THOMSON

Le traçage de traîtres en multimédia

Teddy.furon@thomson.net

Introduction

- **Définition du traçage de traîtres.**
  - Lutter contre la redistribution illégale d'un contenu dématérialisé.
    - Ici : contenu = vidéo H.264 (multimédia)
  - Retrouver l'identité de la (ou les) source(s) parmi les *n* utilisateurs ayant eu le contenu.
  - Arme dissuasive.
  - Aussi connu sous le nom de : *fingerprinting, content serialization, user forensics, transactional watermarking…*

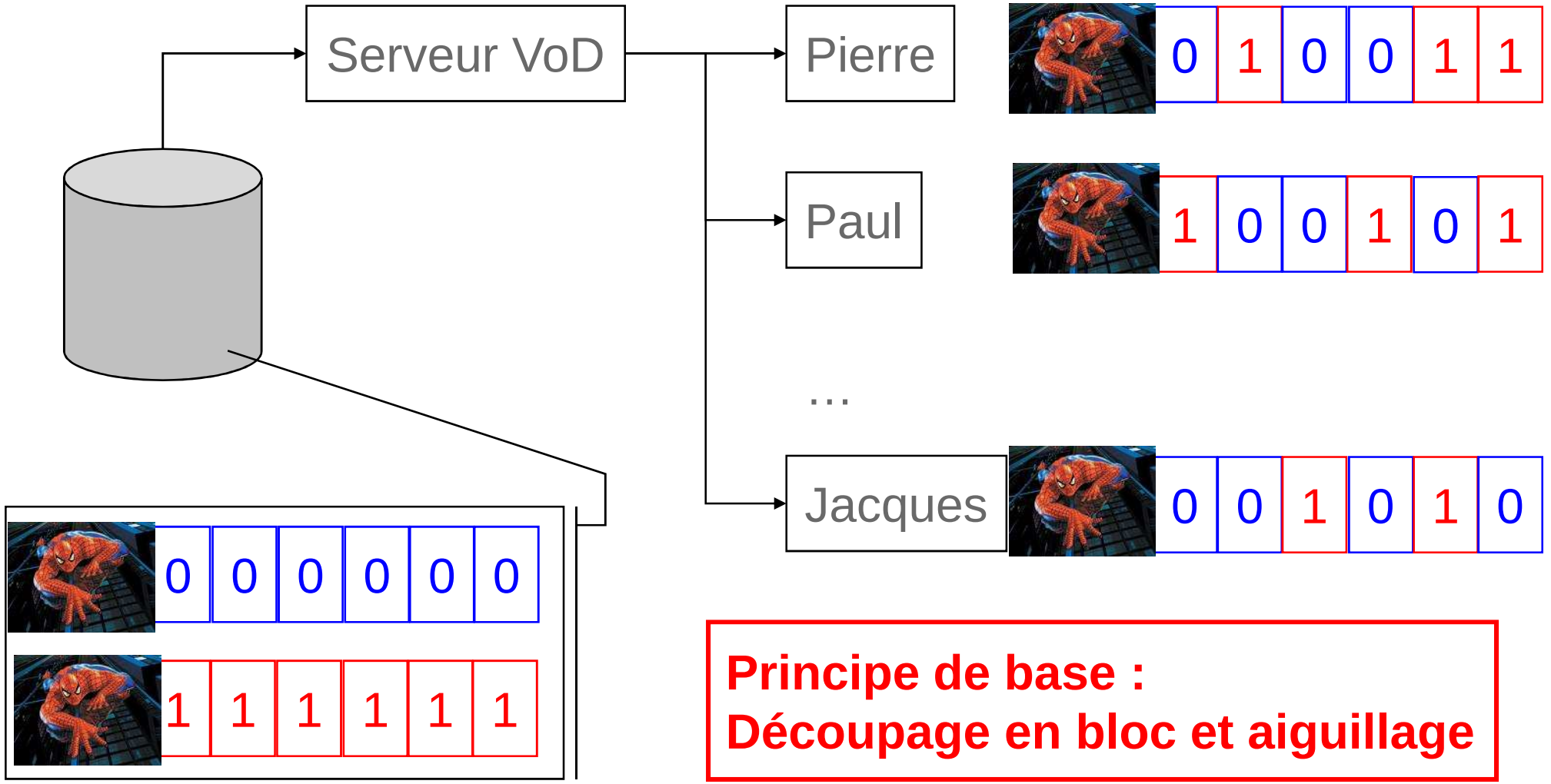
Introduction

- **Quelques faits sur le traçage.**
  - Début XXe s. : Arrondis des valeurs de table de logarithme,
  - 1980 : Qui est la taupe du cabinet de Ms. Tatcher ?
  - 2004 : Screeners des Oscars,
    - La source : 600,000$ d'amende, exclu du jury, banni pdt 5 ans.
    - Le complice : 3 ans de prison.
  - Apple iTunes Store
    - Insertion d'un identifiant dans les fichiers non-DRMisés
    - DRM 2.0 : filtrage de contenus et traçage de traîtres.
  - Prévu dans la norme AACS protection Blu-Ray Disc
    - Sequence-Key, IBM (identification de l'appareil)
    - Associé à du broadcast encryption (révocation de l'appareil)

