

PROTECTION DES DONNÉES VISUELLES

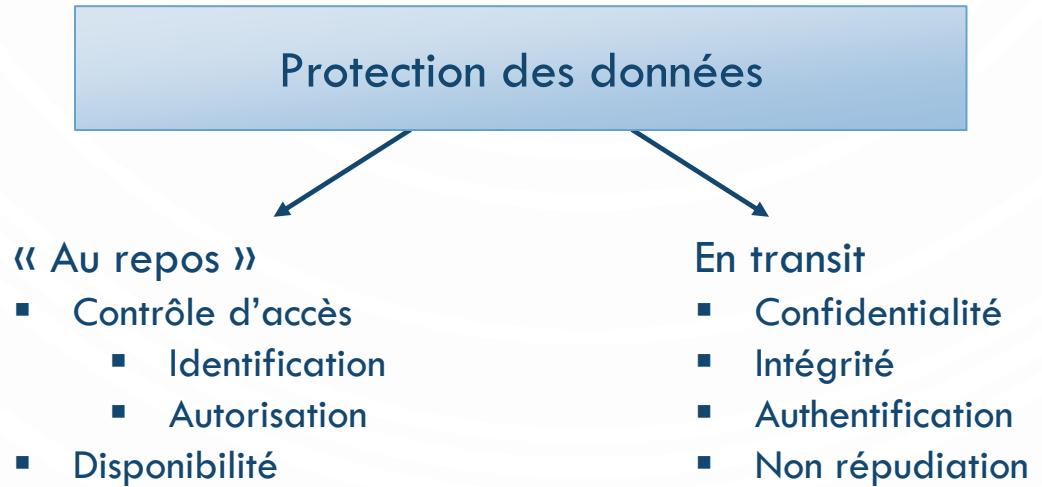
HAI918I - IMAGE, SÉCURITÉ ET DEEP LEARNING (OCTOBRE 2021)

PAULINE PUTEAUX



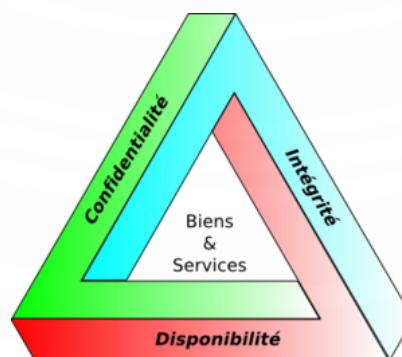
BESOIN IMPORTANT EN SÉCURITÉ

- Essor du « cloud computing »
- De nombreuses menaces potentielles...



PRINCIPES DE SÉCURITÉ

- **Confidentialité** : L'information n'est accessible qu'à ceux dont l'accès est autorisé
- **Authentification** : Chaque personne est bien celle qu'elle prétend être (légitimité)
- **Intégrité** : Le message envoyé n'a pas été altéré de manière volontaire ou involontaire
- **Non-répudiation** : Aucune des deux parties ne pourra assurer ne pas être l'auteur du message
- **Disponibilité** : L'accès à un service ou à des ressources est garanti



PROTECTION DES DONNÉES VISUELLES

- D'après CISCO, les données visuelles = **80% du trafic Internet mondial** en 2019 (contre 67% en 2014).
 - Nécessité de proposer des méthodes efficaces pour protéger ces données visuelles !

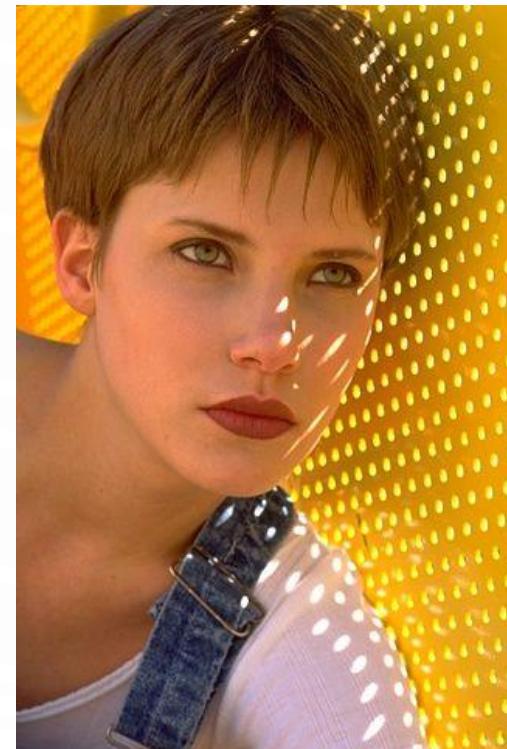
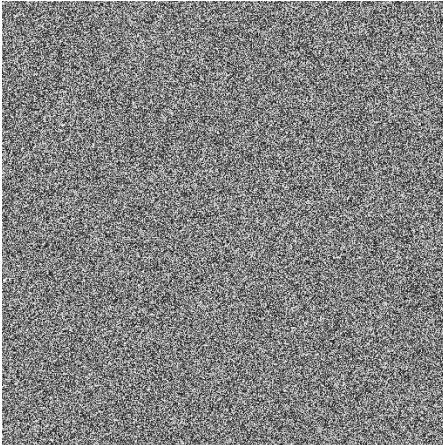
PROTECTION DES DONNÉES VISUELLES

- D'après CISCO, les données visuelles = **80% du trafic Internet mondial** en 2019 (contre 67% en 2014).
 - Nécessité de proposer des méthodes efficaces pour protéger ces données visuelles !
- De nombreux axes :
 - Cryptographie
 - Tatouage
 - Stéganographie
 - Analyse forensique
 - Biométrie



Attention à ne pas confondre tatouage, stéganographie et cryptographie !

CRYPTOGRAPHIE



TATOUAGE



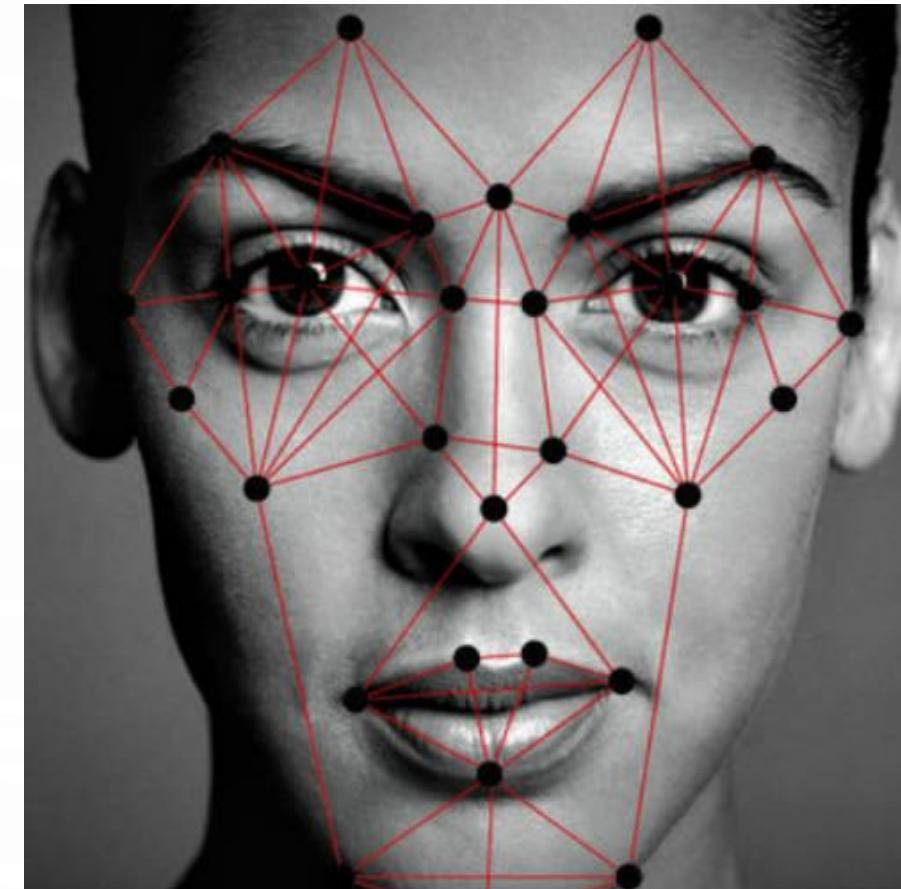
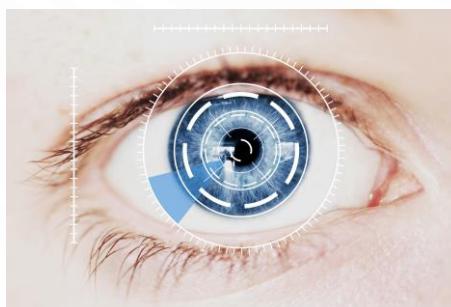
STÉGANOGRAPHIE



