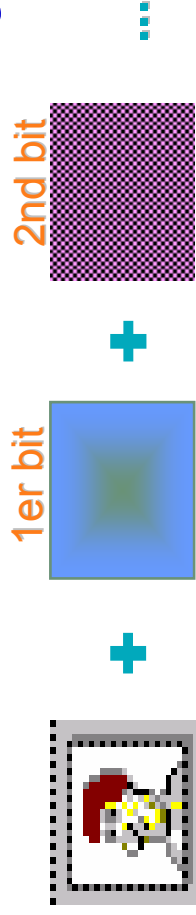
## ► Choix du schéma de modulation

Modulation orthogonale     $\mathbf{w}_j = \mathbf{u}_j$

# fingerprints = # bases orthogonales

Modulation binaire     $\mathbf{w}_j = \sum_{i=1}^B b_{ij} \mathbf{u}_i$   
pour  $b_{ij} \in \{0, 1\}$     ou     $b_{ij} \in \{\pm 1\}$

# fingerprints  $\gg$  # bases orthogonales





# Limites du Fingerprint orthogonal



- ▶ **La taille de collusion maximale est limitée pour deux raisons :**
  - ▶ Les Fingerprints orthogonaux sont fortement atténués quand la taille de la collusion augmente
    - Réduit le taux de détection en présence de collisions de tailles significatives
  - ▶ La probabilité de fausse alarme augmente en fonction du nombre d'utilisateurs
    - Le détecteur détecte une corrélation avec un plus grand nombre de signaux Fingerprintés, chacun correspondant à un utilisateur à identifier
    - Sous l'hypothèse que la détection utilise le même seuil
    - La probabilité de fausse alarme dépend de la probabilité de détection d'un utilisateur qui ne participe pas à la collusion. Ce nombre augmente en fonction du nombre d'utilisateurs

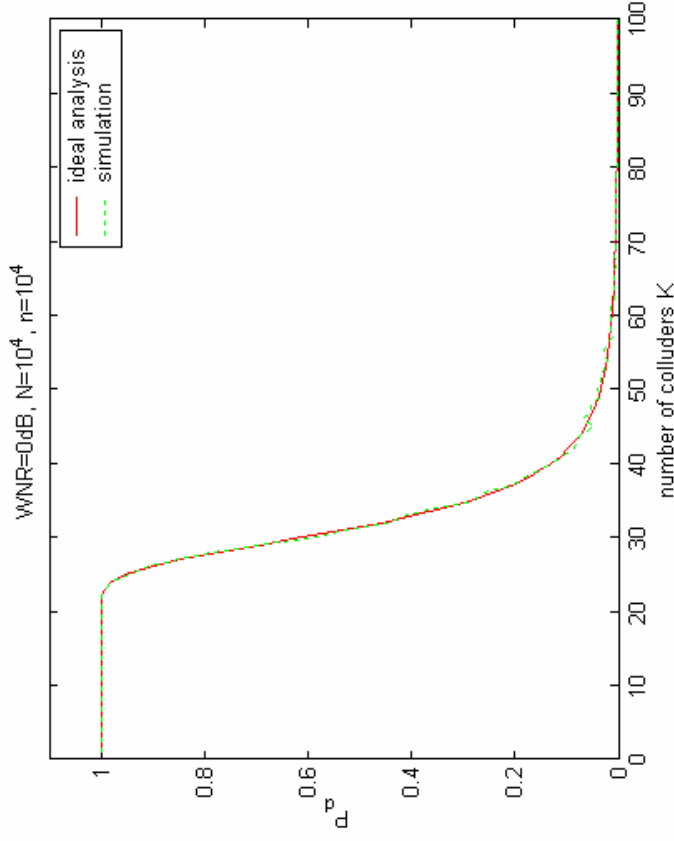
# Limites du Fingerprint orthogonal



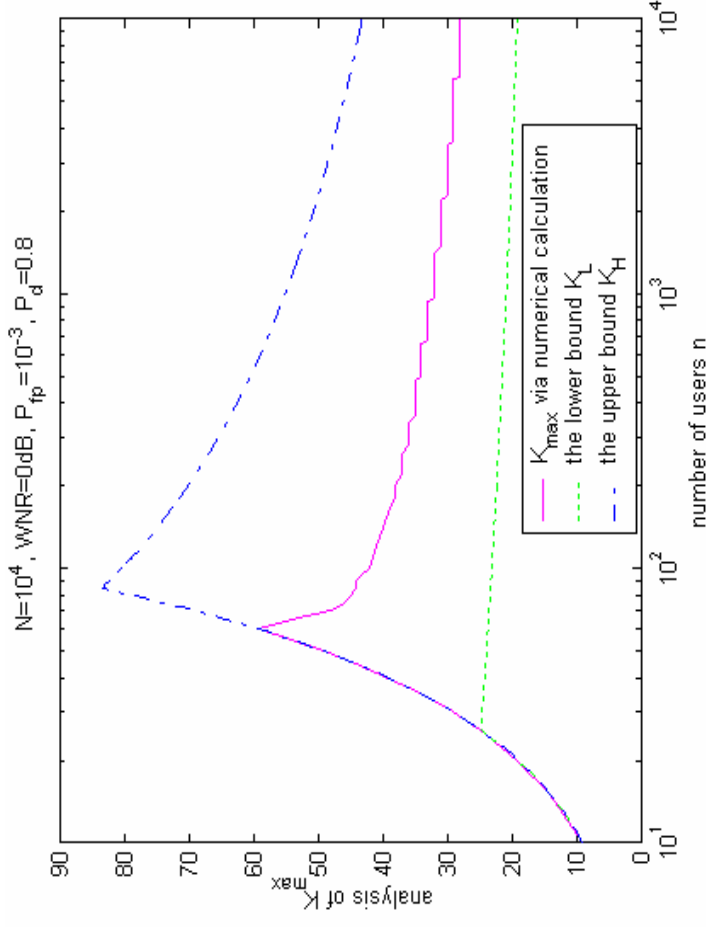
▶ Pour atteindre la probabilité de détection désirée, ainsi que la probabilité de fausse alarme, il est possible d'analyser la taille maximum d'une collusion supportée