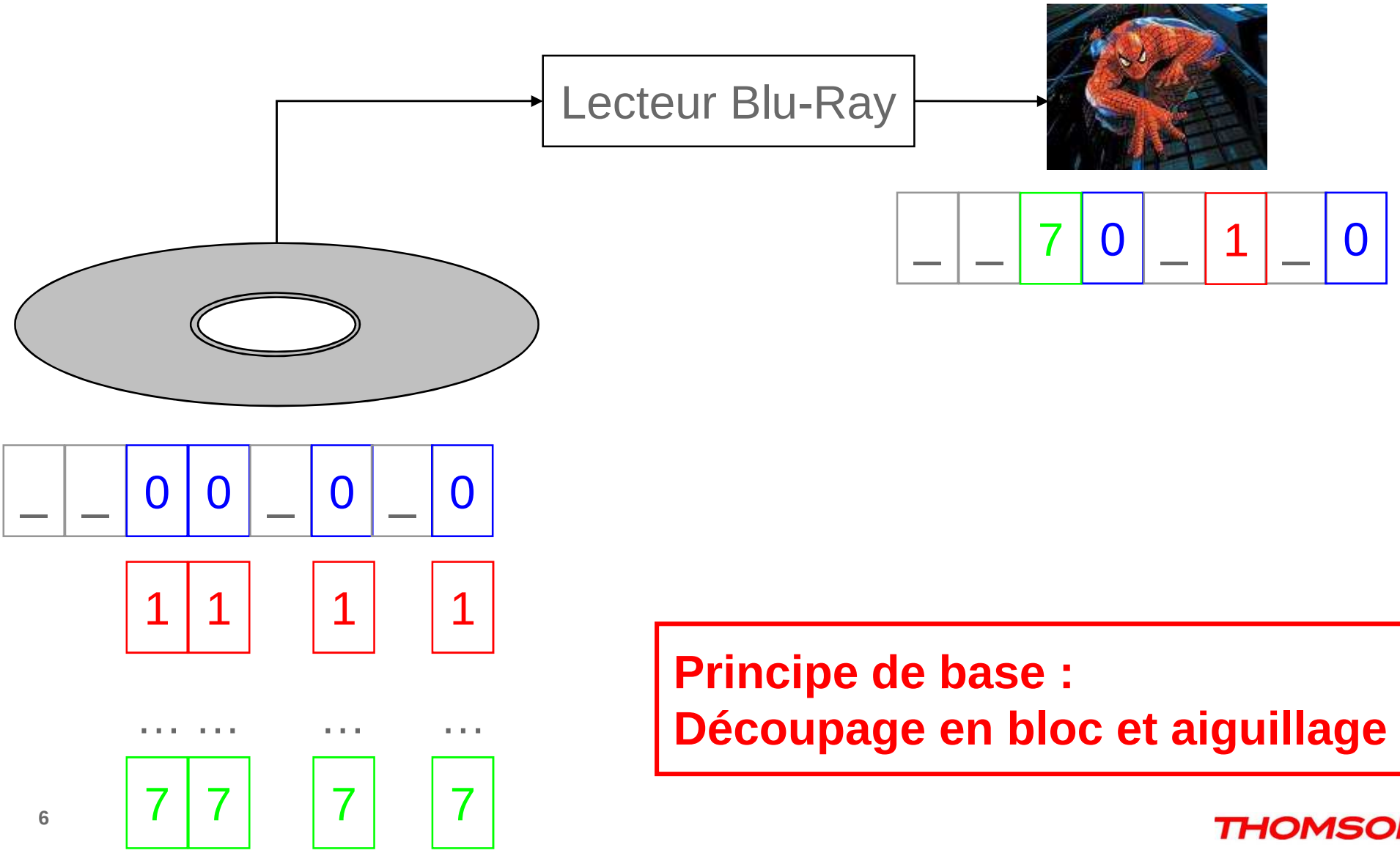
Introduction

- **3 schémas types différents**
  - Video on Demand (VoD)
    - Unicast
    - Le serveur personnalise la copie.
  - Blu-Ray Disc
    - Multicast
    - Le serveur prépare et le lecteur personnalise.
  - Décodeur satellite
    - Broadcast
    - Le décodeur personnalise.

- **Accusation hors-ligne**
  - Labo forensics



Principe de base :
Découpage en bloc et aiguillage



Principe de base :
Découpage en bloc et aiguillage

Introduction

- **La collusion**
  - Plusieurs utilisateurs malhonnêtes produisant une copie pirate à partir de leurs versions.
  - Chimère académique ?
  - Argument de défense de l'accusé : « victime d'une collusion ».