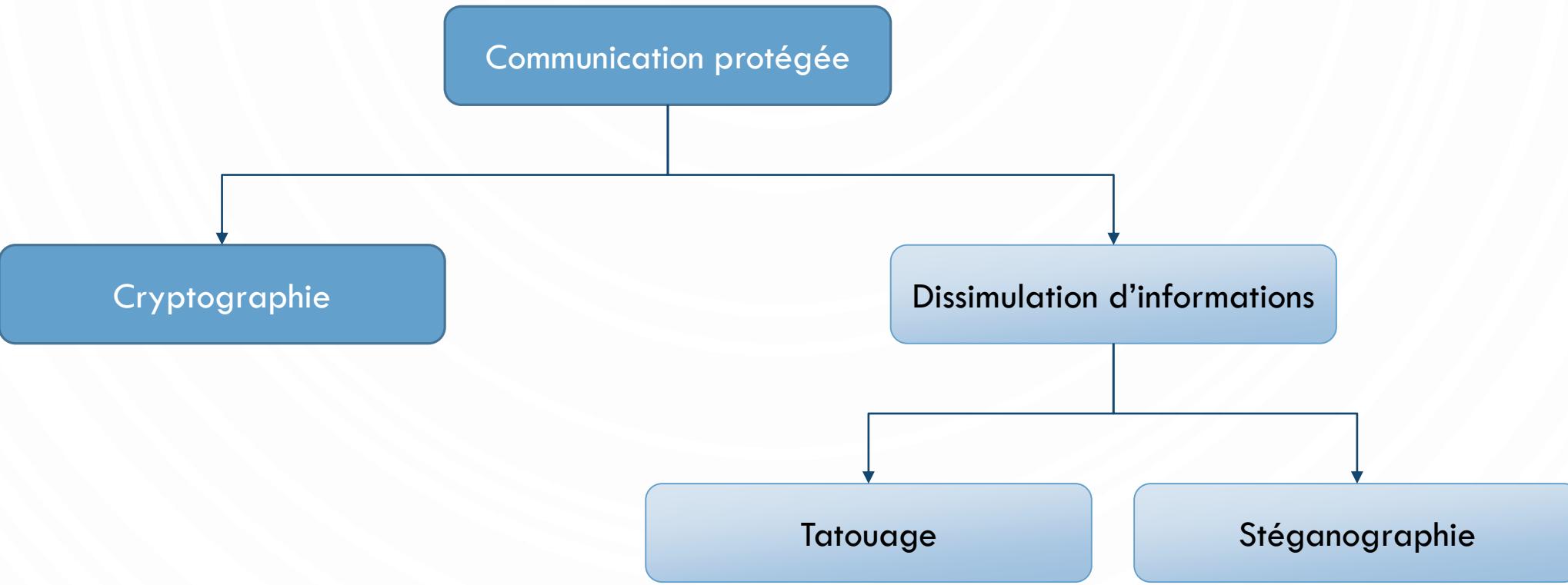
ANALYSE FORENSIQUE



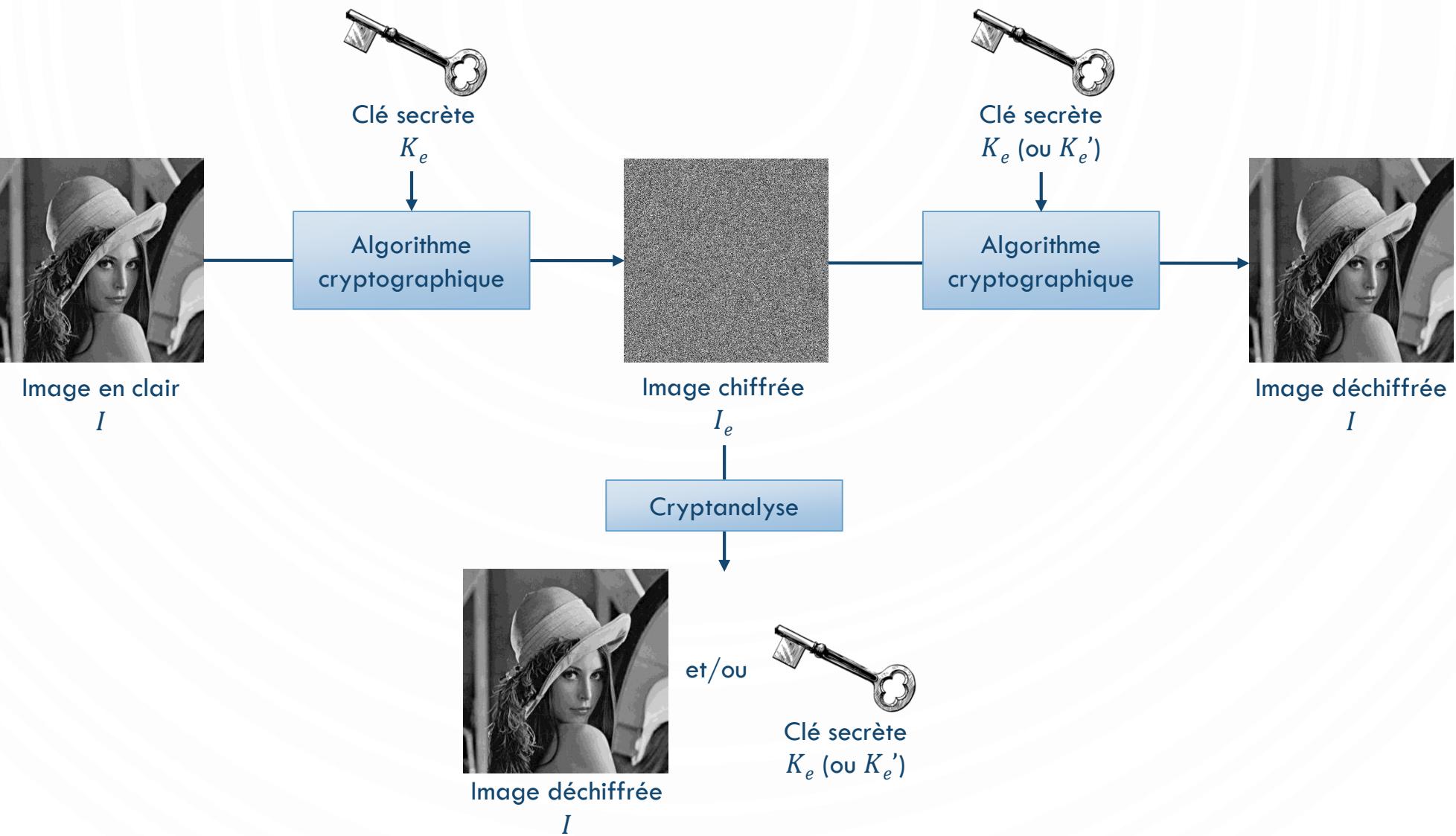
BIOMÉTRIE



TECHNIQUES DE PROTECTION DE DONNÉES



TERMINOLOGIE



PRINCIPE DE KERCKHOFFS

- Aucun secret ne doit résider dans l'algorithme cryptographique utilisé : **tout réside dans la clé !**
 - « La sécurité ne doit pas dépendre de tout ce qui ne peut pas être facilement changé. »
 - Pour un algorithme : **secret \neq robustesse**
- Sans la clé, il doit être impossible de retrouver le message clair à partir du chiffré.
- Si on connaît la clé, on doit pouvoir déchiffrer le chiffré sans problème.



Auguste Kerckhoffs, *La cryptographie militaire*, Journal des sciences militaires, vol. IX, pp. 5–38, jan. 1883, pp. 161–191, févr. 1883.

ALGORITHME PUBLIÉ VS SECRET

ALGORITHME PUBLIÉ

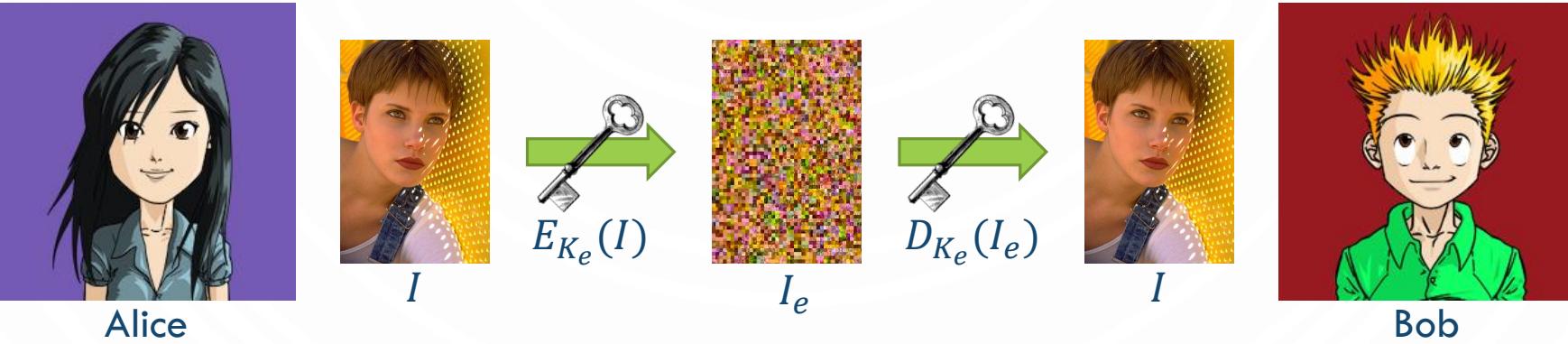
- Possible d'évaluer la sécurité de manière fiable
- Empêche les backdoors cachées par les concepteurs
- Grand nombre d'utilisateurs = Prix réduit + performance élevée
- Pas besoin de protection contre le reverse engineering
- Implémentations logicielles
- Standardisation locale et internationale

ALGORITHME SECRET

- La cryptanalyse doit inclure la récupération de l'algorithme
- Petit nombre d'utilisateurs = Plus petite motivation à essayer de casser l'algorithme
- Indisponible pour un autre pays

CRYPTOGRAPHIE SYMÉTRIQUE

- La **cryptographie symétrique**, également dite **à clé secrète** permet à la fois de chiffrer et de déchiffrer des messages à l'aide d'une même clé secrète



- Deux catégories de méthodes :
 - Chiffrement **par flot** : traitement des données de longueur quelconque, sans besoin de les découper
 - Chiffrement **par bloc** : découpage des données à chiffrer en blocs de taille généralement fixe

CRYPTOGRAPHIE SYMÉTRIQUE

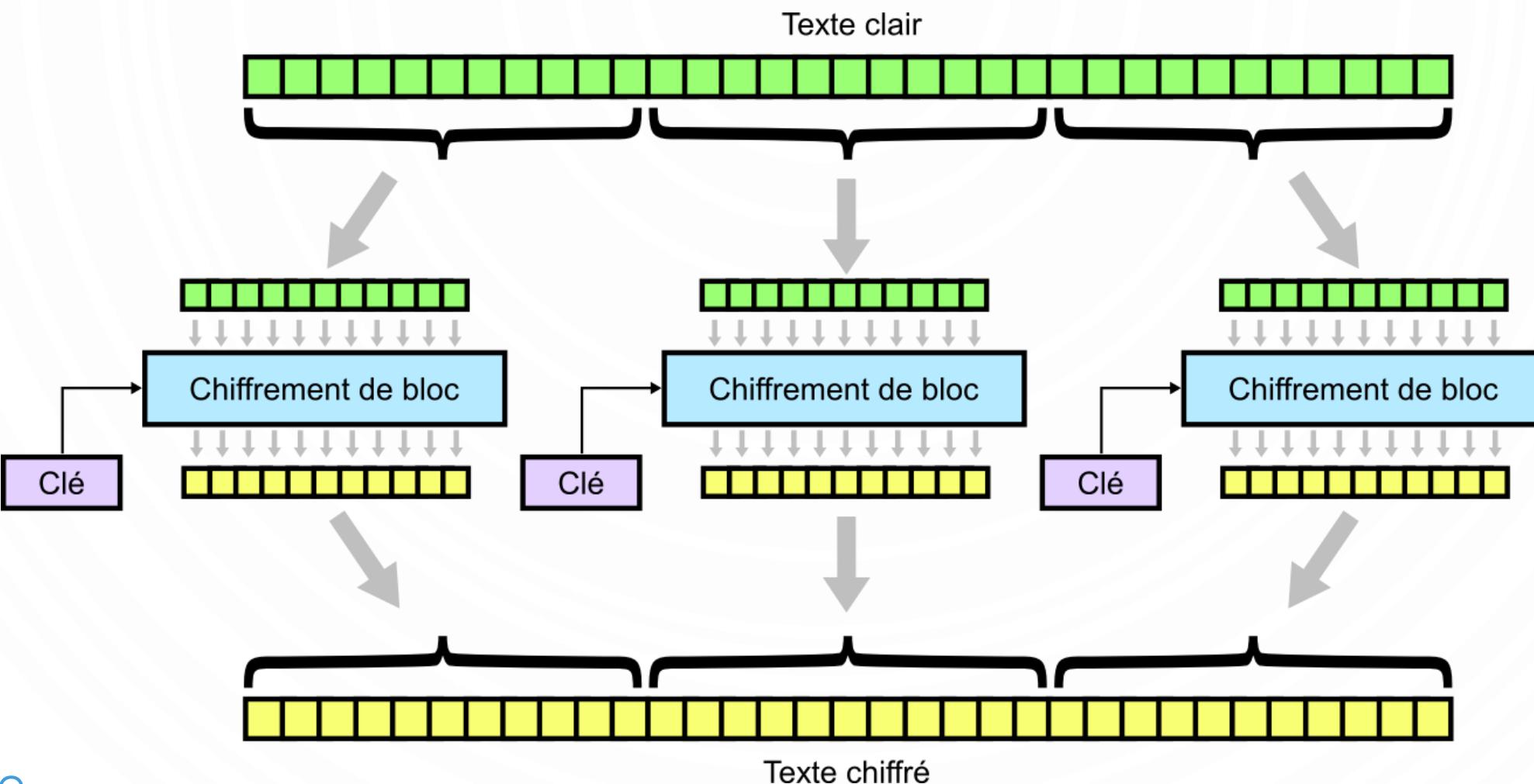
■ Caractéristiques

- Principe : Algorithmes basés sur des opérations en fonction de la clé
- **Transposition/permotion**
 - Propager l'information relative à chaque bit du message en clair dans le message chiffré (**diffusion**)
- **Substitution**
 - Supprimer les relations entre le message en clair et le message chiffré (**confusion**)
- Taille des clés : (standard) 128 bits minimum
- Performances : Très rapide
- Distribution des clés :
 - Très critique
 - Doit s'effectuer de manière sécurisée

