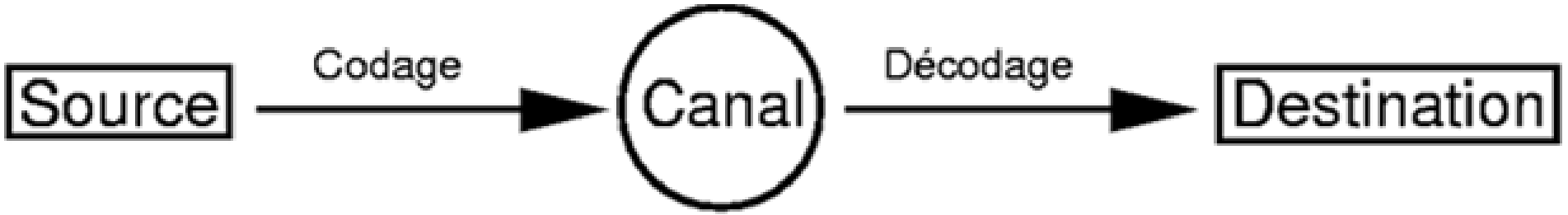
CHAPITRE 2

INTRODUCTION A LA CRYPTOGRAPHIE

Inspiré du document: **Cryptographie & Algorithmique de la Sécurité** **Jean-Guillaume Dumas** 2012

Notion de Code



- Le code doit répondre à différents critères :
  - Sécurité de l'information : cryptage, authentification, etc.
  - Rentabilité : compression des données
  - Tolérance aux fautes : correction/détection d'erreurs

La cryptologie

⇨ Science du secret

- Cryptographie : étude et conception des procédés de chiffrement des informations
  - Stéganographie (στεγανο-γραφην) : écriture couverte
    Dissimule l'existence même d'information secrète
  - Cryptographie (κρνπτο- γραφην) : écriture secrète
    Transforme un message clair en cryptogramme

- Cryptanalyse : analyse des textes chiffrés pour retrouver l'information dissimulée

Cryptographie

- Information chiffrée

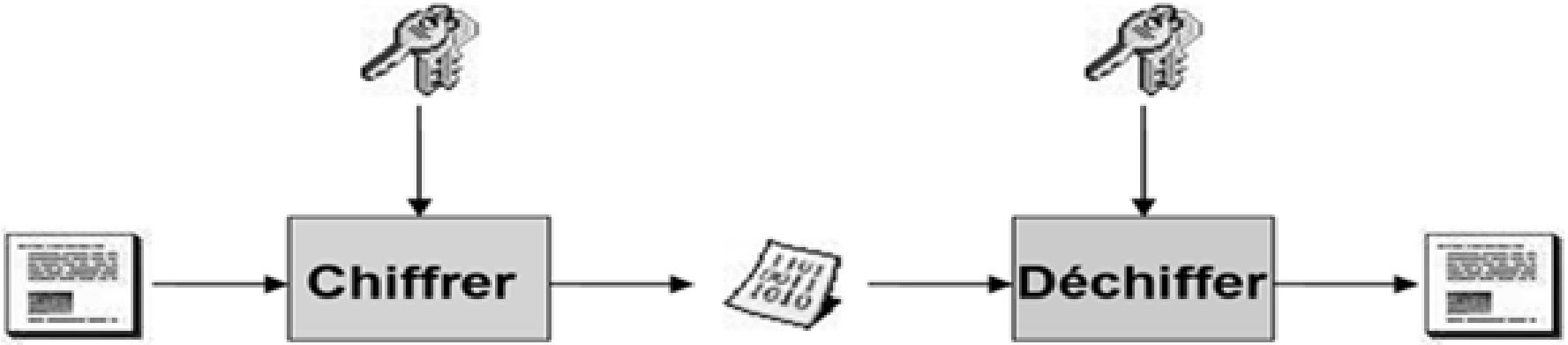
  Connaissance de l'existence de l'information

  ≠

  Connaissance de l'information

- Objectif
  - Permettre à Alice et Bob de communiquer sur un canal peu sûr
    - Réseau informatique, téléphonique, etc.
  - Oscar ne doit pas comprendre ce qui est échangé

Algorithmes de cryptographie



- Propriétés théoriques nécessaires :

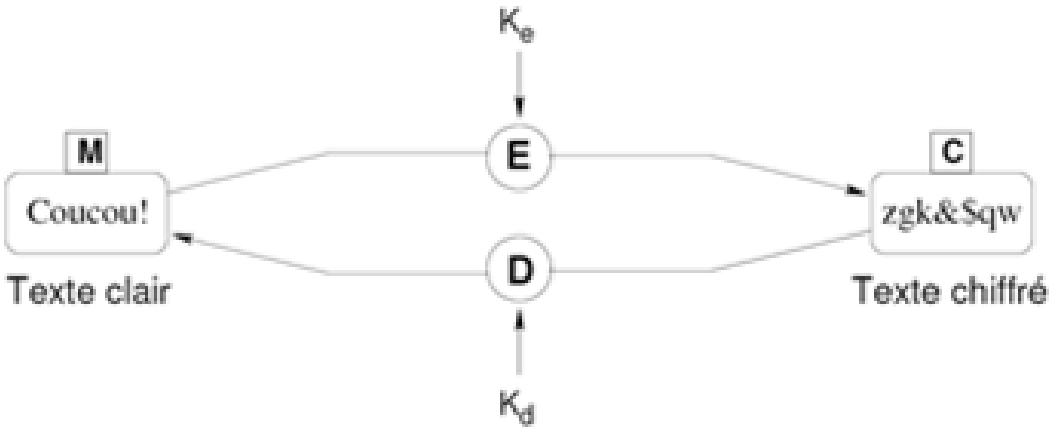
1. Confusion

Aucune propriété statistique ne peut être déduite du message chiffré

2. Diffusion

Toute modification du message en clair se traduit par une modification complète du chiffré