- **2 briques technologiques**
  - Le code anti-collusion (matrice *n* x *m*)
    - Association utilisateur ⇔ séquence de *m* bits
  - Le tatouage
    - Insertion de la séquence binaire dans le contenu

  - Comparaison modèle OSI : couches physique / liaison de données

Plan

- Introduction
- **La collusion :**
  - 3 procédés : échange, fusion et dégradation
- **Le code anti-collusion**
- **Le tatouage**
- **Conclusion**

La collusion : Définition

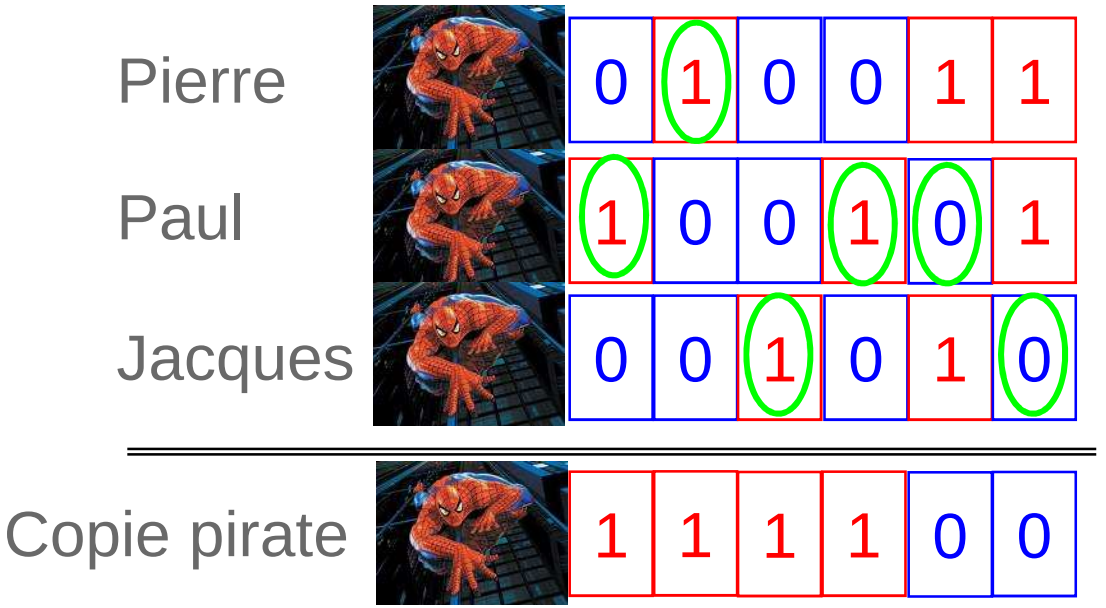
- **Définition :**
  - « Entente secrète entre deux ou plusieurs personnes pour nuire à un tiers », Dico TLFi
  - Les *colluders,* au nombre de *c*

- **Hypothèse de travail :**
  - Découpage en bloc du contenu,
  - Insertion d'un symbole par bloc (cf. partie tatouage),
  - Ce découpage n'est pas secret.

- **Le *i*-ème bloc de la copie pirate est construit à partir des *i*-èmes blocs des colluders.**

La collusion : 3 procédés

- **L'échange de blocs :**
  - Le *i*-ème bloc de la copie pirate est **égal à l'un** des *i*-ème blocs des colluders.
  - Tout se passe comme si les *c* colluders échangent des bits de leurs séquences.



THOMSON

La collusion : 3 procédés

- **Combien d'échanges de bloc possibles ?**
  - $N = c^m$? $N = 2^m$? $N = 2^{m'}$? nb de particules dans univers ~ 2^{300}
  - Stratégies :
    - tirage uniforme : Pierre, Paul, Jacques, Jacques, Pierre, …
    - vote majoritaire : le symbole le plus représentatif ?
    - tout à zéro : mettre un `0' dès que possible
    - …mais combien ?
    - Quelle est la plus méchante ?
  - Règle d'or :

$$X_{i,j1} = X_{i,j2} = \ldots = X_{i,jc} = a \quad \Rightarrow \quad Y_i = a$$

- **La défense est assurée par le code anti-collusion.**

La collusion : 3 procédés

- **La fusion de blocs**
  - Le bloc pirate est la fusion des blocs colluders.
  - Procédés intra-bloc sur des échantillons :
    - moyenne, min, max, médian, …
    - Tiling : découpage en morceaux,
    - Domaine : pixels, coefficients DCT, coefficients ondelette.
  - Il n'y a que 2 versions d'un bloc : `1' ou `0' enfouis.
  - But: ne pas retrouver de bit caché (effacement)

- **La défense est assurée par le tatouage numérique.**

La collusion : 3 procédés

- **La dégradation**
  - La copie pirate est dégradée par un post-traitement : compression DivX, formatage, débruitage, …
  - Un seul colluder peut le faire.
  - Dangers :
    - Effacement : incapable de retrouver le bit caché,
    - Erreur de décodage : bit erroné.

- **La défense est assurée par le tatouage numérique.**

Plan

- Introduction
- La collusion
- **Le code anti-collusion**
  - 2 approches : cryptographique et statistique
- **Le tatouage**
- **Conclusion**

Code anti-collusion cryptographique

- **Code X = matrice binaire n x m**
  - $n << 2^m$
- **Terminologie des codes**
  - Traçabilité faible : *Frameproof* (FP), *Secure Frameproof* (SFP),
  - Traçabilité forte : *Identification Parent Pr.* (IPP), *Traceable* (T).
- **Pour une taille de collusion donnée**
  - « Mon code est c-SFP. »
- **Des codes de plus en plus parfaits**