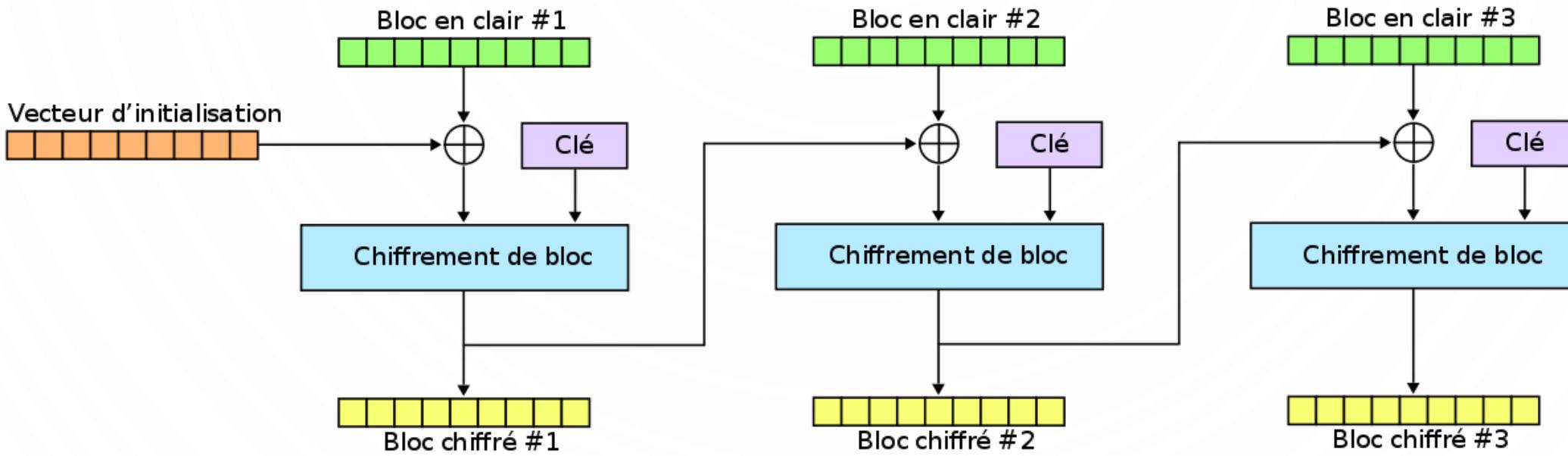
MODES DE CHIFFREMENT

- Généralement défini pour le chiffrement par bloc (peut-être étendu au chiffrement de pixels)
- 5 modes de chiffrement principaux :
 - ECB (« Electronic CodeBook » - Dictionnaire de codes)
 - CBC (« Cipher Block Chaining » - Enchaînement de blocs)
 - CFB (« Cipher FeedBack » - Chiffrement à rétroaction)
 - OFB (« Output FeedBack » - Chiffrement à rétroaction de sortie)
 - CTR (« CounTeR » - Chiffrement basé sur un compteur)

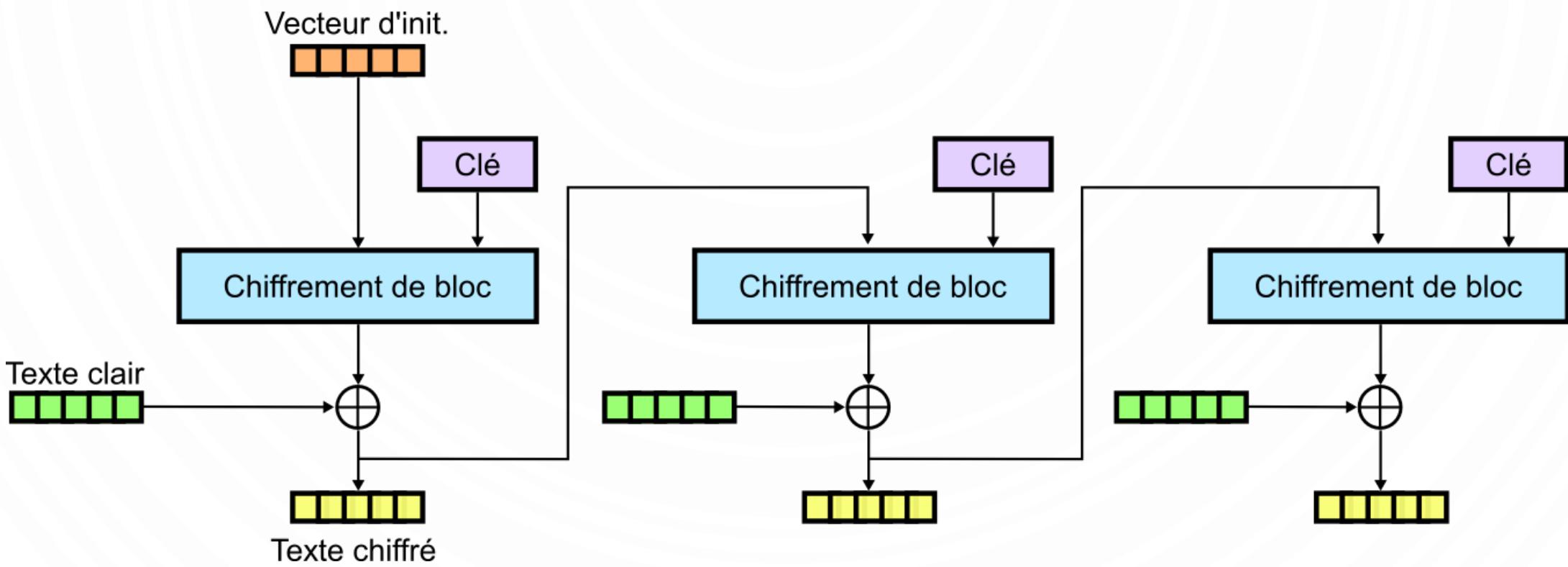
ECB (« ELECTRONIC CODEBOOK »)



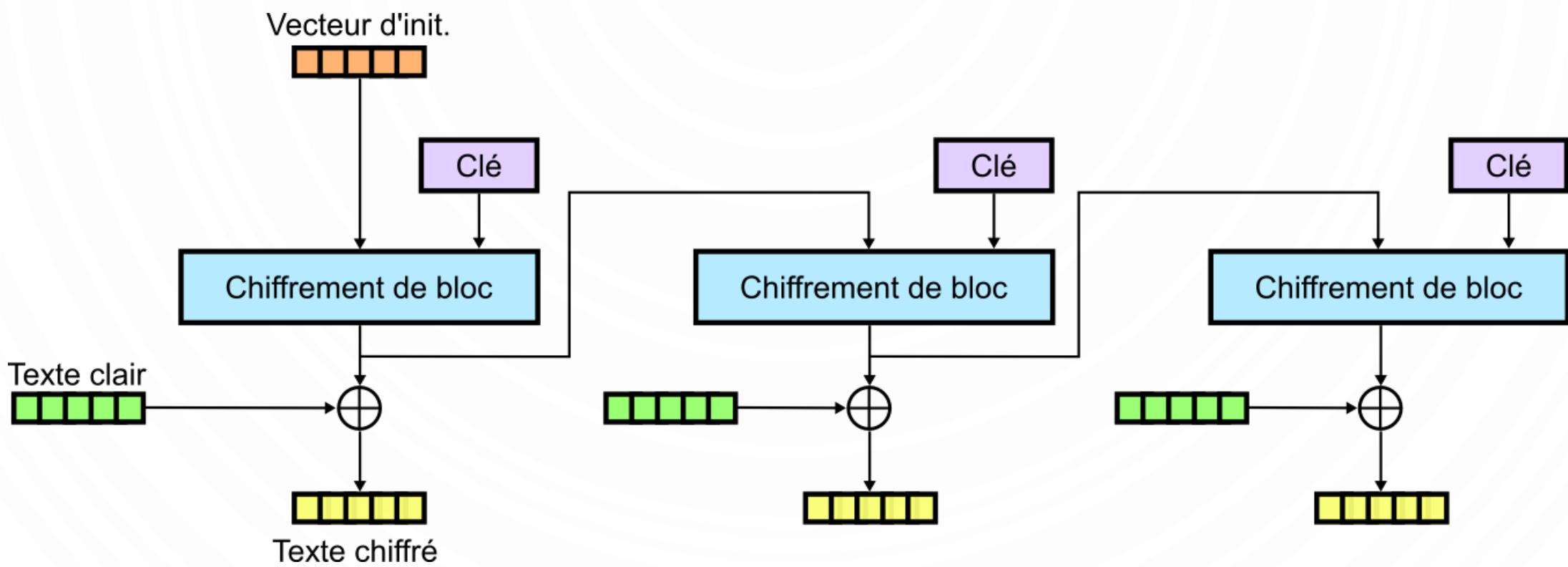
CBC (« CIPHER BLOCK CHAINING »)



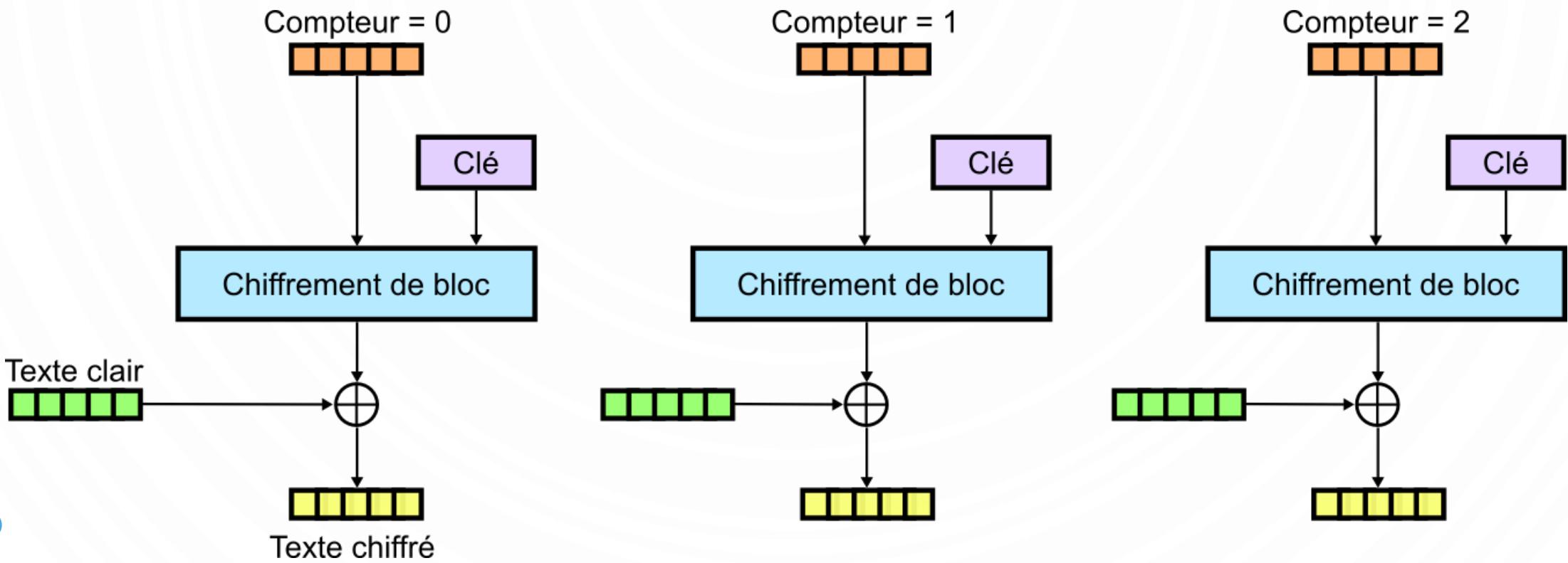
CFB (« CIPHER FEEDBACK »)



OFB (« OUTPUT FEEDBACK »)



CTR (« COUNTER »)



CHIFFREMENT PAR MÉLANGE

- Génération d'une **séquence pseudo-aléatoire** à l'aide d'un **PRNG**
- Clé utilisée comme graine d'initialisation du PRNG
- Utilisation de la séquence pseudo-aléatoire pour **permuter** les pixels (ou les bits) de l'image en clair



$$\sigma = \begin{pmatrix} p(0,0) & p(0,1) & p(0,2) & p(1,0) & p(1,1) & p(1,2) & p(2,0) & p(2,1) & p(2,2) \\ p(2,0) & p(1,2) & p(0,1) & p(2,2) & p(0,2) & p(0,0) & p(1,1) & p(1,0) & p(2,1) \end{pmatrix}$$

Q : Quel est le nombre de permutations possibles ?

CHIFFREMENT PAR XOR

- Génération d'une **séquence pseudo-aléatoire** à l'aide d'un **PRNG**
- Clé utilisée comme graine d'initialisation du PRNG
- Utilisation de la séquence pseudo-aléatoire pour modifier la valeur des pixels de l'image en clair : ou-exclusif entre l'image en clair et la séquence (**substitution**)

Clair	0	1	1	1	0	0	1	1
Séquence	1	0	1	0	0	1	0	1
Chiffré	1	1	0	1	0	1	1	0

Q : Comment déchiffrer le chiffré ?