Types de cryptographie

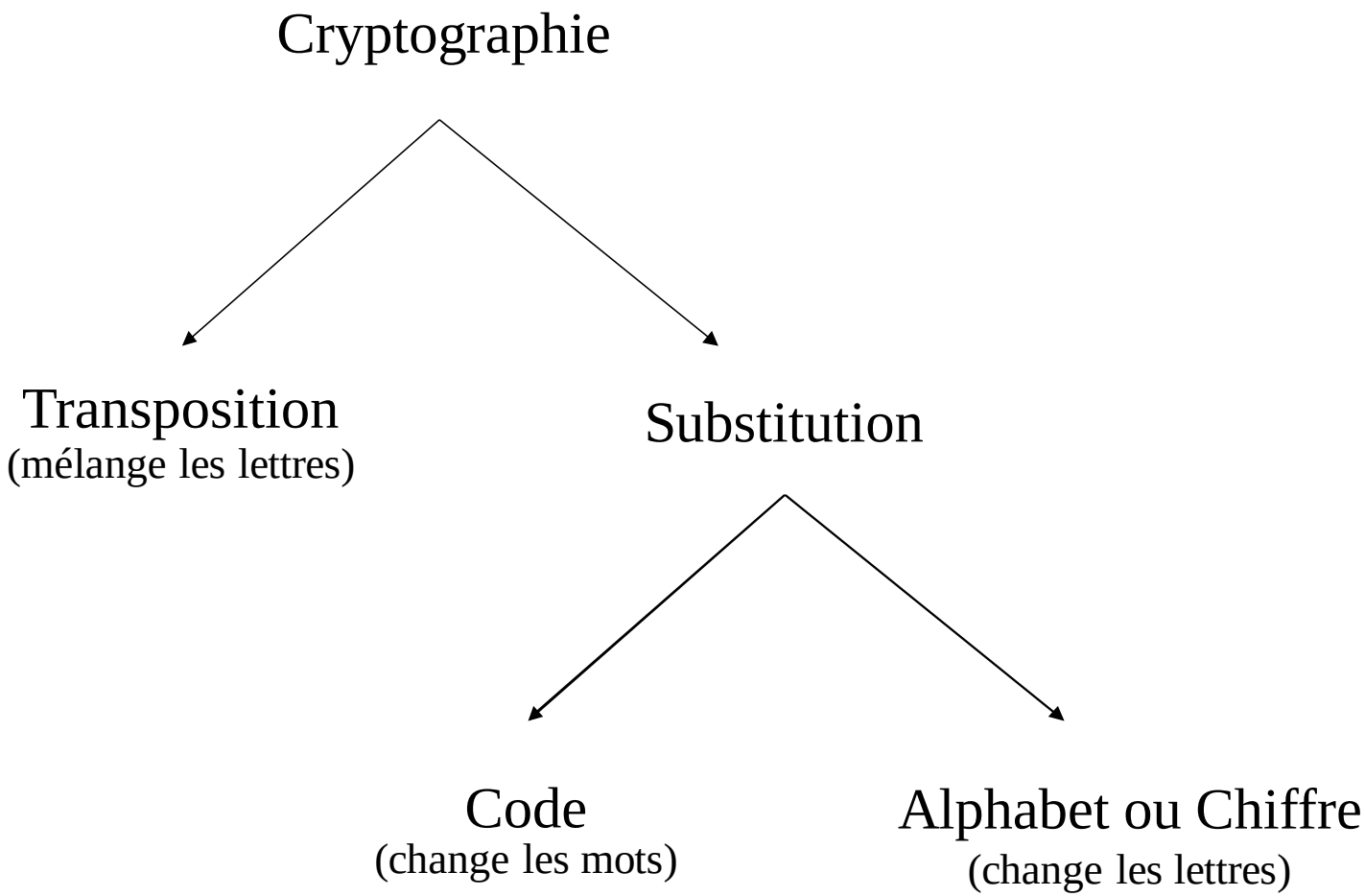


- En pratique E et D sont paramétrées par des clefs K_d et K_e
- Deux grandes catégories de systèmes cryptographiques
  - Systèmes à clefs secrètes (symétriques) : $K_e=K_d=K$
  - Systèmes à clefs publiques (asymétriques) : $K_e \neq K_d$
- Deux types de fonctionnement
  - Par flot : chaque nouveau bit est manipulé directement
  - Par bloc : chaque message est découpé en blocs

Histoire des codes secrets

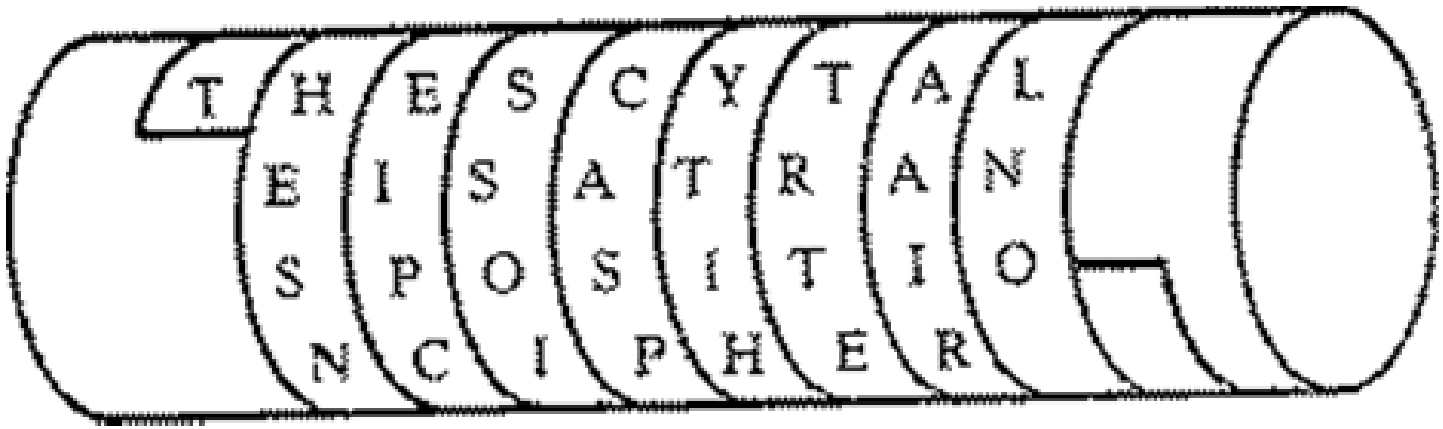
- Cryptographie Ancienne
  - Transposition Sparte (5ème siècle av JC)
  - Substitution César (1er siècle av JC), Vigenère (XVIème siècle),
- Cryptanalyse des codes mono et poly alphabétiques
  - El Kindi (IXème siècle)
  - Babbage/Kasiski (XIXème siècle)
- Mécanisation de la cryptographie et de la cryptanalyse
  - Enigma
  - Les bombes du Biuro Polonais et de Bletchley Park
- Vers un chiffrement parfait
  - Vernam, théorie de l'information