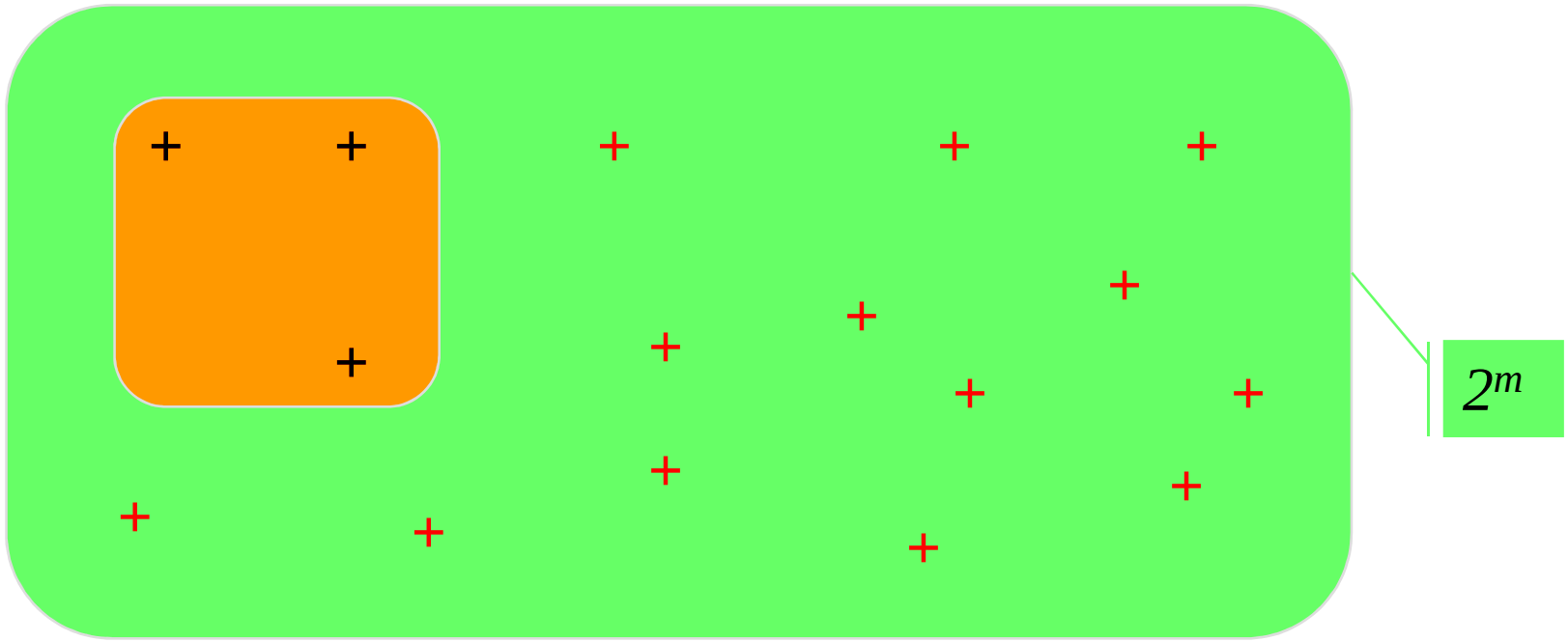
$$c\text{-FP} \subset c\text{-SFP} \subset c\text{-IPP} \subset c\text{-T}$$