CHIFFREMENT PAR XOR

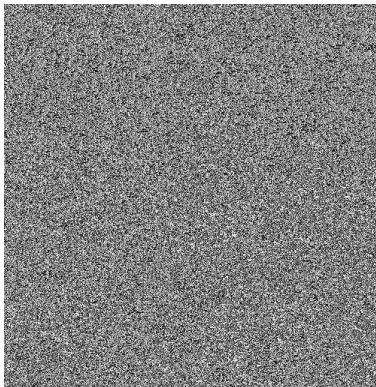
- Utiliser deux fois la même clé ?



I_1



$E_{K_e}(I_1)$



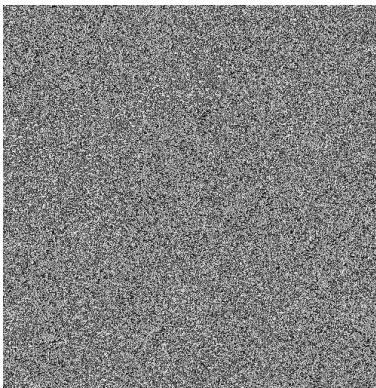
I_{1e}



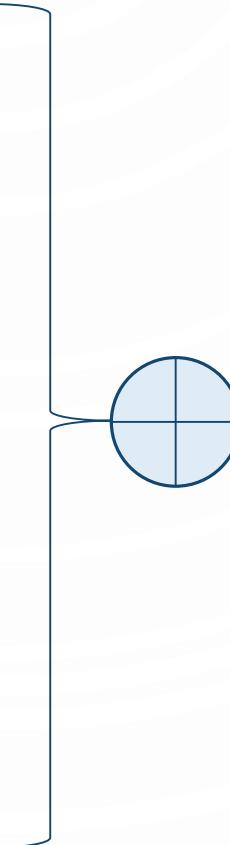
I_2



$E_{K_e}(I_2)$



I_{2e}



CHIFFREMENT PARFAIT

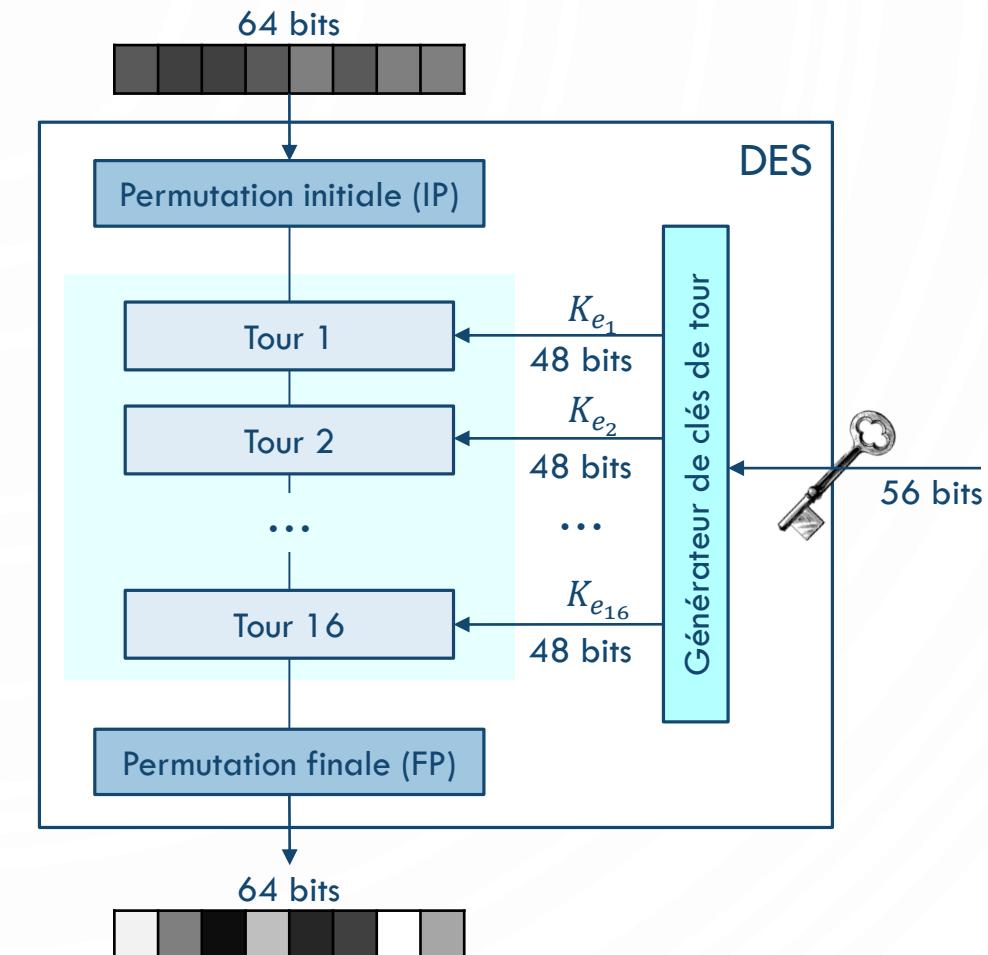
- Chiffre de Vernam (1926) ou « masque jetable »
- Trois impératifs pour la clé :
 - Aussi longue que le message à chiffrer
 - Parfaitement aléatoire
 - Utilisée pour chiffrer **un seul message**, puis **détruite**
- Modèle théorique car très difficile à mettre en place...

Q : Quelle est la probabilité d'apparition d'un niveau de gris dans l'image chiffrée ?

Q : Quelle est la valeur de l'entropie mesurée dans l'image chiffrée ?

DATA ENCRYPTION STANDARD (DES)

- Algorithme de chiffrement symétrique, **par blocs**
- Publié en **1975**, par **IBM**
- Aujourd'hui : considéré comme **non-sécurisé**
- Caractéristiques techniques :
 - Taille de la clé : **56 bits** (+ 8 bits de parité)
 - Taille des blocs : **64 bits**
 - Nombre de tours : **16**



DATA ENCRYPTION STANDARD (DES)

■ Attaques

■ Attaque par force brute possible

- **Diffie-Hellman** en 1977 (US\$ 20M), clé retrouvée en 1 jour → théorique
- **Wiener** en 1993 (US\$ 1M), clé retrouvée en 7h → théorique
- **Electronic Frontier Foundation** en 1998 (US\$ 250k), clé retrouvée en 2 jours → mise en pratique
- **COPACOBANA** (Univ. Bochum & Kiel, en Allemagne) en 2006 (US\$ 10k) → mise en pratique

■ Cryptanalyse différentielle

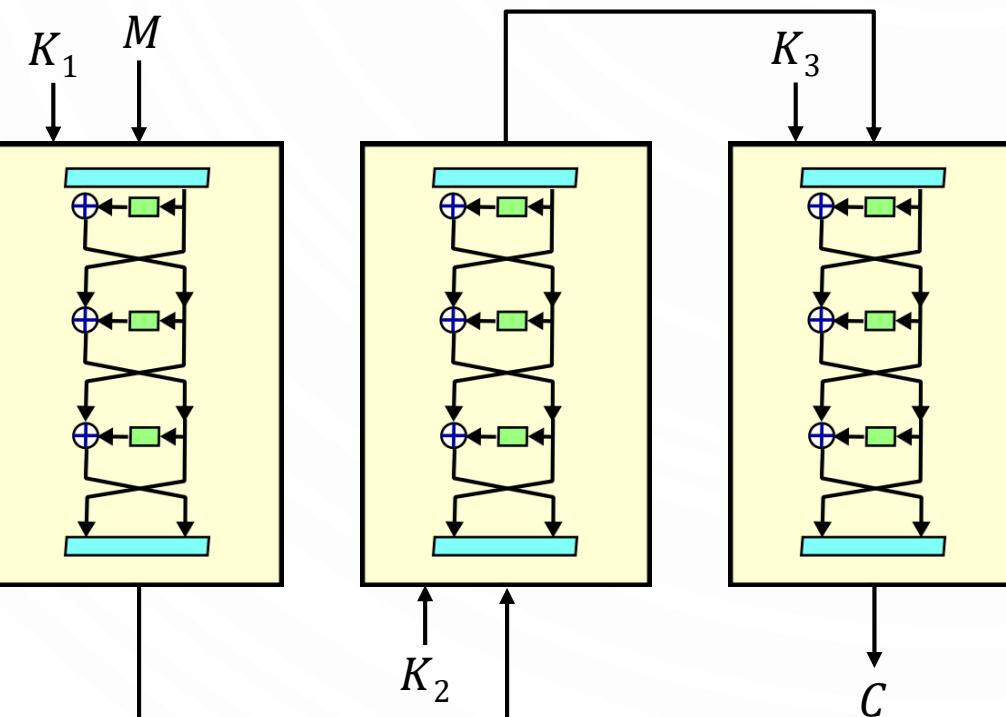
- **Biham-Shamir** en 1980 (CPA –Attaque par 2^{47} clairs choisis)

■ Cryptanalyse linéaire

- **Matsui** en 1994 (KPA –Attaque par 2^{43} clairs connus)
- **Junod** en 2001 (KPA –Attaque par 2^{40} clairs connus)

TRIPLE DES (3DES)

- Publié en **1998**, dérivé de **DES**
- Plus grande taille des clés pour éviter l'attaque par force brute : **168 bits** possible
- Idée : Utiliser 3 clés de **56 bits** chacune



$$C = E_{K_3}(D_{K_2}(E_{K_1}(M)))$$

$$M = D_{K_1}(E_{K_2}(D_{K_3}(C)))$$

TRIPLE DES (3DES)

- Attaque « **Meet-in-the-middle** » sur 2DES (même principe pour 3DES)
 - Attaque **KPA** : M_1, M_2, C_1, C_2 connus tels que $C_1 = E_{K_2}(E_{K_1}(M_1))$ et $C_2 = E_{K_2}(E_{K_1}(M_2))$
 - Chiffrer une fois le message en clair revient à déchiffrer une fois le message chiffré
 - On a donc : $E_{K_1}(M) = D_{K_2}(C)$
 - Calcul et stockage des 2^{56} couples $(K, E_K(M_1))$
 - Pour chaque clé K' , calcul de $D_{k'}(M_1)$ et recherche de correspondance
 - Si couple de clés candidates $K_1 = K$ et $K_2 = K'$, vérification $C_2 = E_{K_2}(E_{K_1}(M_2))$
 - Nombre maximal d'essais : $2 \times 2^{56} = 2^{57} \ll 2^{112}$

ADVANCED ENCRYPTION STANDARD (AES)

- Algorithme de chiffrement symétrique, **par blocs**
- Gagnant d'un concours lancé en 1997
- Standard depuis **2001** (NIST)
- Vrai nom : **Rijndael** → Créé par deux belges **Joan Daemen et Vincent Rijmen**



ADVANCED ENCRYPTION STANDARD (AES)

■ Pourquoi un nouveau standard ?

- DES est devenu attaquantable par force brute
- Développement de systèmes d'évaluation : analyse différentielle et linéaire
- Possible d'avoir une méthode de chiffrement plus rapide en utilisant des instructions processeur

■ Pourquoi un concours public ?