Cryptographie *ancienne*



Transposition

- Chiffrement de type anagramme : mélange les lettres du message
- Sécurité théorique
  - Message de 35 lettres : 35! chiffrés possibles
- Problèmes
  - Confusion sur la syntaxe mais … chaque lettre conserve sa valeur
  - Clé de chiffrement « complexe »
  - Ex: Scytale spartiate (5ème siècle av JC)



Substitution

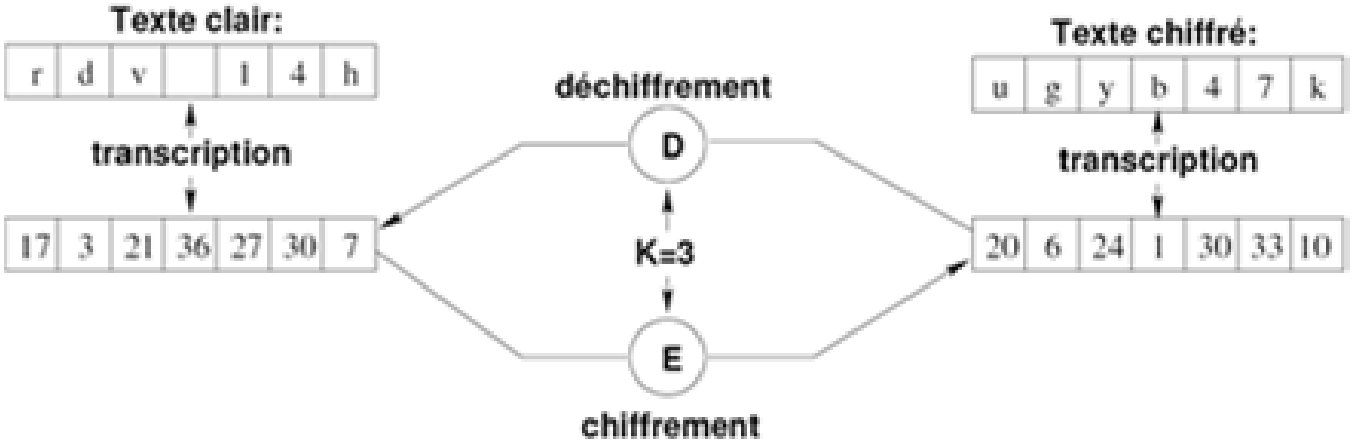
- Chiffrement en changeant d'alphabet
  - Kama sutra : mlecchita-vikalpà (art de l'écriture secrète, 4ème siècle av JC)
- Sécurité théorique
  - Alphabet de 26 lettres : 26! alphabets possibles
- Problèmes
  - Confusion sur l'alphabet mais … chaque lettre conserve sa place d'origine
  - Ex: Chiffrement de Jules César (1er siècle av JC)

Alphabet clair : abcdefghijklmnopqrstuvwxyz
Alphabet chiffré : DEFGHIJKLMNOPQRSTUVWXYZABC

Texte clair : errare humanum est, perseverare diabolicum
Texte chiffré : HUUDUH KXPDQXP HVW, SHUVHYHUDUH GLDEROLFXP

Code de substitution de César

- Chiffrement par décalage avec K=3 :
  - $E_K(M) = M+K \bmod n$
  - $D_K(M) = C-K \bmod n$



∆ Seulement n façons différentes de chiffrer un message
  - Attaque brutale facile de nos jours
  - Réemployé par les officiers sudistes, sur les forums de news (ROT-13)

Carré de Blaise de Vigenère (vers 1560)

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

Utilisation du tableau de Vigenère

- Une clef définit le décalage pour chaque lettre du message
  - Pour chaque paire, le caractère de la clef choisit une rangée du carré de Vigenère et le caractère du message choisit la colonne
  - Pour déchiffrer, c'est symétrique !

Clair	L	A	V	I	E	E	S	T	B	E	L	L	E
Clef	B	O	N	J	O	U	R	B	O	N	J	O	U
Décalage	1	14	13	9	14	20	17	1	14	13	9	14	20
Chiffré	M	O	I	R	S	Y	J	U	P	R	U	Z	Y

  - (A: décalage de 0, B:1, C:2, …, Z:25) : $S+R \equiv J\ [26]$
  - C'est symétrique : $R-J \equiv S\ [26]$

Trouver la longueur de la clef : Kasiski

- Charles Babbage 1854, Kasiski 1863
  Si une même séquence se répète alors la distance entre les deux séquences est probablement un multiple de la taille de la clef

  OOBQB PQAIU NEUSR TEKAS RUMNA RRMNR ROPYO DEEAD ERUNR
  QLJUG CZCCU NRTEU ARJPT MPAWU TNDOB GCCEM SOHKA RCMNB
  YUATM MDERD UQFWM DTFKI LROPY ARUOL FHYZS NUEQM NBFHG
  EILFE JXIEQ NAQEV QRREG PQARU NDXUC ZCCGP MZTFQ PMXIA
  UEQAF EAVCD NKQNR EYCFI RTAQZ ETQRF MDYOH PANGO LCD

- On repère les triplets ou quadruplets se répétant, puis la distance entre eux

PQA	150	DER	57	CZCC	114		
RTE	42	RUN	117	MNB	42		
ROPY	81	UNR	12	ARU	42	UEQ	54

- Pgcd(150,42,81,12,117,57,114,54) vaut 3 => longueur de la clé : 3

Vers un chiffrement parfait

- One-time-pad, Vernam (USA 1917)
  - $C = M \oplus K$ (xor bit à bit)
  - Téléphone rouge

- Vigenère avec : Longueur du mot-clef = longueur du message !
  - Confusion totale : chiffrement de « aaaa…aaa » aléatoire
  - Diffusion totale : si le mot-clef n'est *jamais* réutilisé