=> Ce Fingerprint peut être utilisé pour certaines applications dont les propriétés de robustesse requises correspondent à ceux analysés

# Bornes sur la taille de collusion



$$K_{\max} = 28$$



- Contrainte : identification correcte d'au moins un pirate, sans accuser aucun innocent
- Pour un millier d'utilisateurs, la taille maximale de collusion supportée est bornée par quelques dizaines



# Codes combinatoires

# Codes "Collusion Secure"

- ▶ **Yacobi a adapté efficacement le code de Boneh-Shaw au Fingerprinting de contenus**
  - Codes de Boneh-Shaw et insertion par étalement de spectre
    - Les Bits du code binaire sont remplacés par des séquences d'étalement de spectre
  - Relaxation de l'hypothèse "marking assumption" : après la collusion, les marques ne peuvent pas être changées => après la collusion, des attaques peuvent créer de nouvelles marques
- ▶ **Mais les codes de Boneh-Shaw sont extrêmement longs !**



# Codes "Anti-collision" (ACC)



- ▶ **Construire des Fingerprints corrélés en deux étapes**
  - ▶ Un code binaire Anti-Collision résiste jusqu'à des collisions de taille  $K$ 
    - Chaque sous ensemble de  $K$  utilisateurs partage un ensemble de Bits qui lui est propre
  - ▶ Utilisation d'insertion par modulation pour insérer le Fingerprint
    - Séquences d'étalement de spectre orthogonal
    - Les Bits en commun restent présent lors de la collision et servent à identifier les pirates

# 16-bit ACC pour $\leq 3$ pirates parmi 20



Utilisateur-1 ( -1,-1, -1, 1, 1, 1, 1, ..., 1 )



( -1, 1, 1, 1, 1, 1, ..., -1, 1, 1, 1 ) Utilisateur-4



Plonge le Fingerprint selon un étalement de spectre dans le domaine DCT

Moyenne

Identifie 1 & 4

Mot de code extrait ( -1, 0, 0, 0, 1, ..., 0, 0, 0, 1, 1, 1 )

# Codes ACC et collusion par moyennage



|                      |                |                |               |               |               |               |               |               |               |               |               |               |               |               |               |               |
|----------------------|----------------|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| User 1:              | -1             | -1             | -1            | 1             | 1             | 1             | 1             | 1             | 1             | 1             | 1             | 1             | 1             | 1             | 1             | 1             |
| User 4:              | -1             | 1              | 1             | 1             | 1             | 1             | 1             | 1             | 1             | -1            | -1            | 1             | 1             | 1             | 1             | 1             |
| User 8:              | 1              | -1             | 1             | 1             | 1             | -1            | 1             | 1             | -1            | 1             | 1             | 1             | 1             | 1             | -1            | 1             |
| User(1,4) Average:   | -1             | 0              | 0             | 1             | 1             | 1             | 1             | 1             | 1             | 0             | 0             | 0             | 1             | 1             | 1             | 1             |
| User(1,4,8) Average: | $-\frac{1}{3}$ | $-\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ |
| After thresholding:  | 0              | 0              | 0             | 1             | 1             | 1             | 1             | 1             | 0             | 1             | 0             | 1             | 1             | 0             | 1             | 1             |



Utilisateur 1



Utilisateur 4



Utilisateur 8



La moyenne dans le domaine multimédia correspond à la moyenne dans le code. C'est en fait un opération AND après "discretisation"

Il est possible de distinguer statistiquement les Bits issus de la collusion des Bits originaux avec une modulation et une insertion appropriée et, les Bits originaux étant uniques à un ensemble d'utilisateurs, il est possible de tracer un membre de la collusion



# Codes Fingerprint Anti-Collusion



- ▶ **Simplification de l'hypothèse (peut être relaxée en utilisant une détection appropriée)**
  - En subissant une collusion, le code Fingerprint subit un AND logique
- ▶ **K-resilient AND ACC code**
  - Un code binaire  $\mathbf{C}=\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$
  - Pour tous K mots de code, l'opération de AND logique est distinct de celle effectuée sur tout autre combinaison de K ou moins mots de code