- Rassembler la communauté travaillant sur la cryptographie
- Encourager la recherche autour des systèmes sécurisés
- Prévenir les « backdoors »
- Accélérer l'acceptation et l'adoption d'un standard

ADVANCED ENCRYPTION STANDARD (AES)

■ Description générale

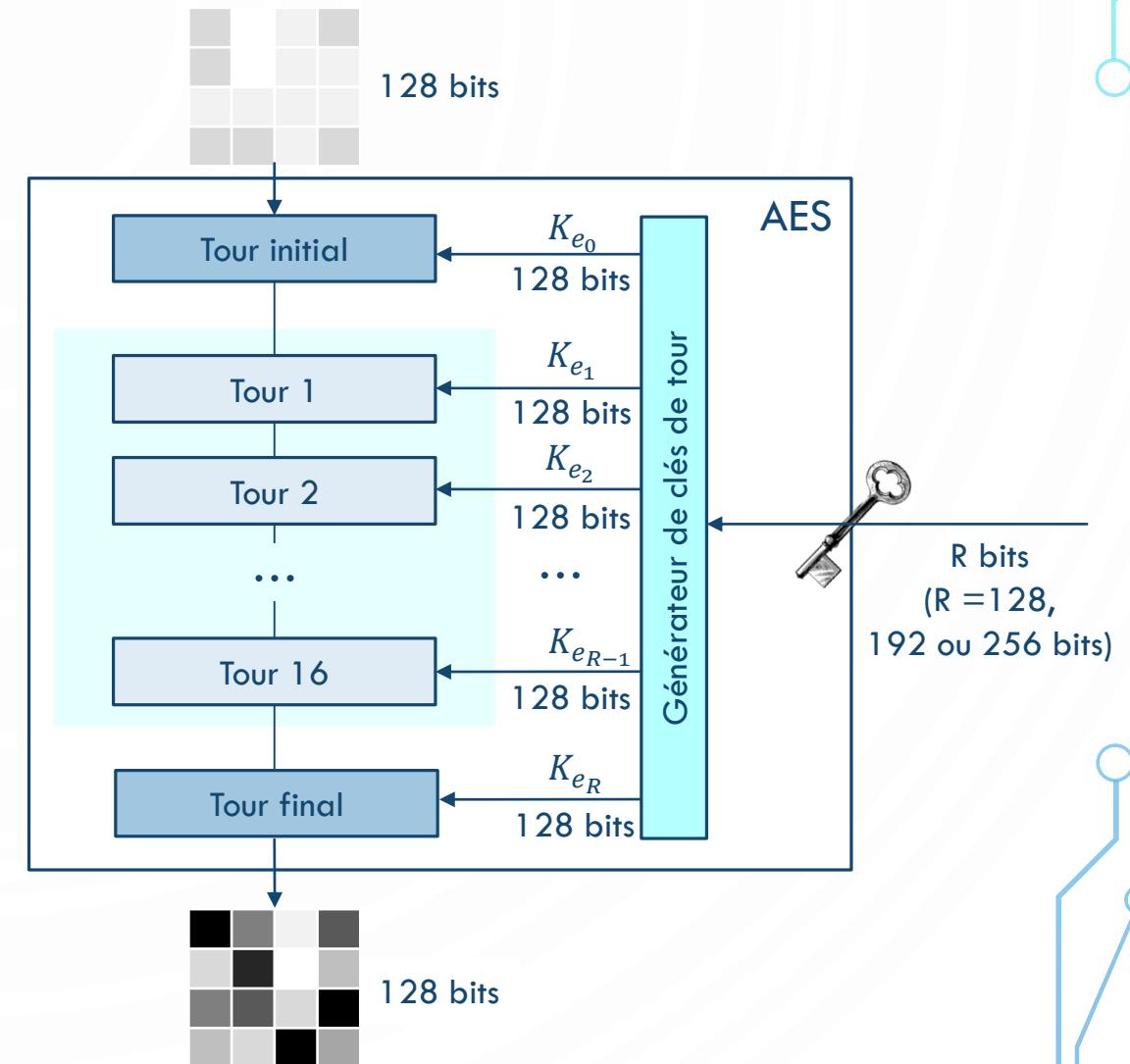
- Nombre de tours : **10, 12 ou 14** (suivant la taille de la clé)
- Chaque tour : **4 opérations**
- Taille des blocs du message : **128 bits** (4 colonnes de 4 octets)
- Taille de la clé de chiffrement : **128, 192 ou 256 bits**

$p(0,0)$	$p(0,1)$	$p(0,2)$	$p(0,3)$
$p(1,0)$	$p(1,1)$	$p(1,2)$	$p(1,3)$
$p(2,0)$	$p(2,1)$	$p(2,2)$	$p(2,3)$
$p(3,0)$	$p(3,1)$	$p(3,2)$	$p(3,3)$

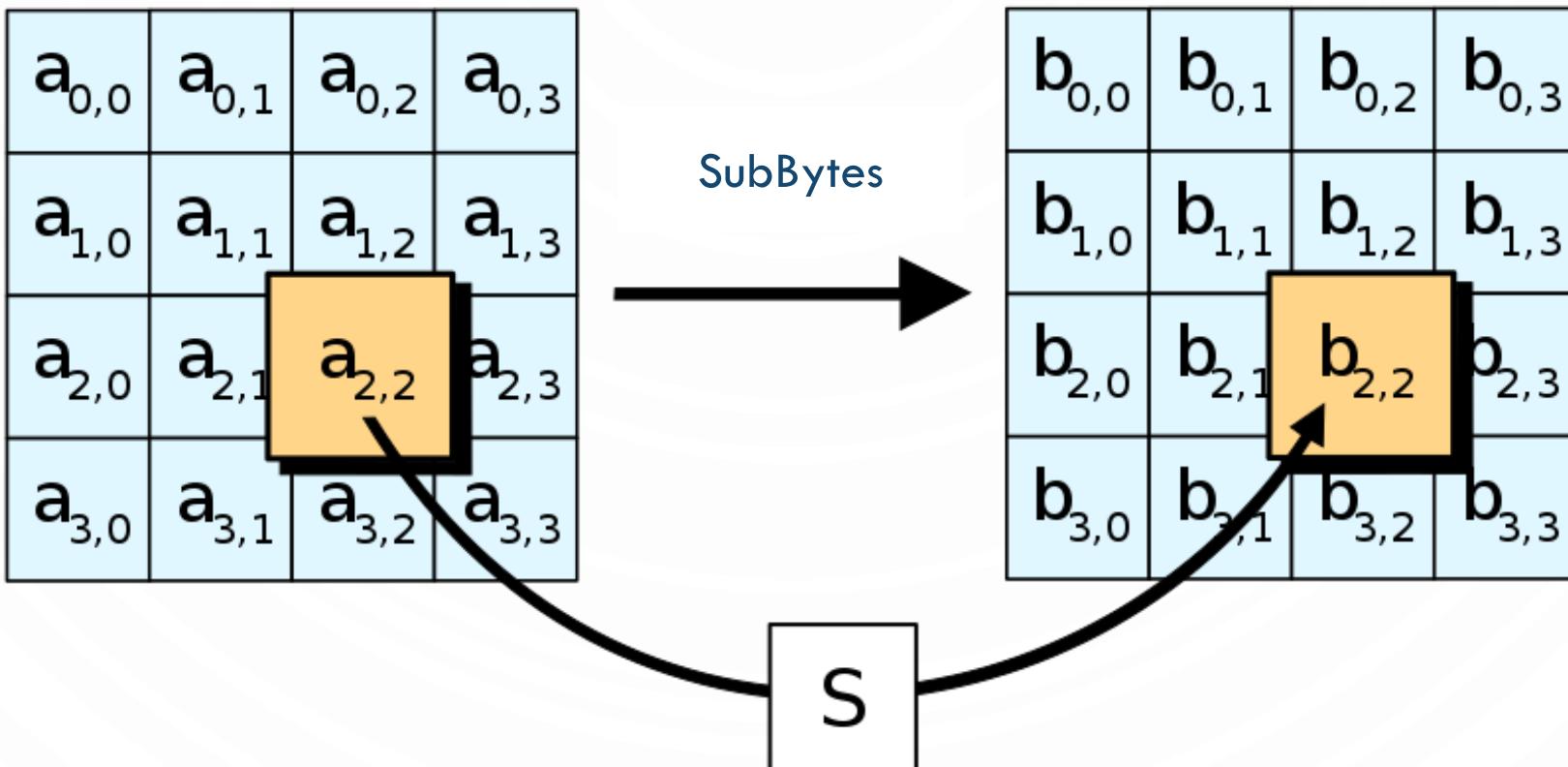
ADVANCED ENCRYPTION STANDARD (AES)

■ Description générale

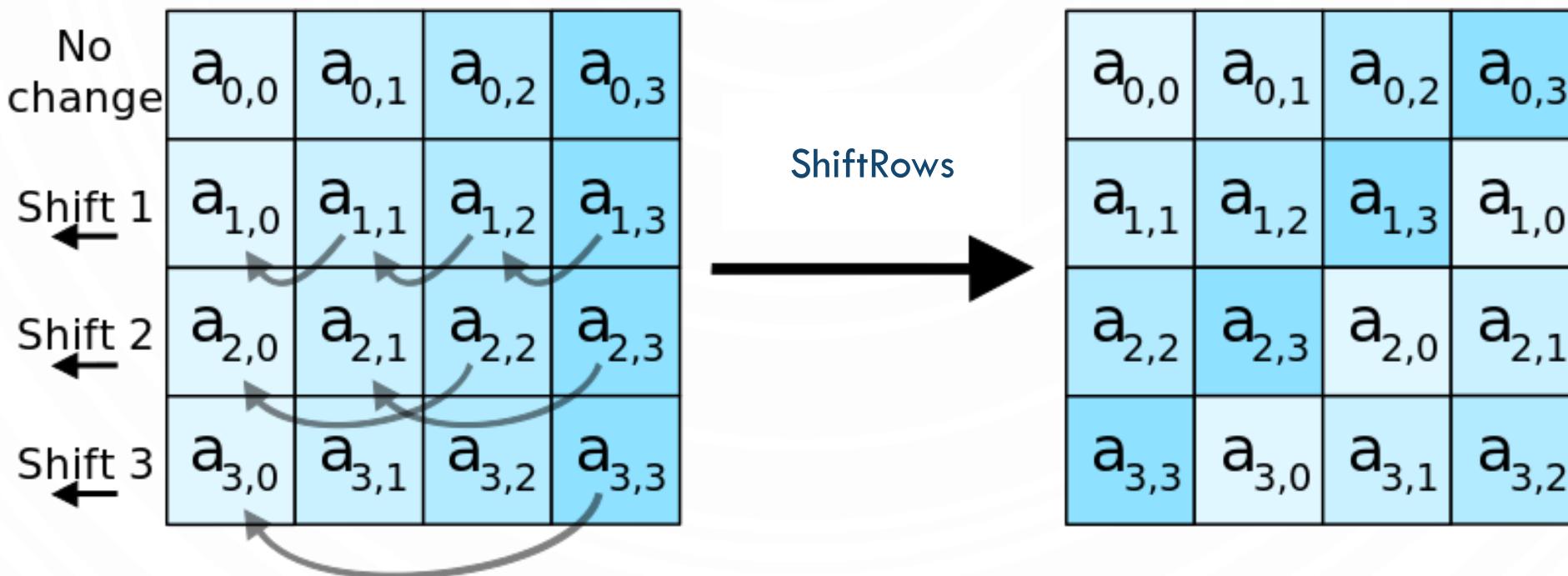
1. KeyExpansion
2. Tour initial
 1. AddRoundKey
3. Pour chaque tour suivant
 1. SubBytes
 2. ShiftRows
 3. MixColumns
 4. AddRoundKey
4. Tour final (pas de MixColumns)
 1. SubBytes
 2. ShiftRows
 3. AddRoundKey



