Cryptographie à clef secrète

Boly SECK

<boly.seck@univ-st-etienne.fr>

d'après un cours de mon directeur de thèse

Pierre-Louis CAYREL

<pierre.louis.cayrel@univ-st-etienne.fr>

Ecole superieure Polytechnique (ESP), Dakar/Université Jean Monnet, Saint-Etienne

2021/2022

Séance 1

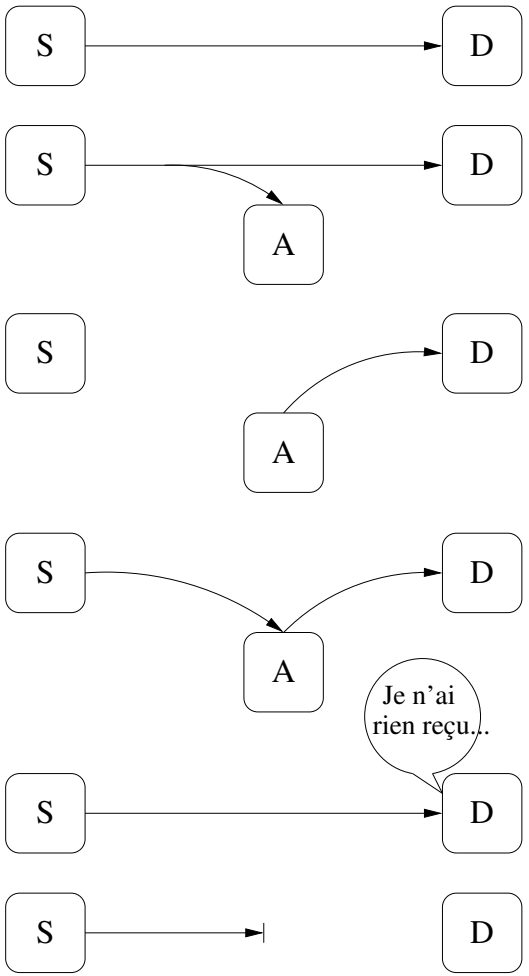
Cryptologie : présentation, historique

Besoins des systèmes d'information

- Menaces, adversaires nombreux et variés
- Comment protéger notre monde interconnecté ?
- Définir des besoins de sécurité : CAIN(D)
  - **Confidentialité** : garantir le secret des données
  - **Authentification** : garantir qu'une entité est celle qu'elle prétend être
  - **Integrité** : garantir la non modification des données
  - **Non-répudiation** : garantir qu'une transaction ne peut pas être niée
  - **Disponibilité** : garantir le bon fonctionnement d'un système d'information

Attaques portées sur un canal de communication

- **Flot normal** :
- **Interception** : confidentialité
- **Fabrication** : authentification
- **Modification** : integrité
- **Négation** : non-répudiation
- **Interruption** : disponibilité



Petit historique

- Apanage des militaires/détenteurs du pouvoir
  - V^e siècle avant notre ère (Grecs) : Scytale

  - IV^e siècle avant notre ère (Indiens) : Kautylia (ministre de Ghandragupta) « il faut se doter d'un service de renseignement. »
  - I^{er} siècle avant notre ère (Romains) : Jules César, communication chiffrée dans l'Empire (par substitution)
  - Moyen âge : Obscurographie
  - XVI^e siècle : Vigenère (résiste jusqu'au milieu du XIX^e, Kasiski/Babbage)
  - Fin XIX^e : Principes de Kerckhoffs
  - Première guerre mondiale : télégramme de Zimmermann (1917) et radiogramme de la victoire (Painvain)
  - 1917 : Vers un chiffrement parfait (Vernam), théorie de l'information
  - Seconde guerre mondiale : Enigma (5 rotors)

Petit historique

- Plus récemment :
  - Standard de chiffrement : DES (Data Encryption System, 1977), AES (Advanced Encryption System, 2000)
  - Principe de cryptographie à clé publique : Diffie et Hellmann (1976)
  - RSA (Rivest, Shamir, Adleman, 1978)

Notion de Cryptologie

Définition

La *cryptologie* est la science du secret, elle est composée de deux composantes complémentaires :

1. **La cryptographie** : étude et conception des procédés de chiffrement des informations
2. **La cryptanalyse** : analyse des textes chiffrés pour retrouver l'information dissimulée

- Attention : Cryptographie $\neq$ Stéganographie
- Stéganographie : dissimule l'information (pas forcément chiffré)
- Exemple célèbre : Lettre de George Sand à Alfred De Musset

Stéganographie : un exemple célèbre

Lettre de George Sand à Alfred de Musset

Cher ami,
Je suis toute émue de vous dire que j'ai
bien compris l'autre jour que vous aviez
toujours une envie folle de me faire
danser. Je garde le souvenir de votre
baiser et je voudrais bien que ce soit
une preuve que je puisse être aimée
par vous. Je suis prête à montrer mon
affection toute désintéressée et sans cal-
cul, et si vous voulez me voir ainsi
vous dévoiler, sans artifice, mon âme
toute nue, daignez me faire visite,
[...]