# Balanced Incomplete Block Design (BIBD)



- ▶ **Code ACC via design combinatoire**
  - ▶ **Balanced Incomplete Block Design (BIBD)**

## Exemple

Code ACC via  $(7,3,1)$  BIBD pour résister à une collusion de taille 2 parmi 7 utilisateurs

$$C = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

# Balanced Incomplete Block Design (BIBD)



## ▶ Exemple d'un code (7,3,1) BIBD

- ▶  $X = \{1, 2, 3, 4, 5, 6, 7\}$
- ▶  $A = \{123, 145, 246, 167, 347, 257, 356\}$

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| x | x | x |   |   |   |   |
| x |   |   | x | x |   |   |
|   | x |   | x |   | x |   |
| x |   |   |   |   | x | x |
|   |   | x | x |   |   | x |
|   | x |   |   | x |   |   |

# Balanced Incomplete Block Design (BIBD)



- ▶  **$(v, k, \lambda=1)$ -BIBD est un  $(k-1)$ -resilient AND ACC**
  - ▶ Défini comme une paire  $(X, A)$ 
    - $X$  est un ensemble de  $v$  points
    - $A$  est une collection de blocs de  $X$ , ayant  $k$  points chacun
    - chaque paire de points distincts est exactement dans  $\lambda$  blocs
  - ▶ # mots de code = # blocs
- ▶ **Longueur du code pour  $n = 1000$  utilisateurs :  $O(n^{0.5})$**   
 **$\sim$  qqs dizaines à qqs centaines de bits**
  - ▶ Beaucoup plus court que le code de Boneh-Shaw
  - ▶  $O((\log n)^6) \sim$  qqs millions de bits

$$n = \frac{\lambda(v^2 - v)}{k^2 - k}$$

# Propriété de Helly, d'identification



► Un ensemble est  $t$ -Helly



Toutes les intersections de cardinal  $t$  sont non vides  
 $\Rightarrow$  l'intersection du tout est non vide

► Un code Fingerprint est  $t$ -identifiant ssi,  
grâce à un mot  $M_i$ , il est possible de retrouver un élément  
de la collusion  $I$ , à la condition que  $|I| \leq t$

# $t$ - Helly, $t$ - identification et TRC



- ▶ Étant donné un mot  $m$ ,  $\mathcal{P}_k(m)$  est l'ensemble des collisions de taille  $k$  ayant pu produire  $m$

$$t\text{-identifiant} \Leftrightarrow \bigcap_{\substack{I \in \mathcal{P}_k(m) \\ k \leq t}} I \neq \emptyset \leftrightarrow t\text{-Helly}$$

$$b_i \equiv b_j \pmod{\gcd(m_i, m_j) \forall i, j} \quad \{k.m_i + b_i\}_{k \in \mathbb{Z}} \cap \{k.m_j + b_j\}_{k \in \mathbb{Z}} \neq \emptyset \forall i, j$$
$$\Leftrightarrow$$

$$\cap_i \{x \equiv b_i \pmod{m_i}\} \neq \emptyset \quad \rightsquigarrow \quad \cap_i \{k.m_i + b_i\}_{k \in \mathbb{Z}} \neq \emptyset$$
$$\Uparrow \qquad \qquad \qquad \Uparrow$$

Théorème des Restes Chinois

2-Helly

# Fingerprint basé sur le TRC dans un anneau de polynômes



- => Il y a un rapport entre le théorème des restes chinois et la propriété de  $t$ - identification
- Construction d'un code basé sur le TRC dans  $\mathbb{Z}$
- Le TRC est valide dans tous les anneaux commutatifs unitaires dont tous les idéaux sont principaux
- Un anneau de polynômes  $K[X]$  possède ces propriétés
- Il est donc possible de construire un code Fingerprint  $t$ - identifiant sur un anneau de polynômes
- Le code résultant est beaucoup plus compact que le code de Boneh-Shaw (H. Muratani)