- Cryptanalyse possible uniquement si
  - Mot-clef trivial
  - Réutilisation du mot-clef : $(M_1 \oplus K) \oplus (M_2 \oplus K) = M_1 \oplus M_2$!!!
    - ⇨ Alors des morceaux de textes en clair donnent $M_1 \oplus M_2 \oplus N \approx M_1$

- Seul code aujourd'hui Mathématiquement prouvé sûr (cf. entropie) …

Cryptographie moderne

- Principes de Auguste Kerckhoffs (1883)
  1. La sécurité repose sur le secret de la clef et non sur le secret de l'algorithme
     - Canal +, Cartes Bleues (LUHN-10) !!!
  2. Le déchiffrement sans la clef doit être impossible (à l'échelle humaine)
  3. Trouver la clef à partir du clair et du chiffré est impossible (à l'échelle humaine)

Théorie de l'information

- Claude Shannon 1948

- Entropie
  - Quantité d'information : $I(p) = \log_2(1/p)$
  - Entropie : quantité moyenne d'information contenue dans un message
    - Nombre de questions binaires pour déterminer la valeur d'un dé
      - Dé normal : 3 questions = $[\log_2(6)] \approx 2.585$
      - Dé pipé : le 1 une fois sur deux et les autres une fois sur 10
        - » $\frac{1}{2} * 1 + \frac{1}{2} * 4 = 2.5$
        - » $\frac{1}{2} * 1 + \frac{1}{2} * (2/5 * 3 + 3/5 * 4) = 2.3$
        - » $\frac{1}{2} * \log_2(2) + 5\ (1/10\ \log_2(10)\ ) \approx 2.161$
    - $H(M) = \sum_p p \log_2(1/p)$

  - L'entropie est maximale lorsque toutes les probabilités sont égales

    Ex: Si l'apparition des lettres est exactement aléatoire, il est impossible d'appliquer l'attaque fréquentielle

Théorie des codes : théorèmes de Shannon 1948

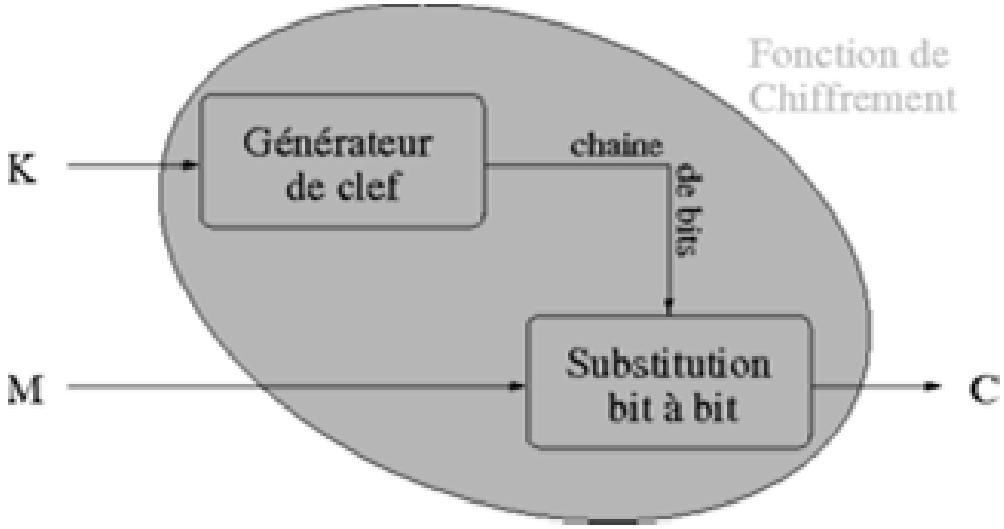
- Compression : « Pour toute source X d'entropie H(x) on peut trouver un code dont la longueur moyenne s'approche de H(X) d'aussi prêt que l'on veut »
  - Algorithme d'Huffman, extensions de sources
- Correction d'erreurs : « Pour tout canal on peut toujours trouver une famille de codes dont la probabilité d'erreur après décodage tend vers 0 »
- Cryptage : « si un chiffrement est parfait, alors il y a au moins autant de clefs que de messages »
  - Un cryptanalyste doit obtenir de l'ordre de H(M) informations pour retrouver M

Complexité et cryptographie

- Niveau de complexité d'une attaque
  - Comparer avec la recherche exhaustive
- Chiffrement idéal (moins que parfait !)
  - L'implémentation est possible : complexité polynomiale au pire
  - Mais toutes les attaques sont de complexité exponentielle au mieux
- Chiffrement sûr
  - Toutes les attaques *connues* sont de complexité exponentielle
- Chiffrement pratique
  - Attaquer coûte plus cher (machines, …) que la valeur du secret
  - Attaquer prend plus de temps que la validité du secret

Chiffrement par flot (stream cipher)

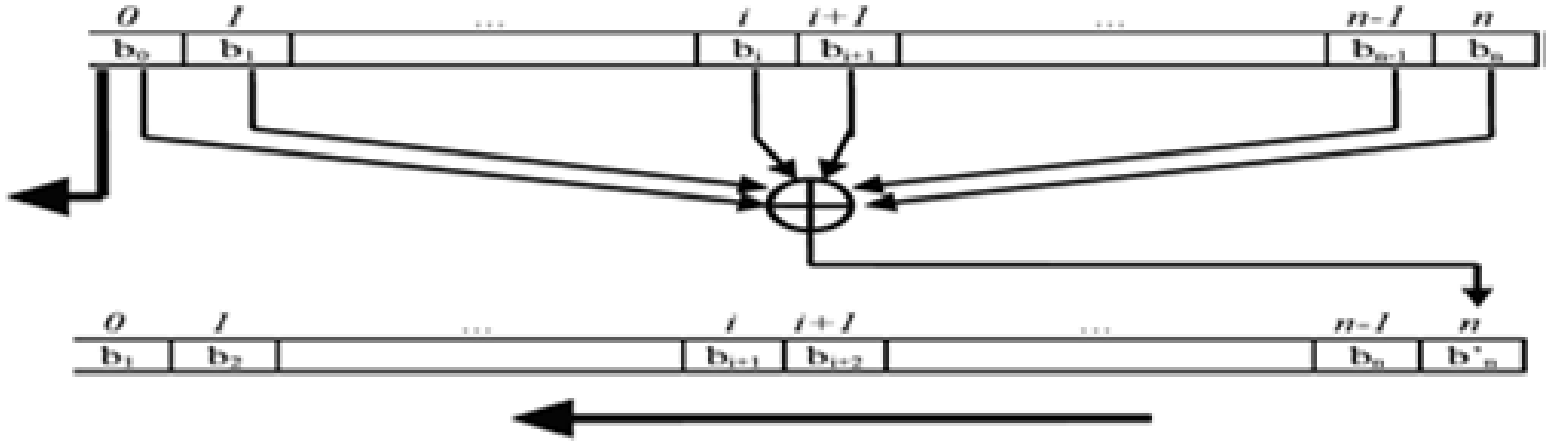
- Chiffrement à la one-time-pad
  - Taille du message M : n
  - Avec une petite clef K, générer K' de taille n