Stéganographie, quelques mots

- Information non chiffrée :

  Connaissance de l'existence de l'information

  =

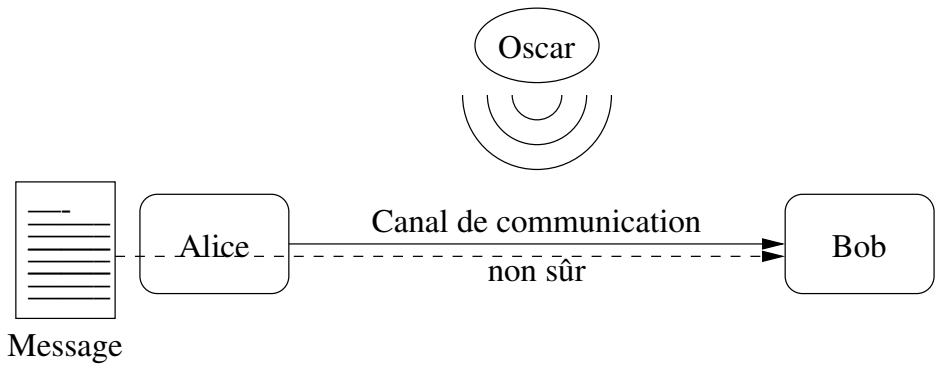
  Connaissance de l'information
- Exemples : Encre sympathique, tablette recouverte de cire, crâne du messager, messages percés de micro trous, traitement de texte des ministres sous Thatcher, etc. . .
- Faible niveau de sécurité
- En pratique fonctionne : 11/09/2001 ?
- Utilisé pour le tatouage numérique (Watermarking) : mp3, mpeg, jpeg, etc. . .

Retour à la crypto : terminologie

- Présentation de 2 (3) personnages célèbres ! : Alice, Bob, (Oscar ou Eve)
  - Alice et Bob veulent communiquer
  - Oscar ou Eve (opposant ou espion) veulent savoir ce que s'échangent Alice et Bob

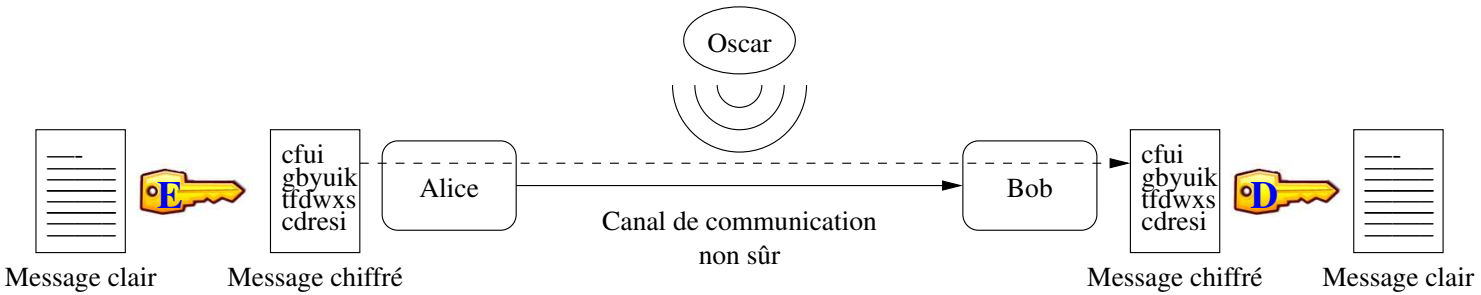
Objectif principal de la cryptographie

Permettre à Alice et Bob de communiquer sur un canal peu sûr sans que Oscar comprenne ce qui est echangé