# Étude du traçage



# Détection de la collusion



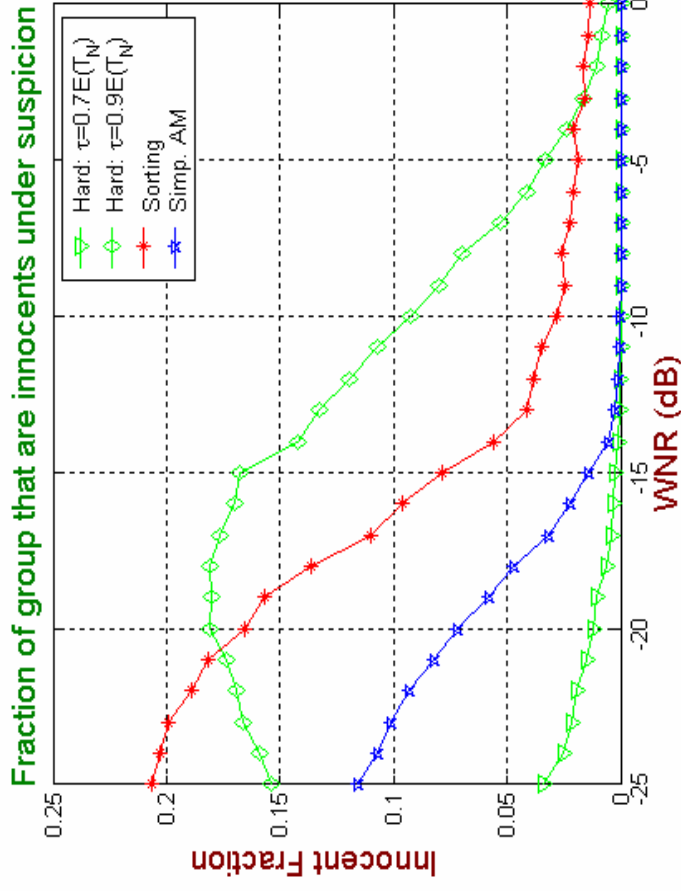
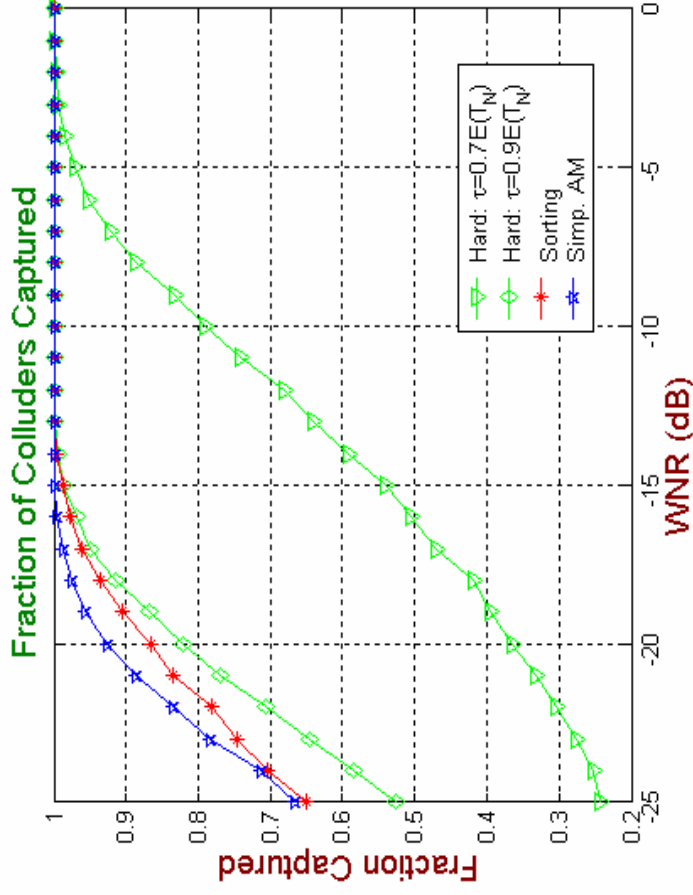
## ▶ Détection Hard :

- Détection des valeurs de bits et **ensuite** cherche les membres de la collusion à partir de ces valeurs
- Tire partie du fait qu'une combinaison de mots de code identifie uniquement les pirates
- Tout le monde est suspect, et chaque Bit à 1 réduit l'ensemble des suspects

## ▶ Détection Soft – candidats possibles :

- Tri : optimisation d'une fonction de recherche, puis détermine **d'abord** la valeur des bits et **ensuite** cherche l'ensemble ayant participé à la collusion
- Méthode séquentielle de recherche : met à jour itérativement la fonction de recherche et identifie **directement** l'ensemble ayant participé à la collusion

# ACC avec un signal Gaussien



- Seuils élevés : identifie plus de pirates, mais suspecte plus d'innocents également
- Détection Soft donne de meilleurs taux de détection que la détection Hard
- Décodage et identification conjointe donne de meilleurs performances qu'en séparant ces deux étapes
- L'identification séquentielle de la collusion (Simp. AM) a un bon équilibre entre performance et temps de calcul



# Travaux récents

# Fingerprint et codes correcteurs d'erreurs



## ► Utilisation des codes correcteurs d'erreurs pour construire un schéma de codage/insertion de Fingerprint

- Les Fingerprints basés sur des symboles (non binaires) sont intéressants pour les contenus à base de trames (vidéo et audio)
- Travaux de comparaisons (avantages/désavantages) des Fingerprints basés sur des codes correcteurs d'erreurs et des Fingerprints orthogonaux

## ► Résultats principaux

- Amélioration significative de la résistance, et meilleur compromis résistance à la collusion / efficacité de la détection
- Démontre l'intérêt des schémas où le codage et l'insertion sont intimement liés

# Fingerprint et communication synchrone mono-canal multi-utilisateurs



- ▶ Le problème de la résistance à la collusion est similaire à celui de la communication synchrone multi-utilisateurs sur un unique canal de transmission

- ▶ le contenu ayant subit une collusion est : le contenu + une mixture de Fingerprints