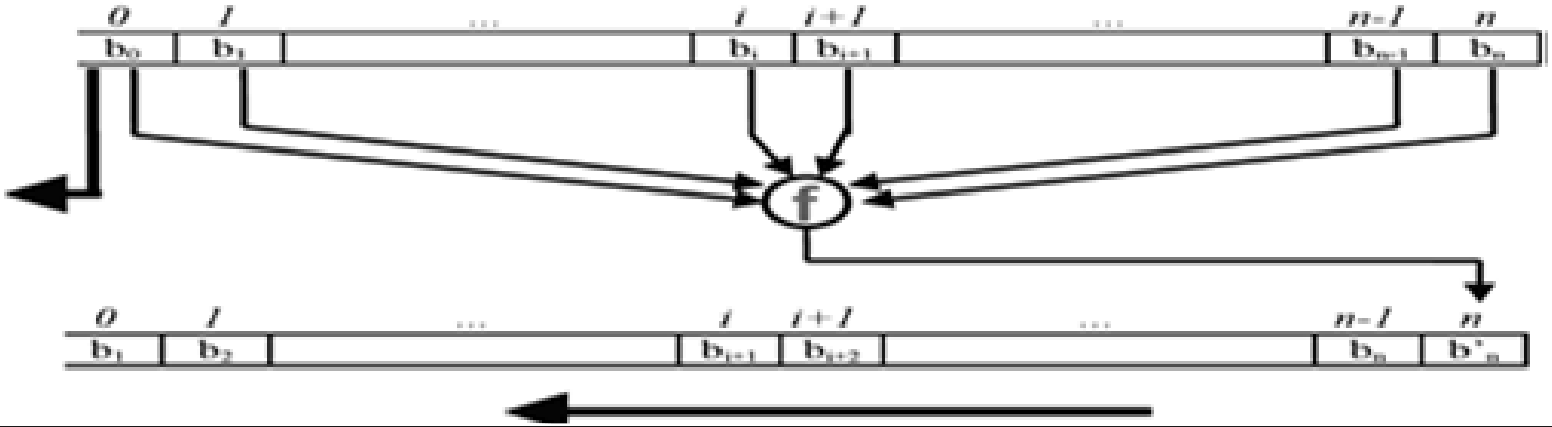
- Sécurité
  - Substitution rapide (xor)
  - Génération de clef
    - Fonction pseudo-aléatoire : devrait être impossible à prédire à l'échelle humaine
  - Principe de Kerckhoffs : la sécurité repose sur la clef

Générateurs Hardware

- Registres linéaires à décalage (LFSR)

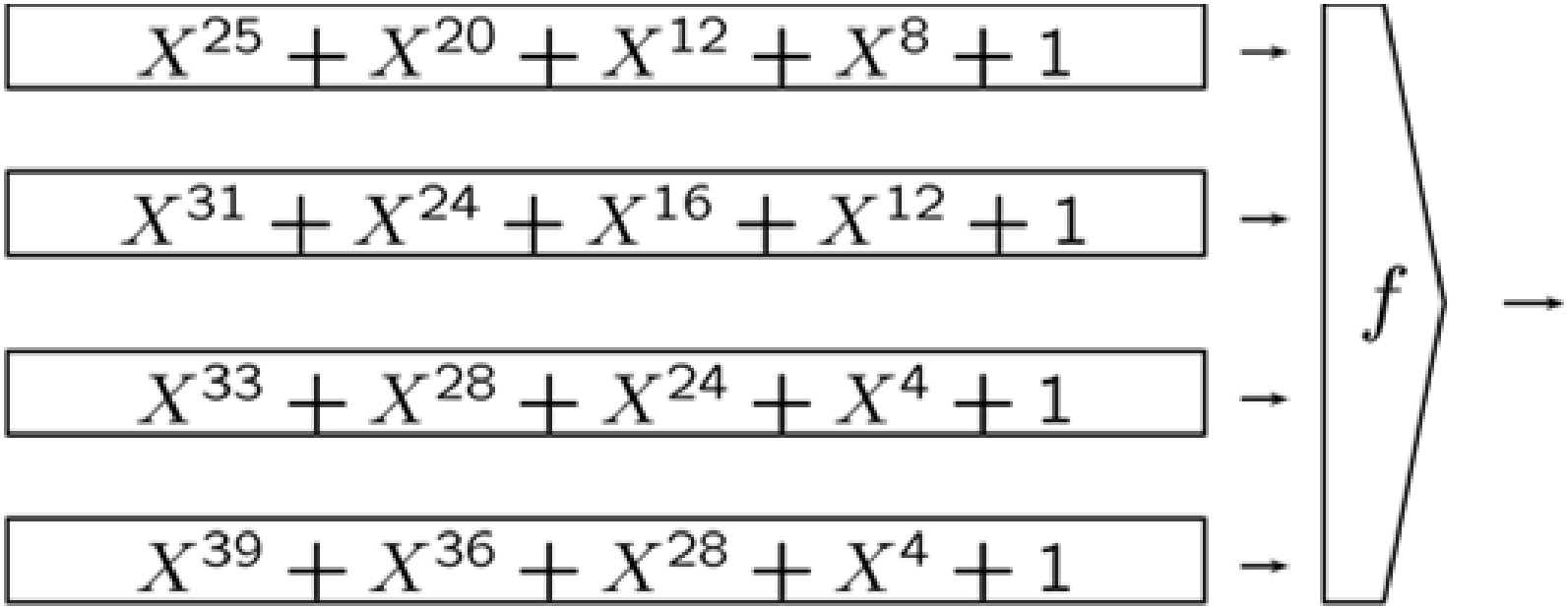


- Registres non-linéaires à décalage (NLFSR)