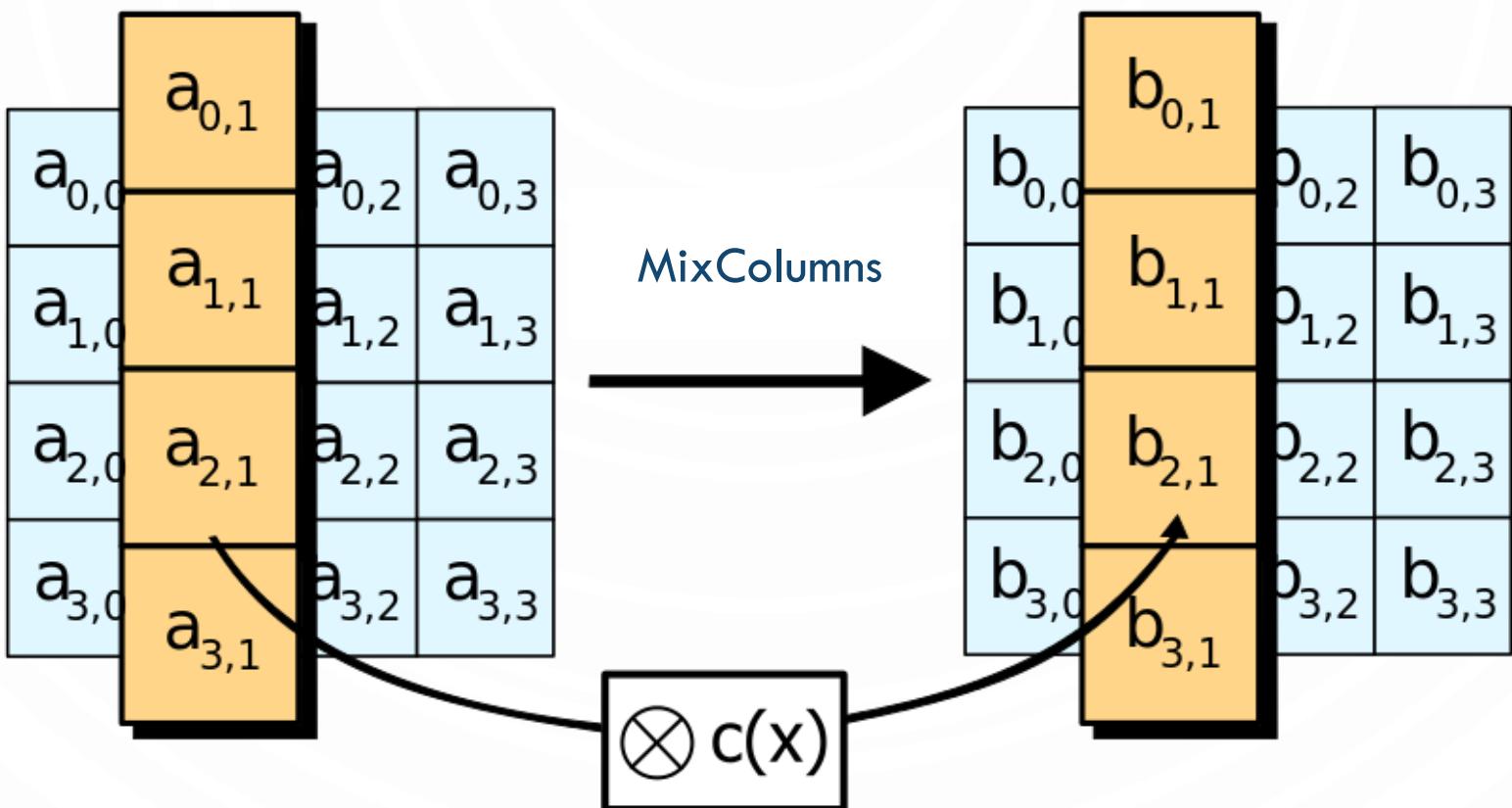
ADVANCED ENCRYPTION STANDARD (AES)



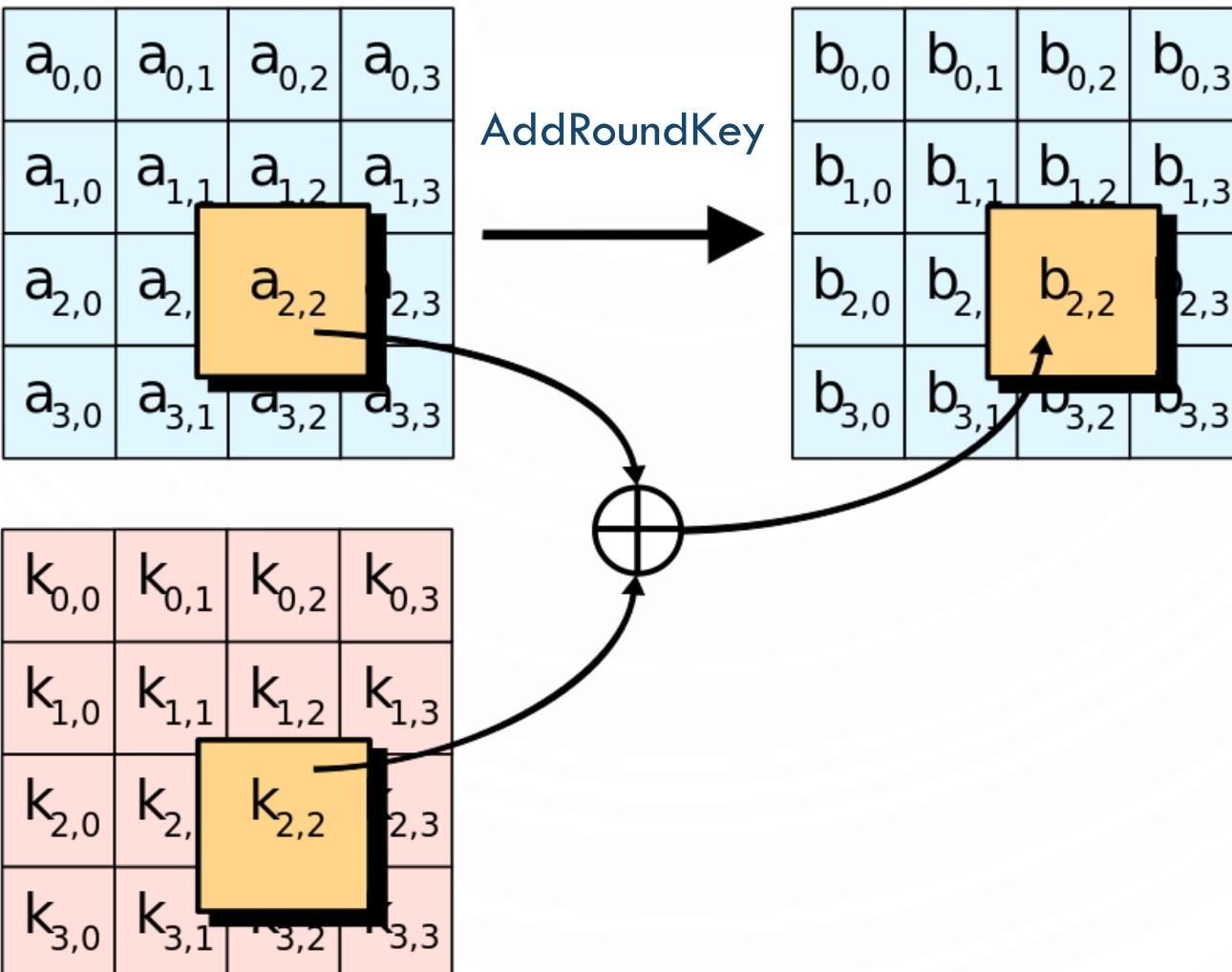
ADVANCED ENCRYPTION STANDARD (AES)



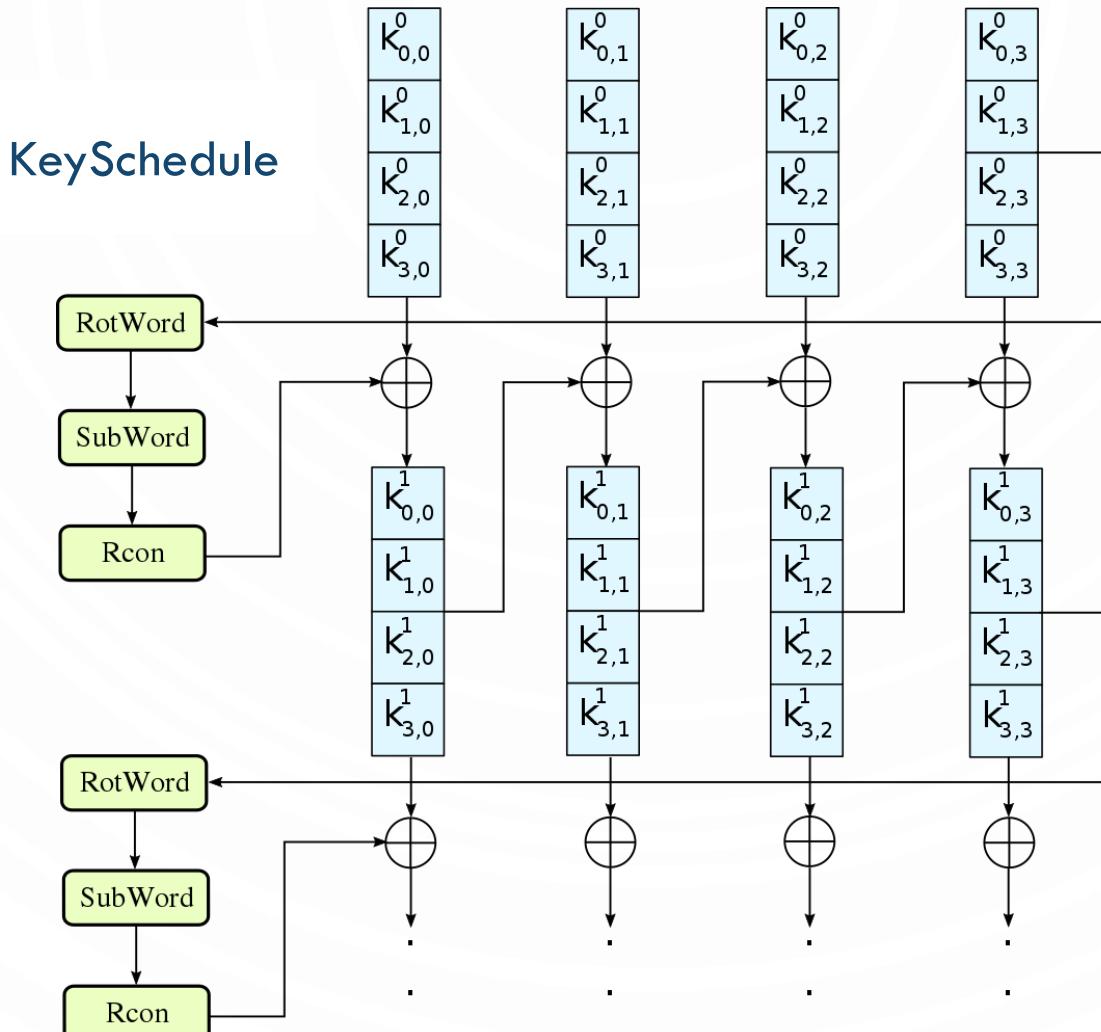
ADVANCED ENCRYPTION STANDARD (AES)



ADVANCED ENCRYPTION STANDARD (AES)



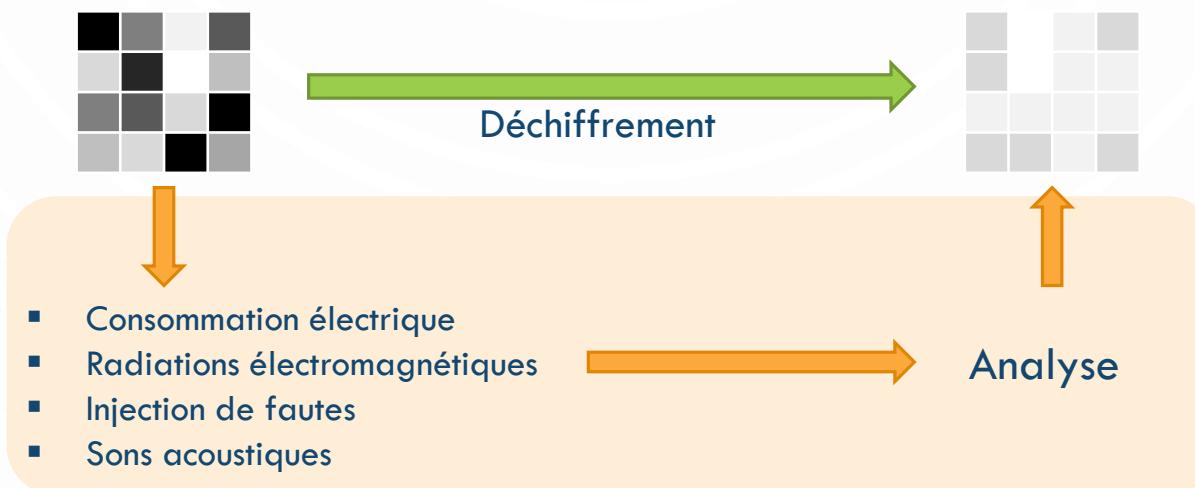
ADVANCED ENCRYPTION STANDARD (AES)