**Problème de la collusion**

$$\mathbf{y}_c = \frac{\alpha}{K} \sum_{j=1}^n \phi_j \mathbf{w}_j + d$$

$\phi_j \in \{0,1\} = 1$  si jeme a participé

$\mathbf{w}_j =$  fingerprint ts

**Canaux CDMA synchrones**

$$y(t) = \sum_{k=1}^n A_k b_k s_k(t) + n(t)$$

$b_k \in \{-1,1\}$

$s_k(t) =$  séquence de signatures

# Collusion et communication synchrone mono-canal multi-utilisateurs

- ▶ Pour avoir des bonnes performances, les séquences CDMA doivent avoir un minimum d'interférences entre elles, cad. de faibles corrélations
- ▶ La similarité entre la collusion et la communication synchrone suggère que de bonnes séquences CDMA seraient de bons Fingerprints
- ▶ => Travaux récents de Trappe



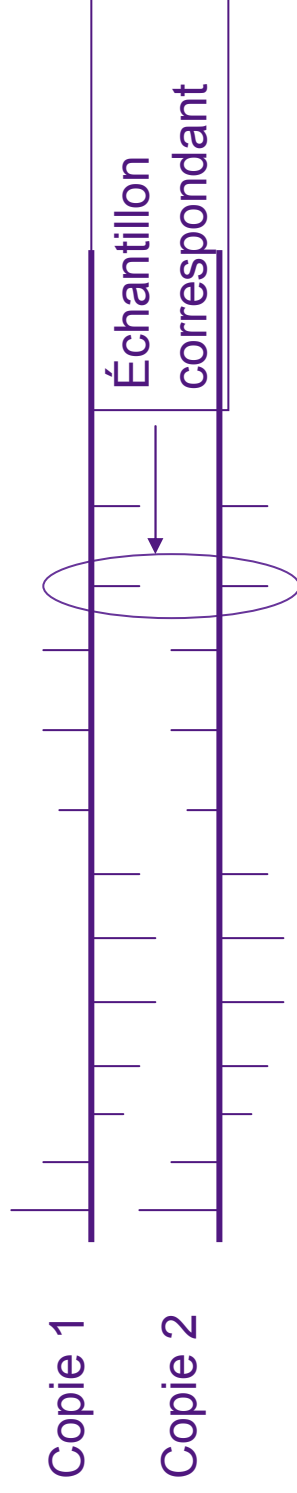


# Désynchronisation

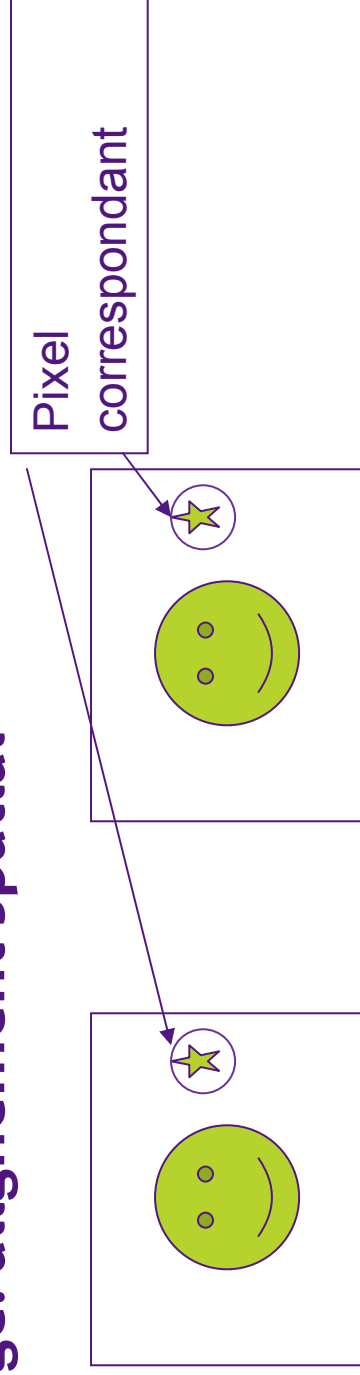
# Alignement et collusion



## ▶ Audio: alignement temporel



## ▶ Image: alignement spatial



## ▶ Vidéo: alignement spatial et temporel



# Désynchronisation intentionnelle de vidéo

- ▶ **Idée : avant l'insertion de la marque, faire subir des distorsions géométriques au contenu (dans chaque trame, et dans l'échantillonnage des trames)**
- ▶ **But : la collusion directe produira des effets visuelles qui dégradent le contenu**