Bluetooth 1/2

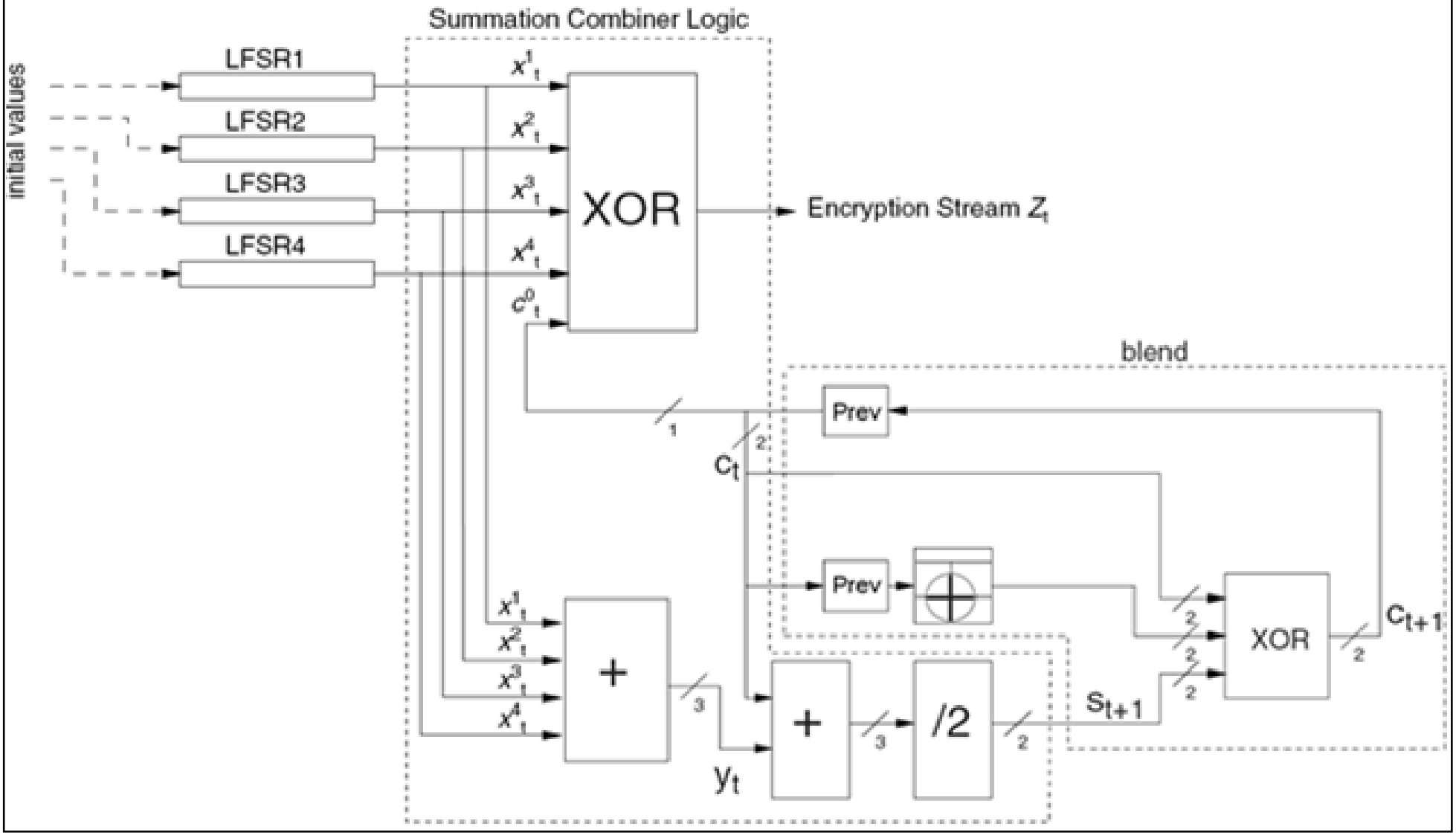
- Sécurisation des communications radio entre unité centrale et périphériques : 4 LSFR puis une fonction f



⇨ Registre à 25+31+33+39=128 bits

⇨ Période : ppcm(2^{25}-1,2^{31}-1,2^{33}-1,2^{39}-1) $\approx$ 2^{125}

Bluetooth

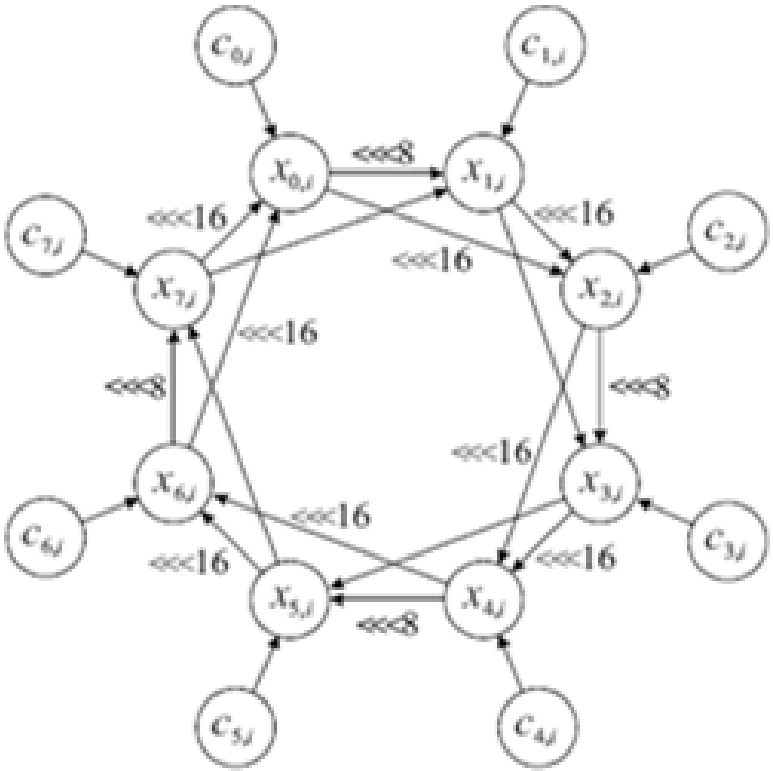

Summation Combiner Logic
initial values
LFSR1
LFSR2
LFSR3
LFSR4
XOR
Encryption Stream Z_t
blend
Prev
Prev
XOR
+
+
/2
c_t
y_t
S_{t+1}
c_{t+1}