ADVANCED ENCRYPTION STANDARD (AES)

■ Attaque par canal auxiliaire

- Recherche et exploitation des failles dans l'implémentation, logicielle ou matérielle
- Ne remet pas en cause la robustesse théorique des méthodes et procédures de sécurité
- Une sécurité « mathématique » ne garantit pas forcément une sécurité lors de l'utilisation en « pratique »



- Beaucoup d'attaques publiées **non faisables** en pratique (2013)
- **Considéré sûr à ce jour**

CRYPTOGRAPHIE ASYMÉTRIQUE

- Caractéristiques

- Une clé privée K_{priv} et une clé publique K_{pub}
- Propriétés :
 - La connaissance de la clé publique K_{pub} ne permet pas de déduire la clé privée K_{priv}
 - $D_{K_{priv}}(E_{K_{pub}}(M)) = M$

- Principe : **Fonction unidirectionnelle à trappe**

- « Facile » à calculer dans un sens, « difficile » à inverser
- Sauf si on connaît une information secrète (**la trappe**)
- Algorithmes basés sur des opérations d'exponentiation en algèbre modulaire

CRYPTOGRAPHIE ASYMÉTRIQUE

■ Caractéristiques

■ Génération des clés :

- A partir de grands nombres premiers $K_{pub} = f(K_{priv})$
- Calcul de $K_{priv} = f^{-1}(K_{pub})$ impossible
- Taille des clés : **512 bits ou 1024 bits**
- Performances : **1000 fois plus lents** que les algorithmes symétriques !
- Nombre de clés : autant de paires que d'entités
- Distribution des clés : Facilitée car **pas d'échange** de clés secrètes
 - Clé secrète conservée par les entités
 - Clé publique échangée

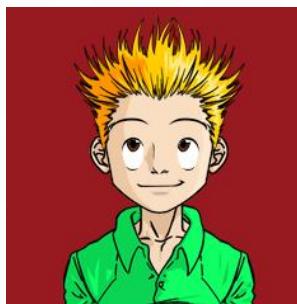
PROTOCOLE DE DIFFIE-HELLMAN (1976)



Alice



Bob

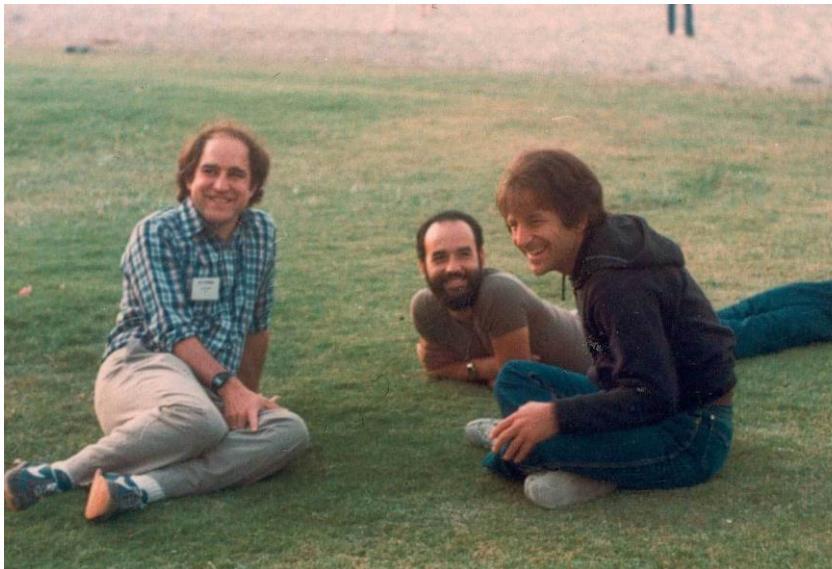


- 1) Alice et Bob choisissent un grand nombre premier p et d'un entier $1 \leq a < p$
- 2) Alice choisit secrètement x_A
- 3) Alice calcule $y_A = a^{x_A} \pmod{p}$
- 4) Alice et Bob s'échangent les valeurs de y_A et y_B
- 5) Alice calcule $y_B^{x_A} = (a^{x_B})^{x_A} = a^{x_B x_A} \pmod{p} = K$
- 2) Bob choisit secrètement x_B
- 3) Bob calcule $y_B = a^{x_B} \pmod{p}$
- 5) Bob calcule $y_A^{x_B} = (a^{x_A})^{x_B} = a^{x_A x_B} \pmod{p} = K$

Q : Sur quoi repose la sécurité de l'échange des clés ?

RIVEST-SHAMIR-ADLEMAN (RSA)

- Crée en **1977** par Ron Rivest, Adi Shamir et Leonard Adleman
- Breveté par le MIT en **1983**
- Basé sur le problème de la **factorisation des grands nombres entiers**



RIVEST-SHAMIR-ADLEMAN (RSA)

■ Génération du couple de clés (K_{pub}, K_{priv})

- Choisir p et q , deux nombres premiers distincts
- Calculer leur produit $n = pq$ (**module de chiffrement**)
- Calculer $\varphi(n) = (p - 1)(q - 1)$
- Choisir un nombre e premier avec $\varphi(n)$ et strictement inférieur à ce nombre (**exposant de chiffrement**)
- Calculer l'entier naturel d , inverse de e modulo $\varphi(n)$ (**exposant de déchiffrement**)
- On a $K_{pub} = (e, n)$ et $K_{priv} = d$

Q : Quel algorithme utilise t-on pour calculer d , l'inverse de e modulo $\varphi(n)$?

RIVEST-SHAMIR-ADLEMAN (RSA)

- L'algorithme de chiffrement/déchiffrement

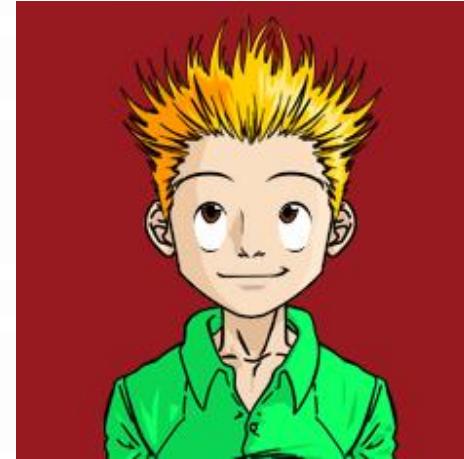
Alice



$$K_{pub}(A) = (e_A, n_A)$$
$$K_{priv}(A) = d_A$$

$$C = M^{e_B} \pmod{n_B}$$

Bob



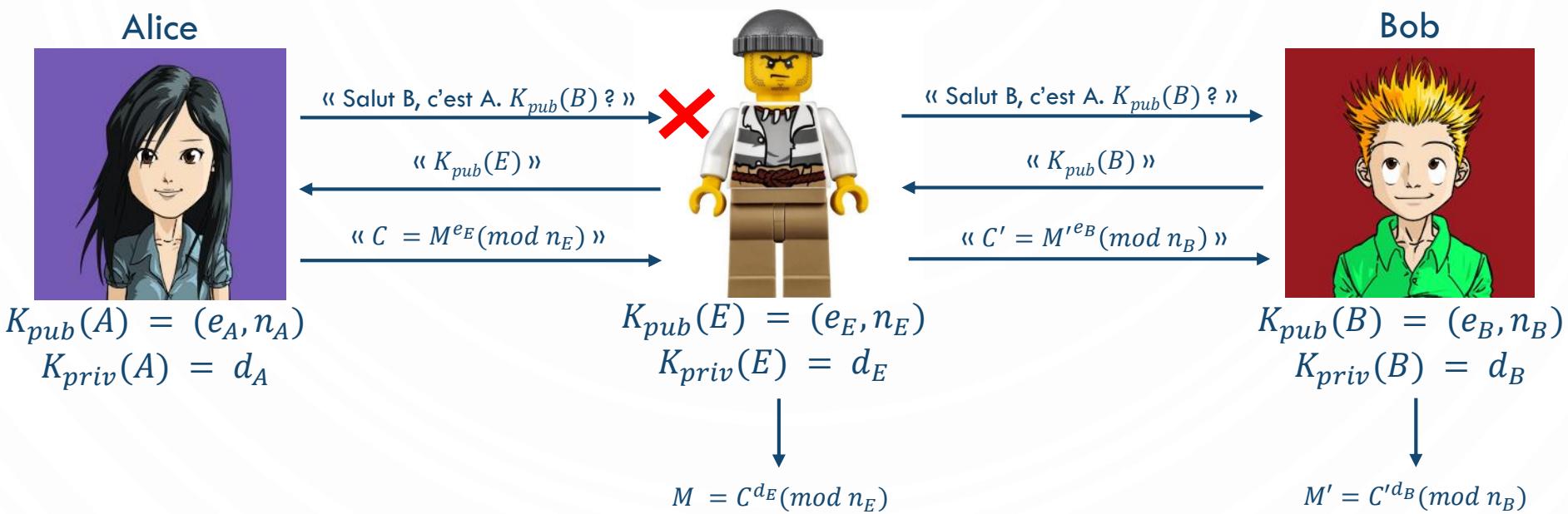
$$K_{pub}(B) = (e_B, n_B)$$
$$K_{priv}(B) = d_B$$

$$M = C^{d_B} \pmod{n_B}$$
$$M = (M^{e_B})^{d_B} \pmod{n_B}$$

Q : Comment calculer efficacement l'exponentiation modulaire ?

RIVEST-SHAMIR-ADLEMAN (RSA)

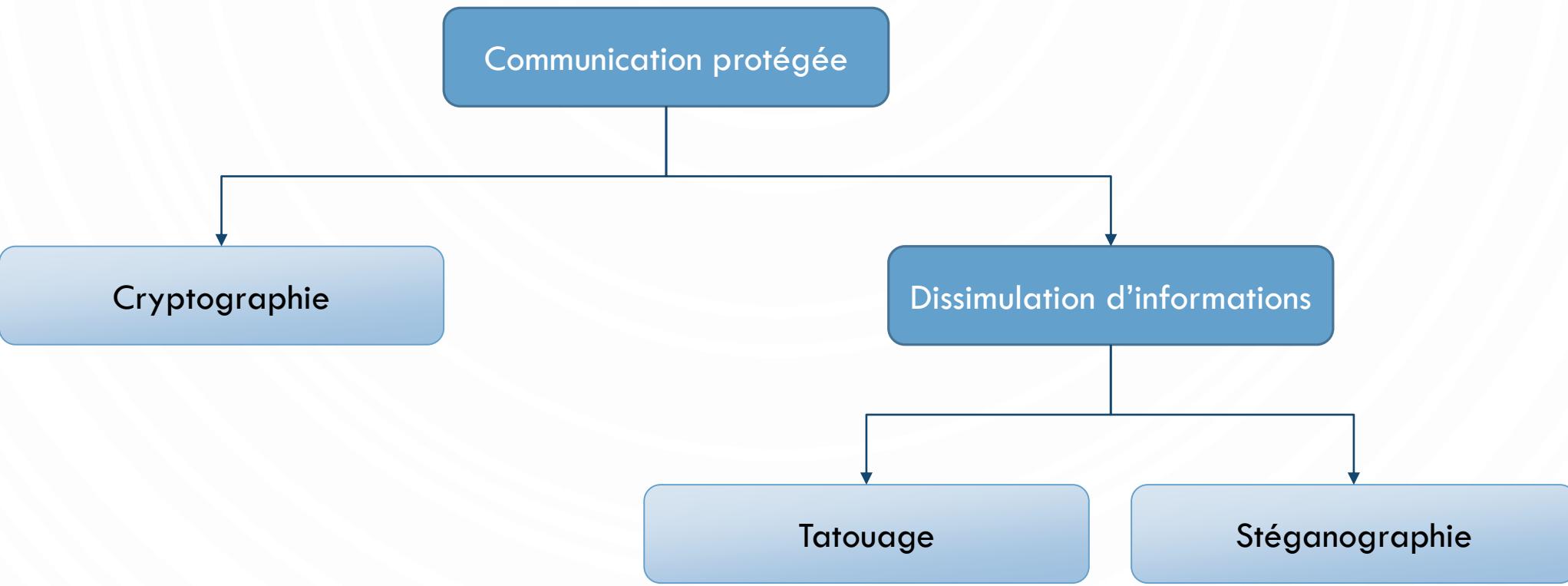
■ Attaque de « l'homme du milieu »



Par exemple, M = « Viens me chercher à la gare » et M' = « Viens me chercher au stade »

Q : Comment remédier à ce problème ?