- **Un critère important**
  - La longueur m

Code anti-collusion cryptographique

Frameproof

{+}: code *X* **{+}:** collusion *C* : descendance



$$desc(C) \cap X = \varnothing$$

THOMSON

Code anti-collusion cryptographique

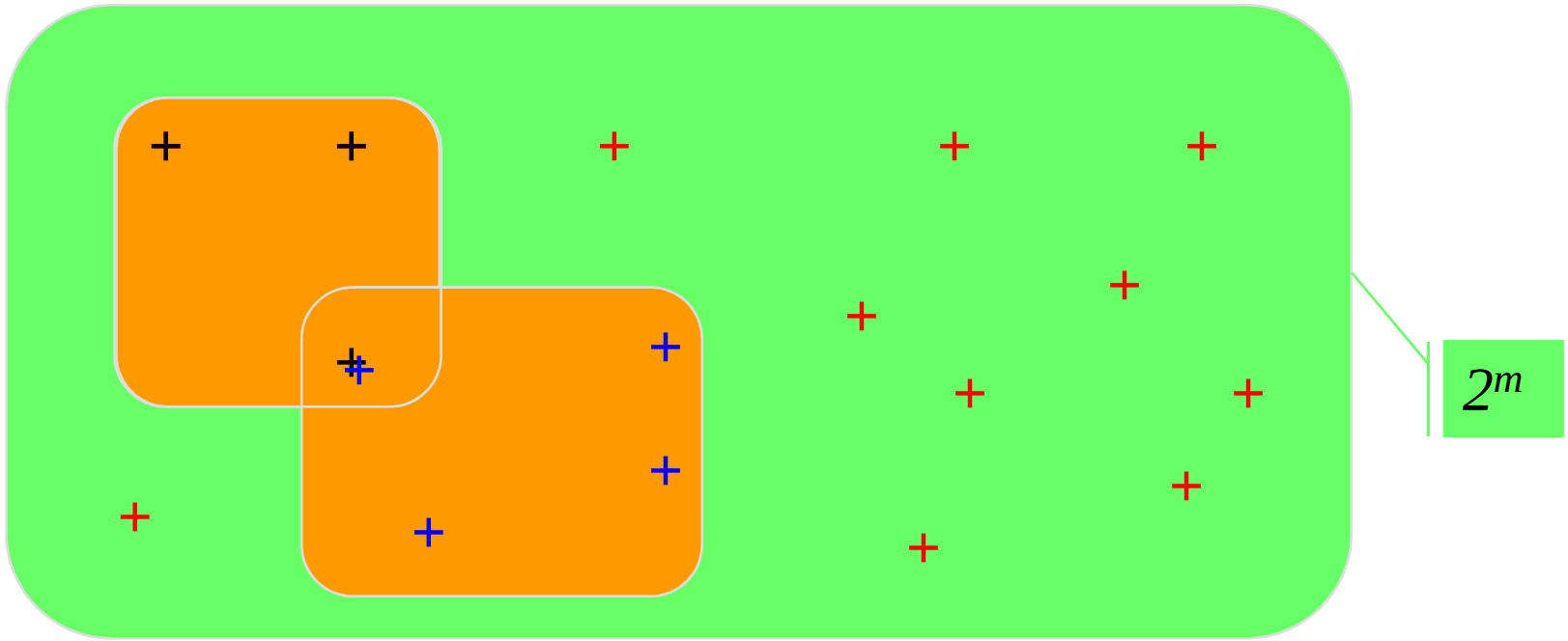
Secure Frameproof

{+}: code X

{+}: collusion C_a

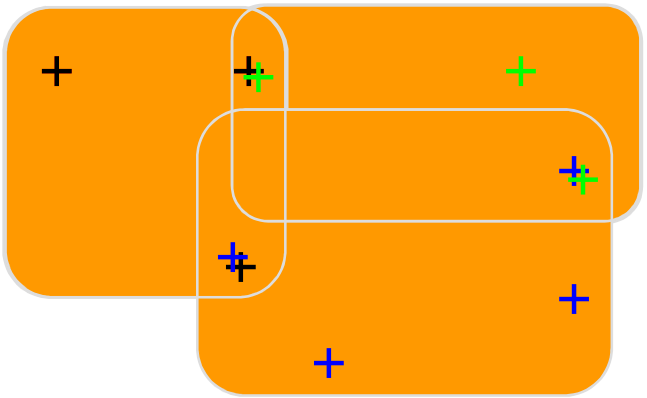
{+}: collusion C_b

: descendance



$$\text{desc}(C_a) \cap \text{desc}(C_b) \neq \emptyset \Rightarrow C_a \cap C_b \neq \emptyset$$

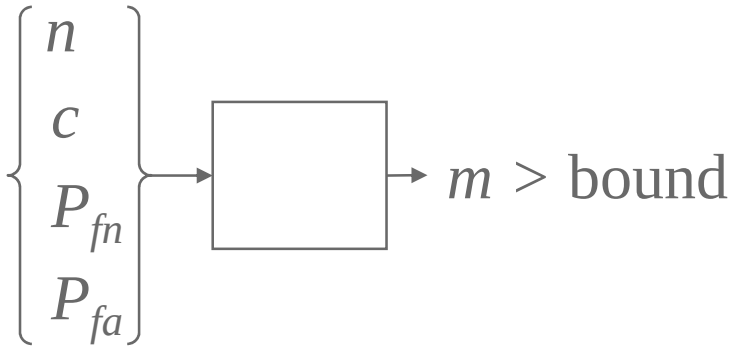
Code anti-collusion cryptographique



- **FP et SFP sont nécessaires mais pas suffisantes**
  - « On accuse parfois, mais jamais à tort. » P_{fn}
  - « On accuse toujours, mais parfois à tort. » P_{fa}
- **Traçabilité forte (IPP et traceable)**
  - Trop contraignant
  - En binaire, impossible si $c > 2$.
  - Longueur de code très très longue.
- **Outil de base : code correcteur d'erreurs**