Registres à décalage

- Implémentation facile et rapide en hardware
  - Période : n registres, $T \leq 2^n-1$
  - Sûr : RC4 avec clef de 256 bits, 2^{1700} états
  - Rapide : 10 fois plus rapide que DES par exemple
- LFSR cassés par l'algorithme de Berlekamp-Massey 1969
  - Polynômes générateurs
  - Prévoient les prochains bits à partir des premiers
- NLFSR, Problème : fonction f généralement secrète en contradiction avec le principe de Kerckhoffs
  - RC4, variante des NLFSR, f publiée en 1994

Chiffrement cryptographique par flot : Rabbit

- 128 bits de clef, 64 bits de valeur initiale IV génèrent
  - 512 bits de compteurs internes (8 $\times$ 32 bits) : $c_{i,j}$
  - 512 bits d'états internes (8 $\times$ 32 bits) : $x_{i,j}$



$$g_{j,i} = (x_{j,i} + c_{j,i})^2 \oplus ((x_{j,i} + c_{j,i})^2 >> 32) \mod 2^{32}$$
$$x_{0,i+1} = g_{0,i} + (g_{7,i} <<< 16) + (g_{6,i} <<< 16) \mod 2^{32}$$
$$x_{1,i+1} = g_{1,i} + (g_{0,i} <<< 8) + g_{7,i} \mod 2^{32}$$
$$x_{2,i+1} = g_{2,i} + (g_{1,i} <<< 16) + (g_{0,i} <<< 16) \mod 2^{32}$$
$$x_{3,i+1} = g_{3,i} + (g_{2,i} <<< 8) + g_{1,i} \mod 2^{32}$$
$$x_{4,i+1} = g_{4,i} + (g_{3,i} <<< 16) + (g_{2,i} <<< 16) \mod 2^{32}$$
$$x_{5,i+1} = g_{5,i} + (g_{4,i} <<< 8) + g_{3,i} \mod 2^{32}$$
$$x_{6,i+1} = g_{6,i} + (g_{5,i} <<< 16) + (g_{4,i} <<< 16) \mod 2^{32}$$
$$x_{7,i+1} = g_{7,i} + (g_{6,i} <<< 8) + g_{5,i} \mod 2^{32}$$

Propriétés des « stream ciphers »

- Très proches de Vernam (pseudo-random)
- Les plus rapides ($\times$10 symétriques/blocs , $\times$ 100 pub-key)

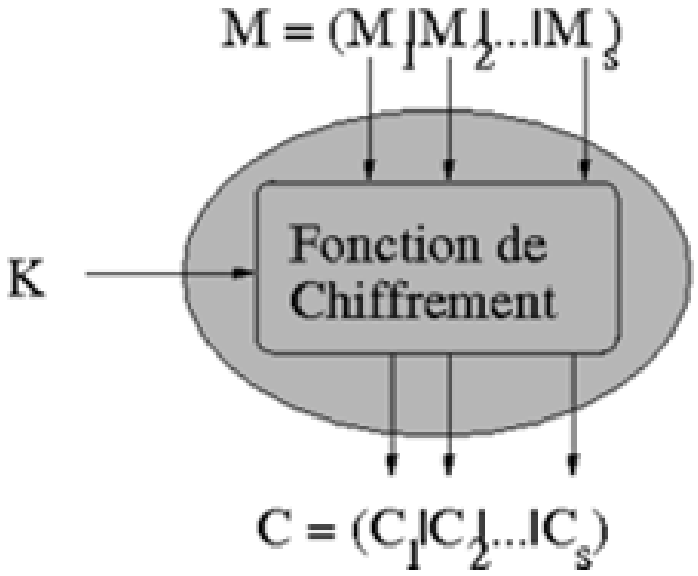
- A5-GSM, RC4-WEP cassés …
- Pas de standard de remplacement
  - eStream 2008 → 7 recommandations, dont Rabbit

- Synchronisation ou pré-génération de clefs
  ✚ État

⇨ Besoin de parallélisation → chiffrement par blocs
  - Tout chiffrement par bloc + mode type CTR peut-être utilisé comme chiffrement par flot …

Chiffrement par bloc

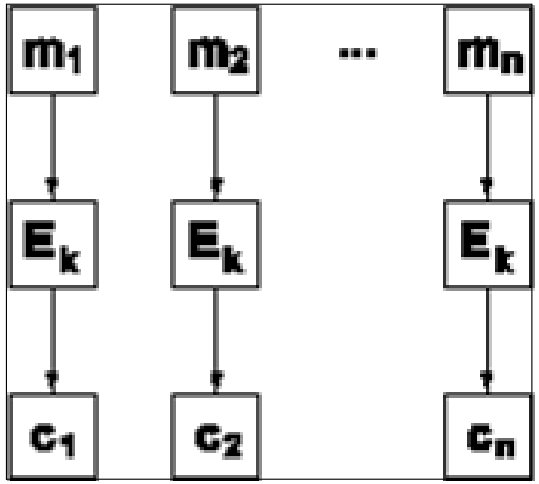
- $M = (M_1|M_2|...|M_s)$: s blocs de r=n/s bits chacun



- Sécurité
  - Sur chaque bloc : $C_i = E_K(M_i)$, dépend de la fonction E
  - Pour le message : dépend également du mode de chiffrement