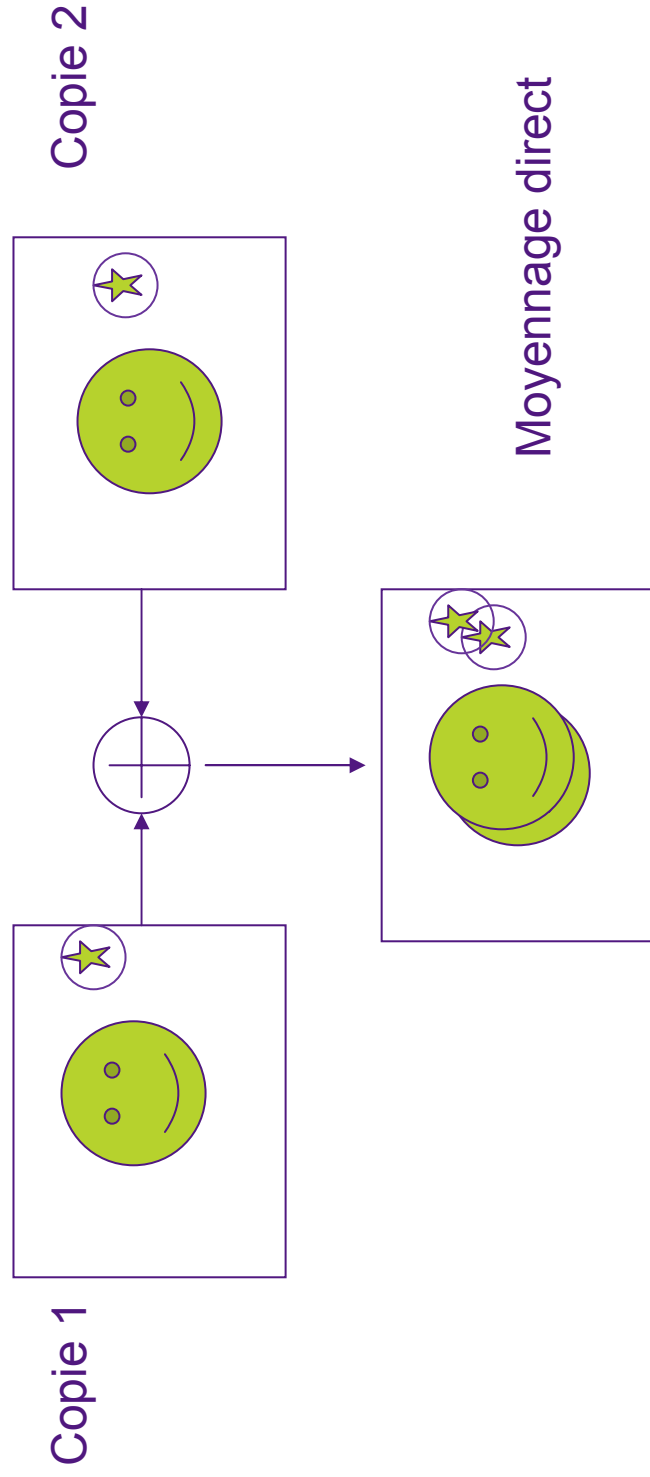


# Désynchronisation



## ▶ Objectifs

- Doit créer des artefacts perceptibles en cas de collusion
- Doit rendre la tâche de resynchronisation des pirates "calculatoirement inaccessible" (la resynchronisation quasi-parfaite est possible avec une quantité de calculs infinie)



# La désynchronisation



## ▶ **Contraintes de qualité**

- ▶ La désynchronisation intentionnelle ne doit pas altérer le contenu perceptuellement
- ▶ Elle ne doit pas non plus affecter les performances du détecteur de fingerprint

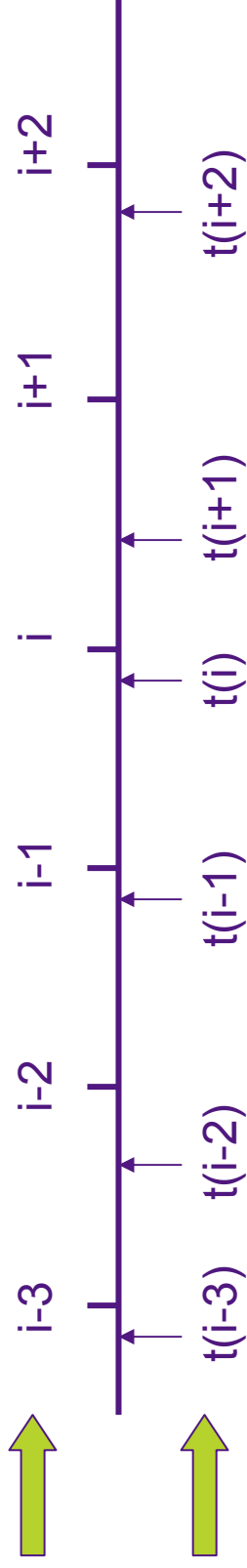
## ▶ **Techniques de désynchronization vidéo**

- ▶ Échantillonnage aléatoire par interpolation vidéo
- ▶ Rotation, translation, changement d'échelles aléatoires
- ▶ Filtre volumétrique (sur la luminance, le contraste, le volume) variant au cours du temps