Terminologie

- **Texte clair** : information M qu'Alice souhaite transmettre à Bob
  - Texte, image, vidéo, etc. . .
- **Chiffrement** : processus permettant de rendre le message M incompréhensible
  - Basé sur une **fonction de chiffrement :** E
  - Génération d'un message **chiffré :** $C = E(M)$
- **Déchiffrement** : processus permettant la reconstruction du message clair à partir du message chiffré
  - Basé sur une **fonction de déchiffrement :** D
  - Propriété : $D(C) = D(E(M)) = M$



Terminologie

- E et D paramétrées par des **clés** : K_E et K_D

$$\begin{cases} E_{K_E}(M) &= C \\ D_{K_D}(C) &= M \end{cases}$$

- $K_E, K_D \in \{\text{Espace de clés}\}$
- Deux grands types de systèmes de chiffrement :
  1. $K_D = K_E$: chiffrement **symétrique**
  2. $K_D \neq K_E$: chiffrement **asymétrique**

Cryptologie : deux grands principes de chiffrement

1. Chiffrement par transposition : change l'ordre des mots et/ou des lettres en fonction d'une convention secrète

Exemple :

Message clair :	M	E	S	S	A	G	E
Message chiffré :	E	S	M	S	G	E	A

2. Chiffrement par substitution : remplace les lettres du texte clair, sans changer l'ordre, à l'aide d'un code (permutation des lettres de l'alphabet par exemple)

Exemple : $A \rightarrow B$

Message clair :	M	E	S	S	A	G	E
Message chiffré :	N	F	T	T	B	H	F

Chiffrons !

- Un texte clair : “hello”
- Correspondance lettres $\leftrightarrow$ chiffres : codage de l’information

a	b	$\ldots$	x	y	z
00	01	$\ldots$	23	24	25

(ou table ASCII)

“hello” $\leftrightarrow$ 07 04 11 11 14

- Fonction de chiffrement : décalage des lettres par exemple

$$E : y = x + 13 \mod 26 \qquad \text{Arithmétique modulaire}$$

$07\,04\,11\,11\,14 \overset{E}{\longmapsto} 20\,17\,24\,24\,01 \leftrightarrow$ “uryyb”

- Fonction de déchiffrement :

$$D : y = x - 13 \mod 26 \qquad (D(E(x)) = x)$$

$20\,17\,24\,24\,01 \overset{D}{\longmapsto} 07\,04\,11\,11\,14 \leftrightarrow$ “hello”

Remarques

- Codage de l'information $\neq$ cryptage
- Chiffrement utilisé : définition de la structure mathématique dans laquelle on fait les opérations (ici : $\mathbb{Z}/26\mathbb{Z}$)
- Mode de chiffrement : lettre par lettre ? bloc de x lettres ? ...

Cryptanalyse - Principe de Kerckhoff

- Cryptanalyse : Etude de la sécurité des procédés de chiffrement cryptographique

Principe de Kerckhoff

La sécurité d'un système cryptographique ne doit pas reposer sur l'ignorance de la méthode de chiffrement employée.

–Auguste Kerckhoff, 1883

Corollaire

1. Oscar connaît le système cryptographique
2. La sécurité du système ne doit reposer que sur la clé secrète

Cryptanalyse - Types d'attaques

- Oscar connaît le cryptosystème utilisé.
- 4 types d'attaques :
  1. Texte chiffré connu : Oscar ne connaît que C
  2. Texte clair connu : Oscar connaît M et C correspondant
  3. Texte clair choisi : Oscar choisit M et chiffre M pour obtenir C
     - Nécessite l'accès à une machine chiffrante.
  4. Texte chiffré choisi : Oscar choisit un texte chiffré C et peut obtenir son texte clair M
     - Nécessite l'accès temporaire à une machine déchiffrante.
- Classés par ordre décroissant de difficulté.
- But d'Oscar : retrouver la clé.
- Confidentialité garantie si Oscar ne peut pas :
  - Trouver M à partir de $E(M)$
  - Trouver la clé de déchiffrement à partir d'une famille de $\{M_i, E(M_i)\}_i$

Les algorithmes d'attaque

1. Par force brute : tester toutes les clés possibles
   - Clé de 64 bits : $2^{64} \approx 1.8 \times 10^{19}$ possibilités
   - 1 milliards de tests / sec : 1 an sur 584 machines
2. Par séquence connue :
   - Deviner la clé si une partie du message est connue (En tête de fichiers, de courriels, . . .)
3. Par séquence forcée :
   - Faire chiffrer par la victime un bloc dont l'attaquant connait le contenu, puis appliquer l'attaque précédente.
4. Attaque par analyse différentielle :
   - Utiliser les faibles différences entre plusieurs messages pour deviner la clé.