TECHNIQUES DE PROTECTION DE DONNÉES



DÉFINITION DU TATOUAGE

- Art d'altérer un média (texte, image, vidéo...) de sorte qu'il contienne un message le plus souvent **en rapport avec le média et de manière imperceptible et robuste**



Tatouage visible



Tatouage invisible

DÉFINITION DE LA STÉGANOGRAPHIE

- Art de la communication secrète
- Dissimuler un message secret dans un **médium anodin** (image, vidéo, son...) **de sorte qu'il ne puisse être détecté** (visuellement, mais aussi **statistiquement**)



Image hôte

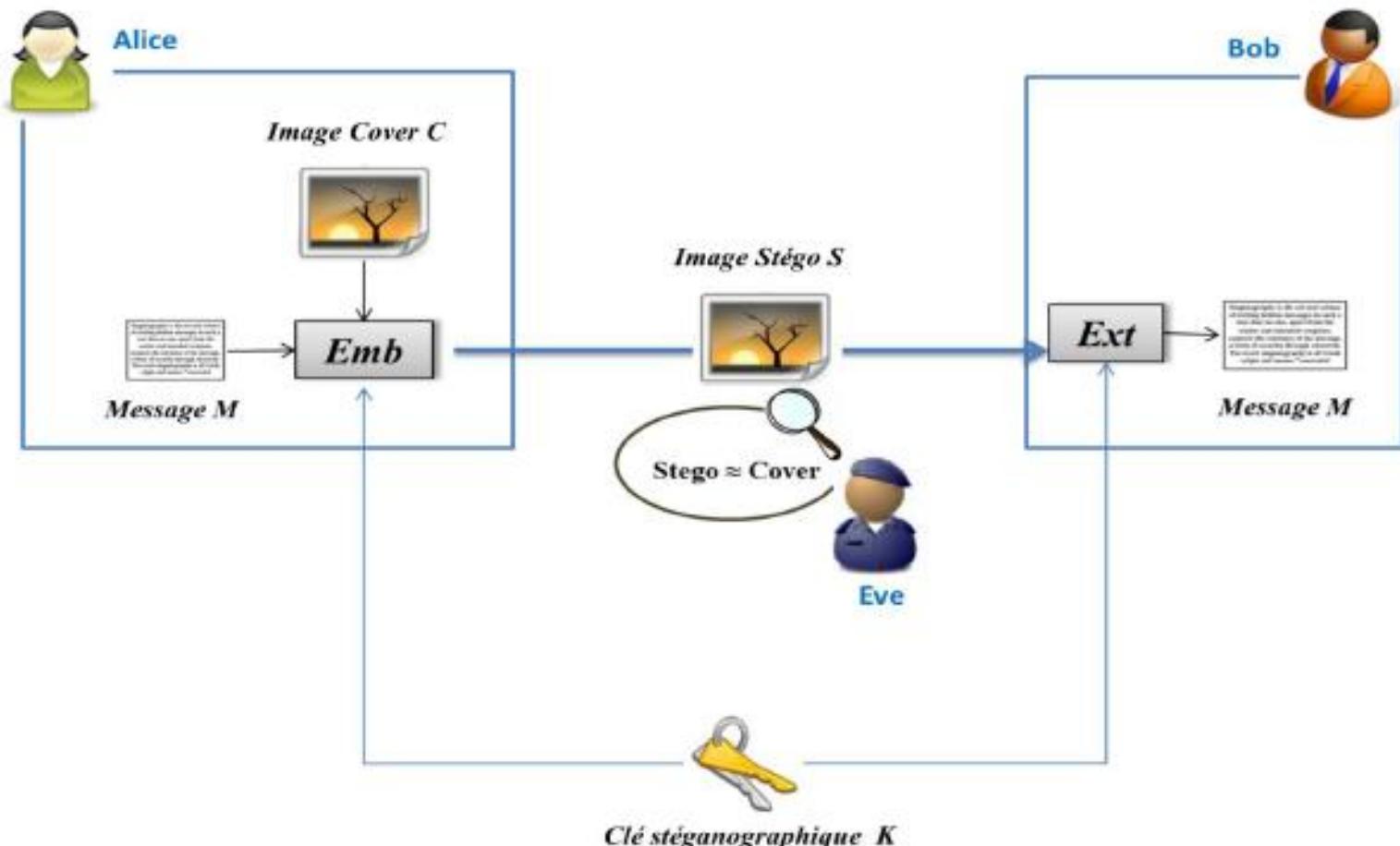


Image stego

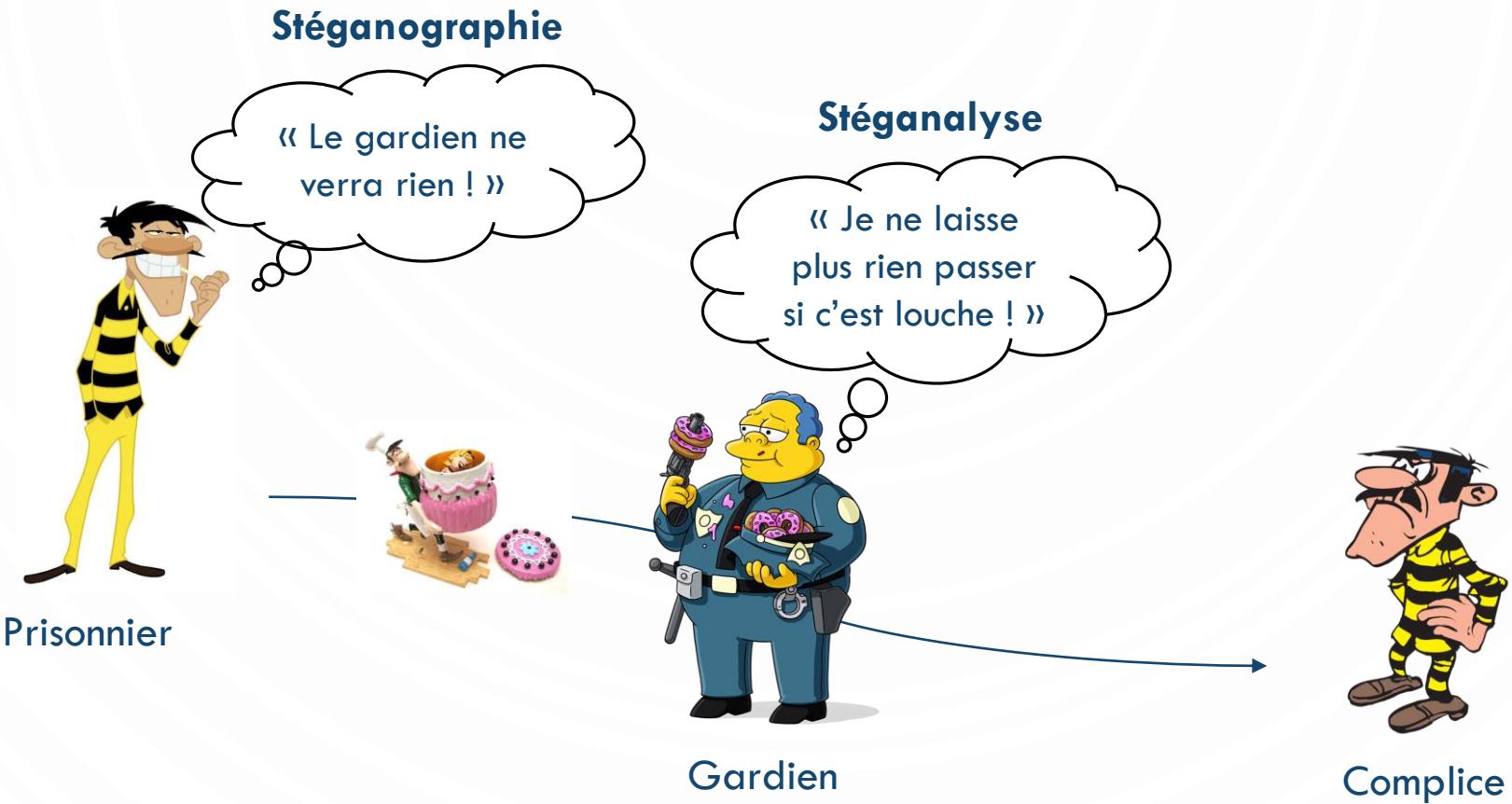
SCHÉMAS STÉGANOGRAPHIQUES

- Sélection du médium de couverture
 - + quasiment indétectable, facile à implémenter, complexité faible
 - - capacité d'insertion limitée
- Synthèse de médium de couverture
 - + parfaitement sûre
 - - très théorique, difficile à mettre en œuvre
- Modification du médium de couverture
 - + très efficace, plus pratique à mettre en place

MODIFICATION DU MÉDIUM DE COUVERTURE



STÉGANOGRAPHIE VS STÉGANALYSE



STÉGANALYSE

- Déetecter la présence de données dissimulées à l'aide d'un algorithme stéganographique
- Discipline duale de la stéganographie
- Différents types de gardien
 - Passif : observe le trafic entre Alice et Bob
 - Actif : essaye d'apporter des modifications sur le médium (compression, filtrage) pour détruire le processus stéganographique s'il existe
 - Malicieux : essaye de comprendre la technique stéganographique et extraire le message, pour le contourner pour ses propres fins

SÉCURITÉ D'UN SCHÉMA STÉGANOGRAPHIQUE

- Théorie (Cachin 1998) : comparaison de la distribution du support avant et après l'insertion en utilisant la divergence de Kullback Leibler

$$D_{KL}(P_c || P_s) = \sum_{x \in C} P_c(x) \log_2 \frac{P_c(x)}{P_s(x)}$$

- Si $D_{KL}(P_c || P_s) = 0$, parfaitement sûr
- Si $D_{KL}(P_c || P_s) < \varepsilon$, dit ε -sûr

SÉCURITÉ D'UN SCHÉMA STÉGANOGRAPHIQUE

- Pratique : définir ce qu'est une distribution peut être difficile
- Un attaquant ne dispose que d'une approximation de ces distributions (puissance limitée : ressources matérielles, temps et capacité de calcul)
- Plusieurs modèles proposés suivant le type de schéma et adversaire (passif, actif, malicieux...)

SÉCURITÉ D'UN SCHÉMA STÉGANOGRAPHIQUE

- Règles de base :

- S'assurer que le support de couverture est utilisé une seule fois, et qu'il est détruit dès son utilisation, afin d'éviter toute attaque par différence
- Vérifier que la taille de la clé est assez grande (éviter recherche exhaustive)
- S'assurer que le processus de dissimulation est imperceptible à l'œil nu (attaque visuelle)
- Essayer de préserver les propriétés statistiques originales du support hôte

PROPRIÉTÉS

- Capacité d'insertion : nombre maximal de bits pouvant être cachés dans le médium de couverture (souvent en bpp)
- Efficacité d'insertion : nombre de bits insérés par unité de distorsion
 - « Moins il y a de modifications, moins le schéma est détectable » ? FAUX !
- Capacité stéganographique : nombre maximal de bits pouvant être modifiés dans le médium de couverture pour que la probabilité de détection soit insignifiante.
 - Bien plus petite que la capacité d'insertion
 - L'estimer est très difficile
 - Manque de modèles statistiques précis représentant les images naturelles

EXEMPLE : SUBSTITUTION DES LSB

- Dans le domaine spatial, substituer les LSB des pixels par les bits du message à insérer
- Sens du parcours des pixels défini par une clé secrète

Message (M): 0 0 1 1 0 1 0 0

Image en niveau de gris

01001011	00101011	10010100	11101000
10011111	01110110	00110111	00110101

Message (M): 0 0 1 1 0 1 0 0

Image en niveau de gris

01001010	00101010	10010101	11101001
10011110	01110111	00110110	00110100