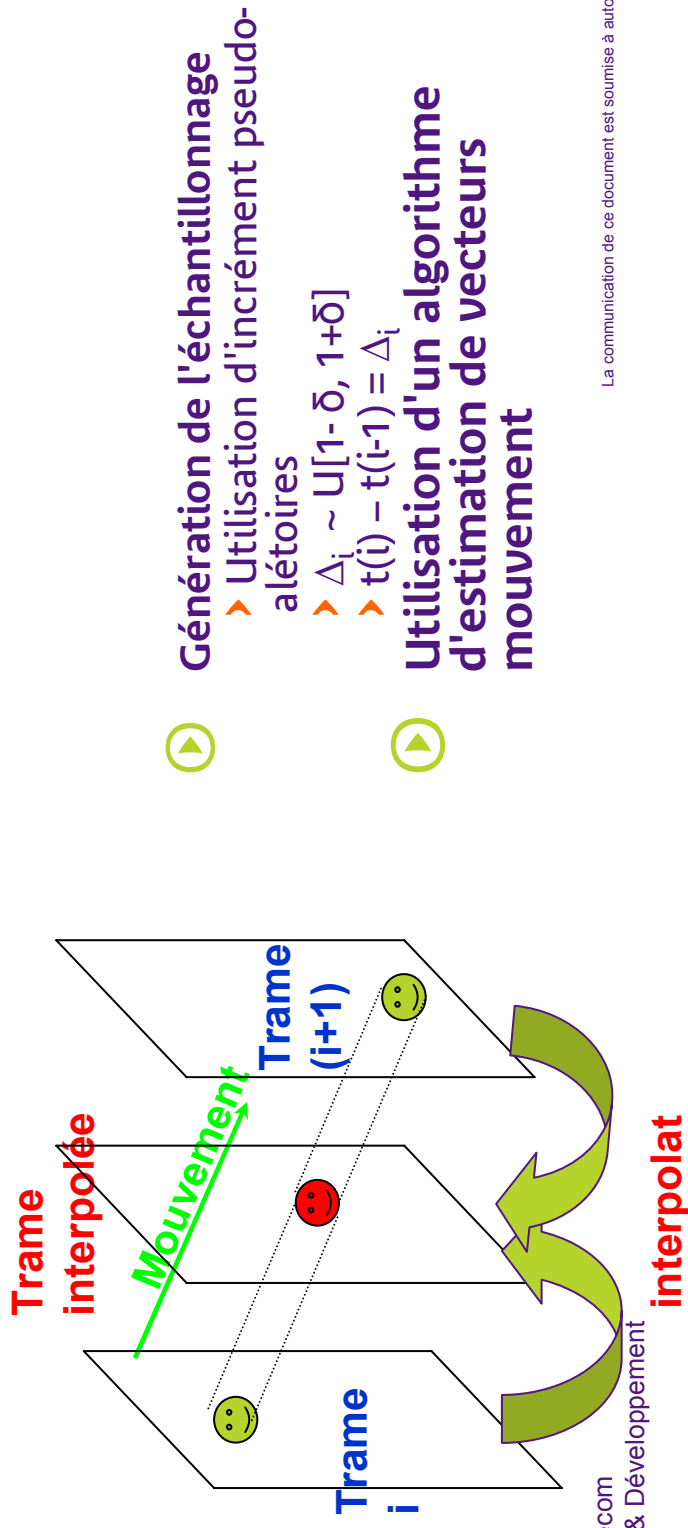
# Mouvement aléatoire basé sur un ré-échantillonnage



Position des trames originales uniformément échantillonnées



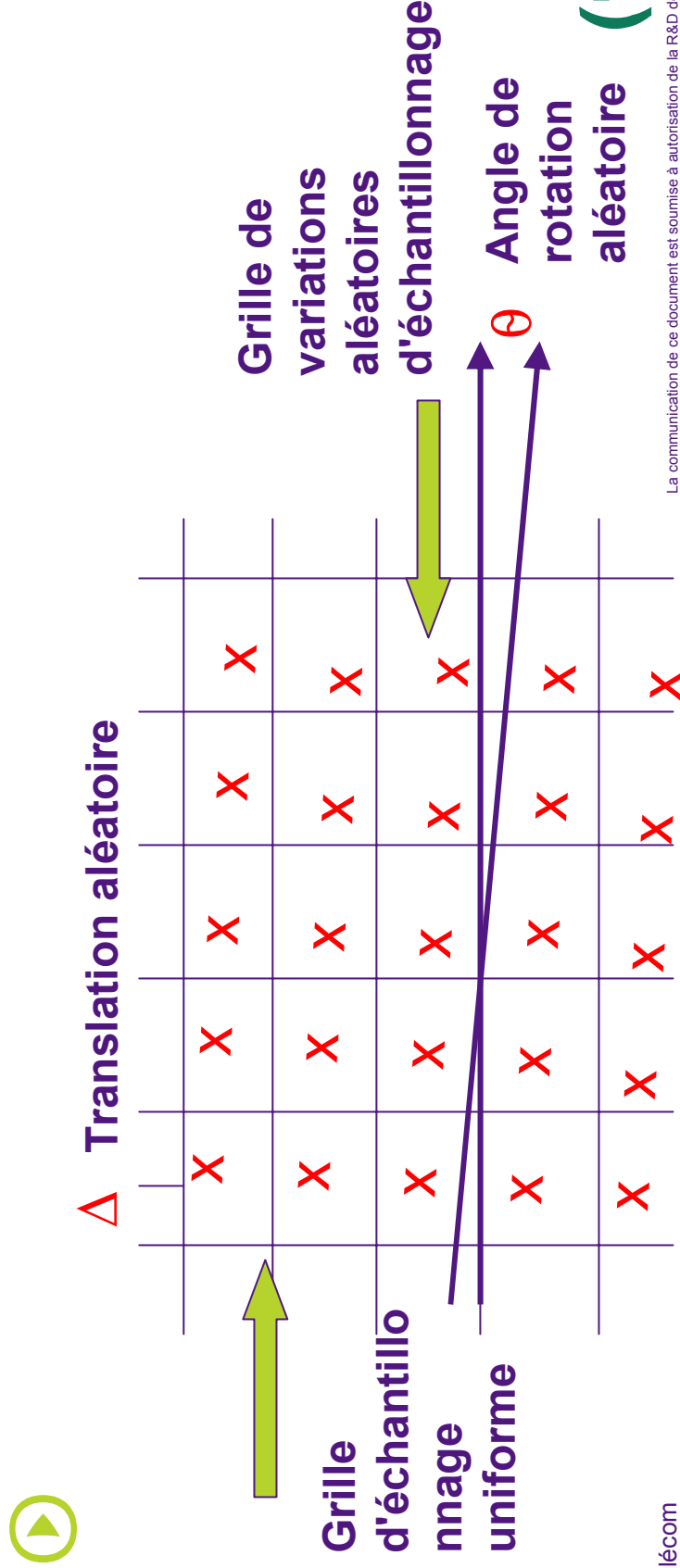
Trames échantillonnées aléatoirement (avec contrainte)