Code anti-collusion cryptographique

- **Le jeu de la traçabilité faible**



- **Borne de Peikert, Shelat, Smith (2003)**
  - Théorème non constructif

$$m \geq O\left[c^2.\log(n.P_{fa}^{-1}) \right]$$

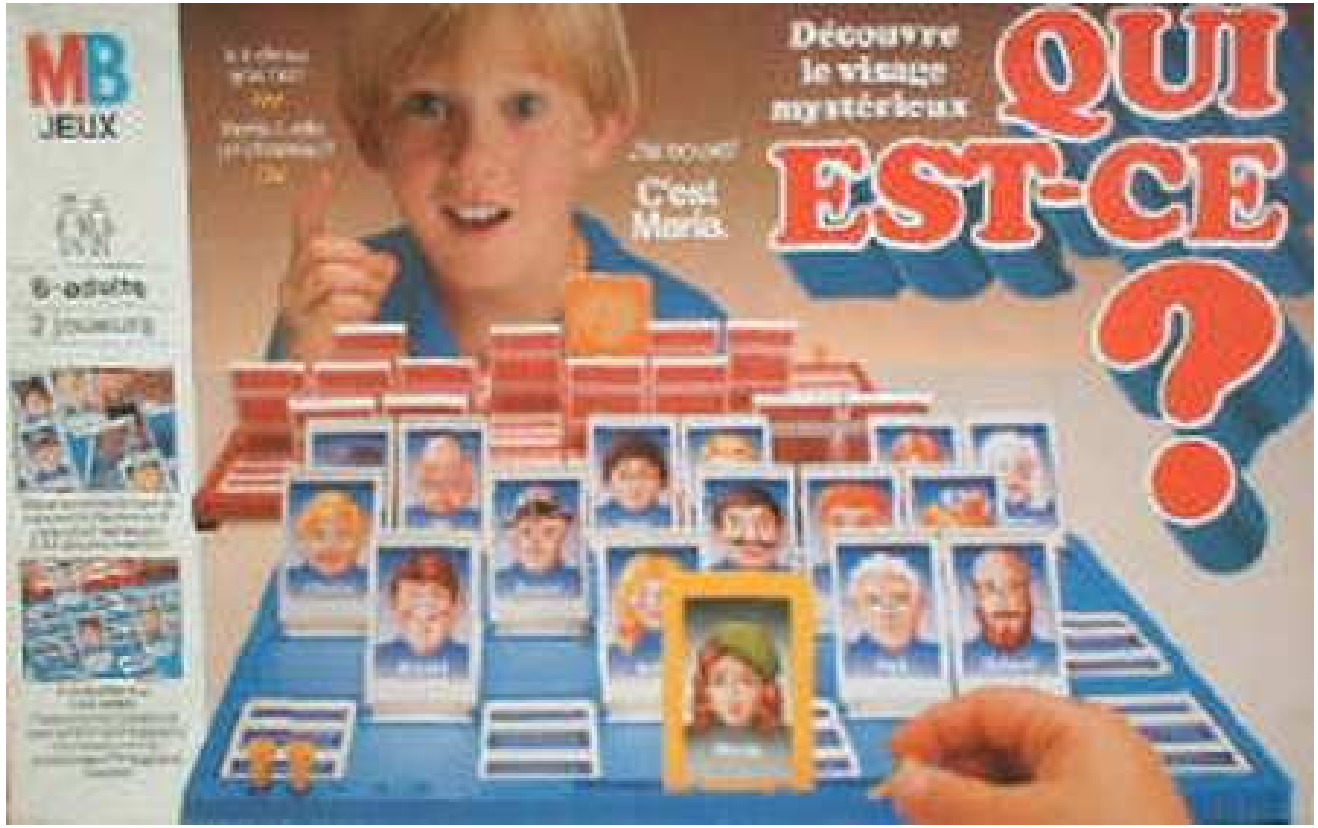
- **Vous avez dit « probabilités » ?**

Code anti-collusion statistique

- **Gabor Tardos**
  - Le génie inconnu
  - Il montre un code atteignant la borne inférieure
  - Moins de 10 lignes de code en Matlab
  - Il n'explique rien :
    - « Soit un tel code X , alors $m \sim 100.c^2.\log(n.P_{fa}^{-1})$] . »
  - Que se cache - t - il derrière un code de Tardos ?
    - Des jeux « Qui est-ce ? » « Cluedo » : Faisceau de preuves
    - De la théorie des jeux

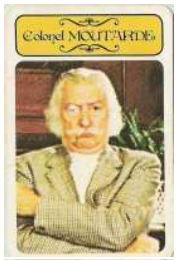
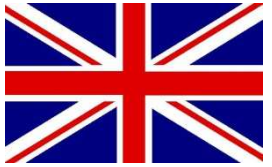
Code anti-collusion statistique

- « Qui est-ce ? » a plusieurs inconnus



THOMSON

Le docteur Le Noir retrouvé mort

				♂		Somme
+1	-1	+1	+1	+1	+3
+1	+1	-1	-1	+1	+1
-1	- 1	+1	-1	+1	- 1
-1	-1	-1	+ 1	-1	-3