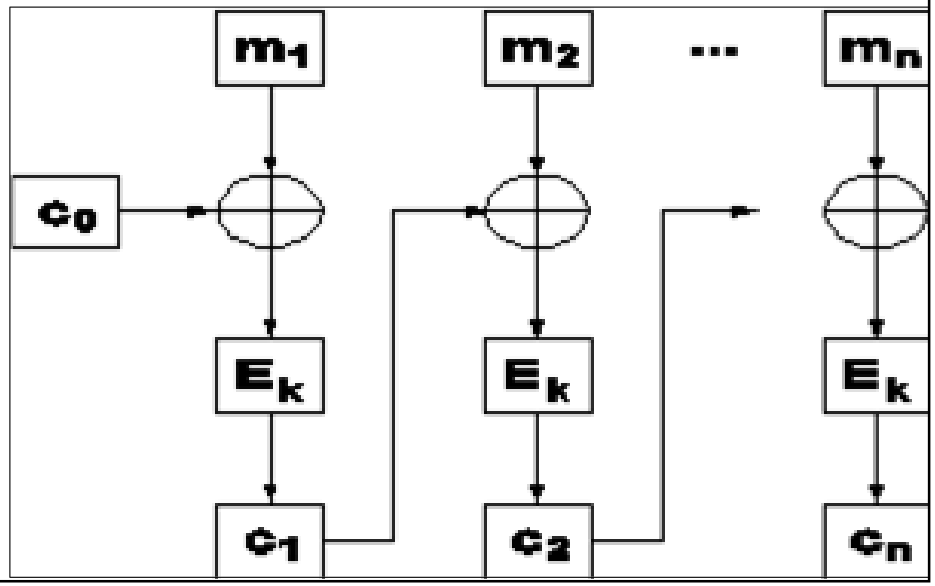
Cinq Modes de chiffrement (1/3)

1. ECB : Electronic Code Book
   - ⇨ $C_i = E_k(M_i)$
   - ⇦ $M_i = D_k(C_i)$
   - – Un bloc est toujours crypté identiquement
   - – Aucune sécurité, pas d'utilisation



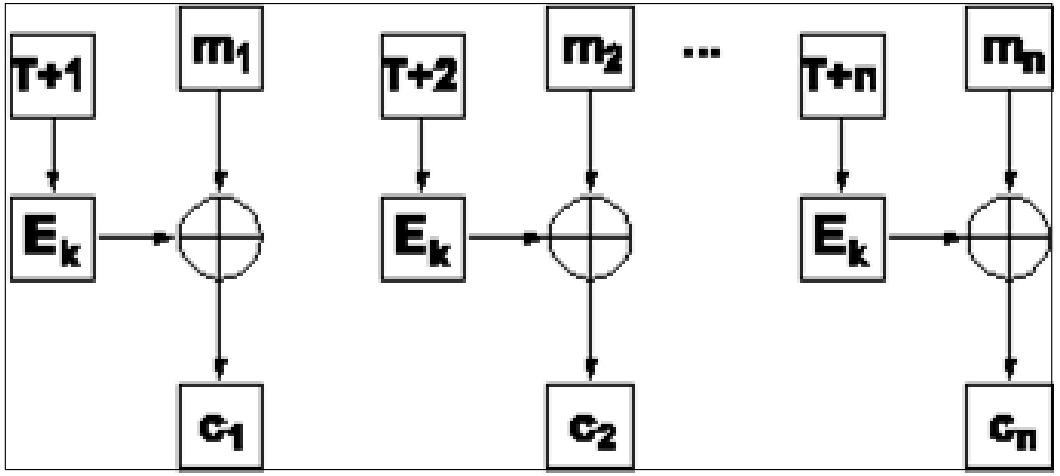
2. CBC : Cipher Bloc Chaining
   - ⇨ $C_i=E_k(M_i \oplus C_{i-1})$
   - ⇦ $M_i=C_{i-1} \oplus D_k(C_i)$

   - – Le plus utilisé



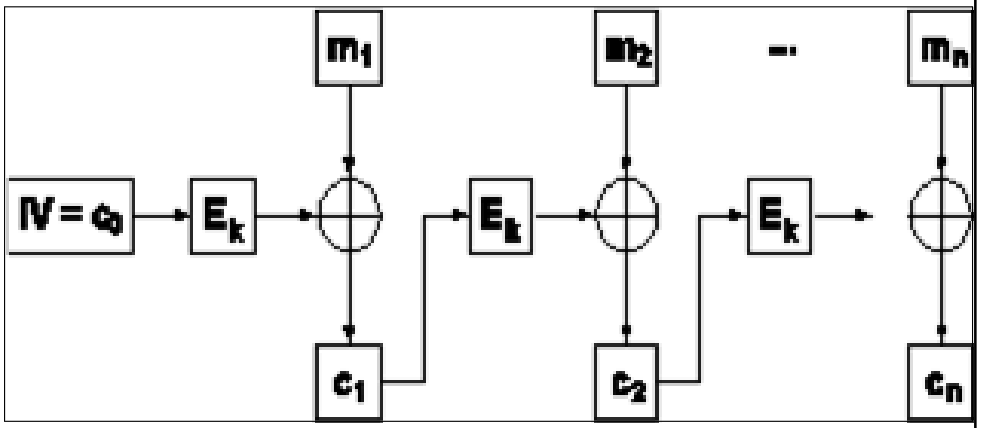
Cinq Modes de chiffrement (2/3)

3. CTR : Counter Encryption Mode
   - ⇨ $C_i = M_i \oplus E_k(T+i)$
   - ⇦ $M_i = C_i \oplus E_k(T+i)$
   - – Fait intervenle chiffrement d'un compteur de valeur initiale T
   - – Totalement symétrique, facilement parallélisable
   - – Un même bloc n'est a priori jamais codé de la même façon



(3/3)

4. CFB : Cipher Feedback
   - ⇨ $C_i=M_i \oplus E_k(C_{i-1})$
   - ⇦ $M_i=C_i \oplus E_k(C_{i-1})$
   - – Pas besoin de D_k
   - – Moins sûr, parfois plus rapide
   - – Réseaux



5. OFB : Output Feedback
   - ⇨ $Z_i = E_k(Z_{i-1})$; $C_i = M_i \oplus Z_i$
   - ⇦ $Z_i = E_k(Z_{i-1})$; $M_i = C_i \oplus Z_i$
   - – Totalement symétrique
   - – Moins de cablage
   - – Satellites