# Désynchronisation spatiale



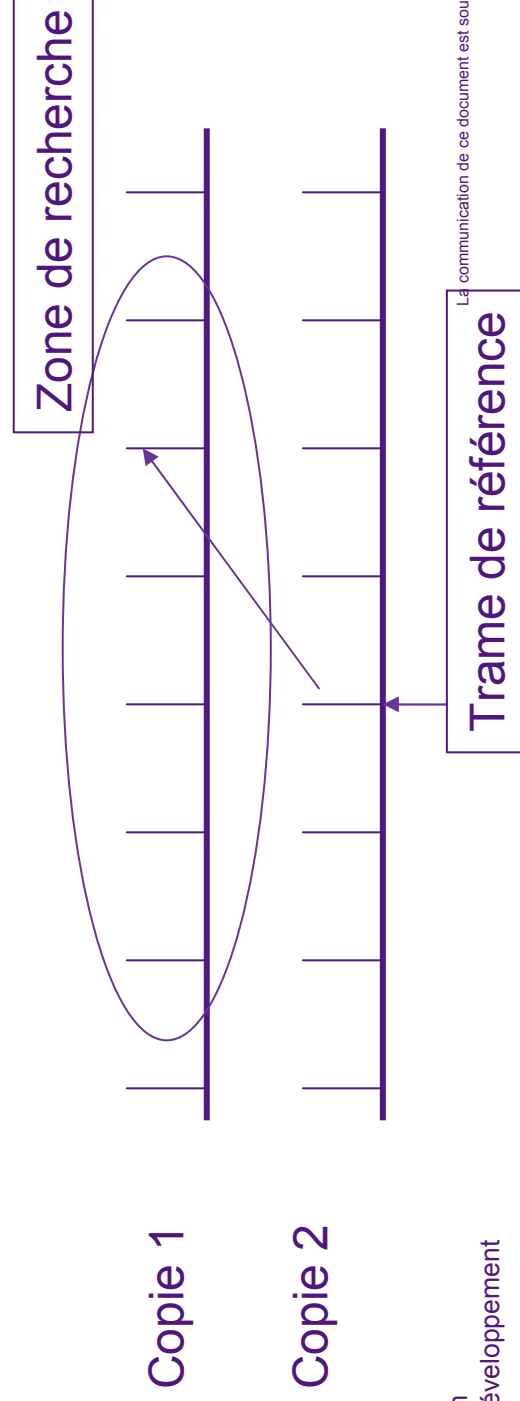
- ▶ Rotation aléatoire, changement d'échelle, translation
- ▶ Variations aléatoires en utilisant une grille d'échantillonnage (temporelle et spatiale)
- ▶



# Attaque par collusion



- ▶ **Attaque simple : moyennage direct**
- ▶ **Plus astucieux : trouver la trame la plus "proche", puis moyennage**
  - Utiliser une copie comme référence, les autres comme cibles
  - Pour chaque trame de référence  $F_a$ , déterminer une zone de recherche  $S$  parmi les trames cibles
  - Trouver la trame cible  $F_t$  qui minimise  $SVA(F_a, F_t)$  (SVA : Somme des différences de Valeurs Absolues)



# Tatouage – de-synchronisation



**Original**



**Après de-synch.**



**Moyennage direct**



**Moyennage après alignement**