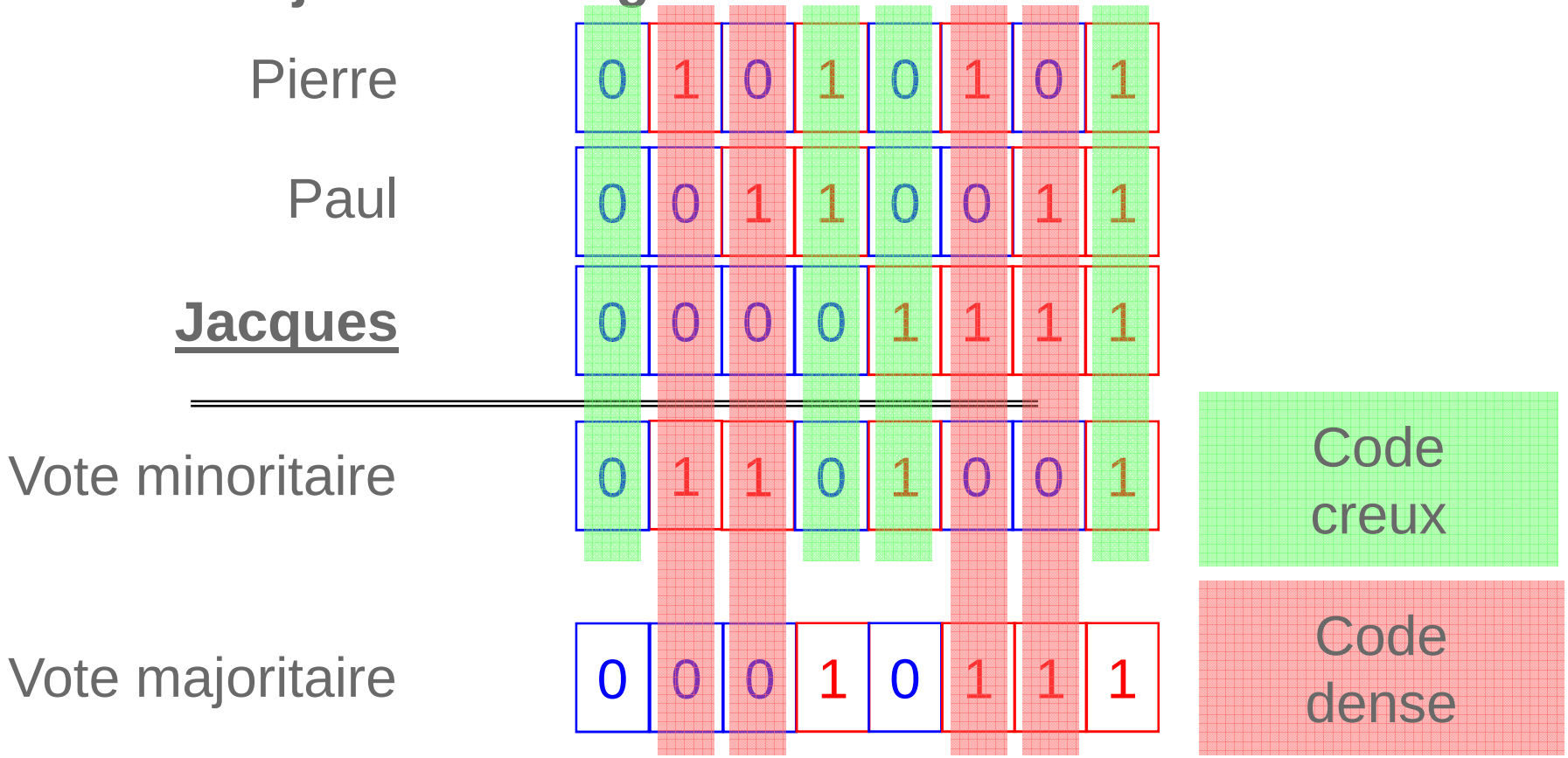
Une copie pirate est retrouvée

Y_1=0	Y_2=1	Y_3=1	Y_4=0	Y_5=1	Somme
+1	-1	+1	+1	+1	+3
+1	+1	-1	-1	+1	+1
-1	- 1	+1	-1	+1	- 1
-1	-1	-1	+ 1	-1	-3

THOMSON

Code anti-collusion statistique

- **Théorie des jeux : échange de blocs à 3 colluders**



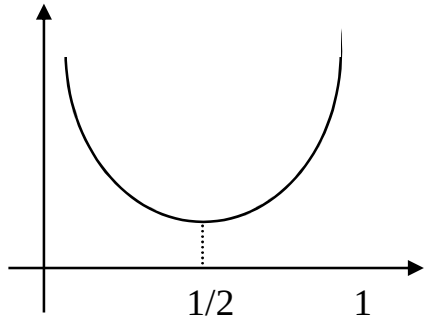
THOMSON

Code anti-collusion de Tardos

- **Initialisation**
  - Tirer au hasard m variables: p_i suivant distribution $f(p)$.
  - Ces variables constituent la clé secrète du code.
- **Construction**
  - Tirer au hasard les X_{ji} t.q. Prob(X_{ji} = 1) = p_i
- **Accusation**
  - Pour chaque utilisateur #j:

$$S_j = \sum \text{ -/+ } g(p_{yi})$$

  - Soit on accuse le plus gros score,
  - Soit on accuse les scores supérieurs à un seuil.



Code anti-collusion de Tardos

- **Théorie des jeux**
  - Décodeur simple

$$\max \min E_P [I (Y ; X | P)]$$

  - Décodeur joint

$$\max \min E_P[I (Y ; \{X_1,.., X_k\} | P)]$$

- **Meilleur score ?**
  - Décodeur simple (*n* scores)

$$S_j = \sum \text{ -/+ } g(p_{yi})$$

  - Décodeur joint ($O(n^k)$)

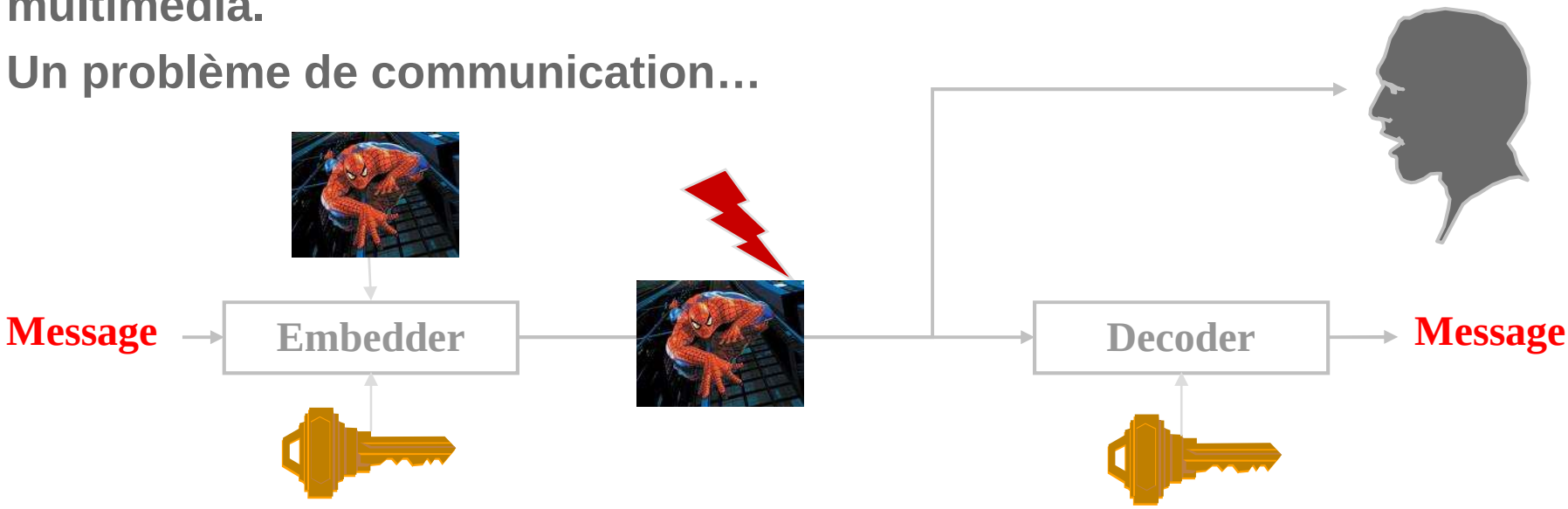
$$S_{\{j, \ldots ,k\}} = ?$$

Plan

- Introduction
- La collusion
- Le code anti-collusion
- **Le tatouage**
- Conclusion

Le tatouage numérique

- **Data hiding: l'art de cacher des données dans des contenus multimédia.**
- **Un problème de communication…**

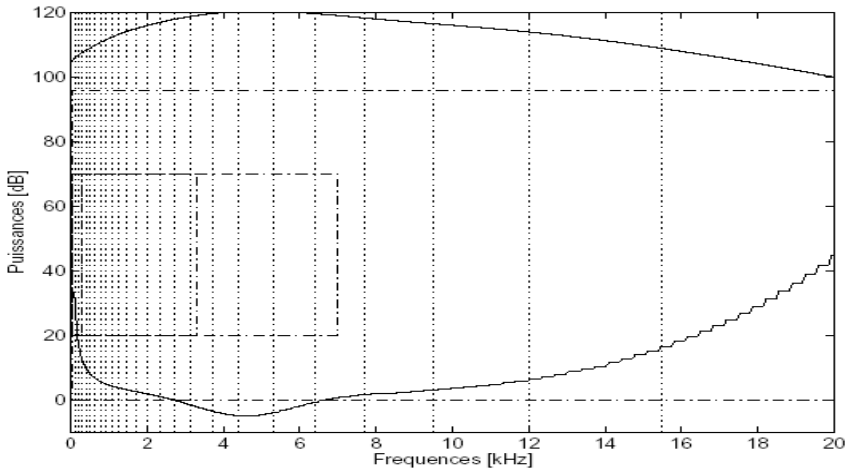


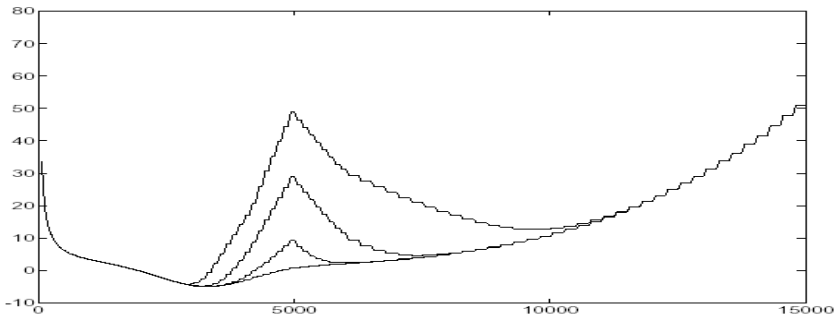
- **… sous constraintes:**
  - non perceptible, (watermark, *filigrane* en Français)
  - capacité,
  - robustesse, (*tatouage* en Fançais)
  - sécurité.

THOMSON

Le tatouage numérique

- **Comment être imperceptible ?**
- **En tirant profit les défauts de la perception humaine**
  - Sensibilité : hautes fréquences, couleur sombre, couleur claires.
  - Masquage :
    - La présence d'un signal fort masque un signal faible
    - Temporel : l'oreille insensible aux échos courts
    - Fréquentiel : l'oreille ne perçoit pas deux sons trop proches





Le tatouage numérique

- **Comment être robuste ?**
  - En étant partout présent : spatial, temporel, fréquentiel,
  - Tous les échantillons sont modifiés mais un peu seulement.
  - L'étalement de spectre (WW II: SIGSALY)

Règle d'or : compromis

Conclusion

- **Maturité**
  - Bornes théoriques
  - Codes / algorithmes d'accusation efficaces
  - Les pires attaques

- **Futur**
  - Anonymat
  - Asymétrique
  - Séquentiel

THOMSON

Thank you for your attention

This document is for background informational purposes only. Some points may, for example, be simplified. No guarantees, implied or otherwise, are intended