Avant de poursuivre : grands nombres

Proba de mourir foudroyé (par jour)	1 chance sur 9 milliards (2^{33})
Proba de gagner le gros lot (US)	1 chance sur 4 millions (2^{22})
Proba de gagner le gros lot et de mourir le même jour	1 chance sur 2^{61}
Proba de se noyer (US/an)	1 chance sur 59000 (2^{16})
Proba d'être tué dans un accident de voiture (US/vie)	1 chance sur 88 2^{7}
Temps d'ici la prochaine glaciation	14000 ans (2^{14})
Temps d'ici à ce que le soleil explose en nova	10^{9} années (2^{30})
Âge de la terre	10^{9} années (2^{30})
Âge de l'univers	10^{10} années (2^{34})
Nb atomes constituant la terre	10^{51} (2^{170})
Nb atomes constituant le soleil	10^{57} (2^{190})
Nb atomes constituant la galaxie	10^{67} (2^{223})
Nb atomes constituant l'univers	10^{77} 2^{265}
Volume de l'univers	10^{84} cm^3 (2^{280})
Durée de vie de l'univers (si fermé)	10^{11} années (2^{37}) 10^{18} secondes (2^{61})
Temps d'ici à ce que les étoiles peu massives se refroidissent	10^{14} années (2^{47})
Temps d'ici à ce que les planètes quittent leur étoile	10^{15} années (2^{50})
Temps d'ici à ce que les étoiles quittent leur galaxie	10^{19} années (2^{64})
Temps d'ici à ce que toute la matière soit liquide à $0K$	10^{65} années (2^{216})
Temps d'ici à ce que toute la matière se transforme en fer	$10^{10^{26}}$ années

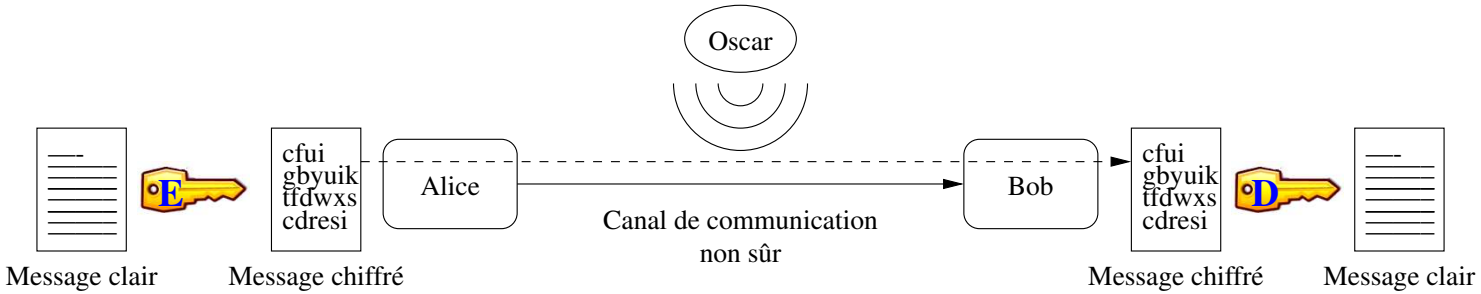
Source : *–Applied Cryptography, Bruce Schneier*

Avant de poursuivre : limitations thermodynamiques

- *2nd* principe thermodynamique : « *Il faut une certaine quantité d'énergie pour représenter de l'information* »
  - Enregistrer un simple bit : $E \geq kT$ où $k = 1,38.10^{-23}$ constante de Boltzman et T température fondamentale du système.
  - Température ambiante de l'univers : $3,2K$
- Ordinateur idéal fonctionnant à $3,2K$ consomme $4,4.10^{-23}J$ à chaque effacement ou modification d'un bit.
- Soleil dégage : $1,2.10^{34}J$ en un an
- Assez pour effectuer $2,8.10^{63}$ changements de bits (soit assez pour qu'un compteur de 210 bits prenne toutes ses valeurs)
- 64 ans à capter **toute** l'énergie du soleil : compteur de 216 bits passe par tous ses états
- Attaque exhaustive contre des clés de 256 bits impossible jusqu'à ce que les ordinateurs soient fabriqués avec autre chose que de la matière et fonctionnent ailleurs que dans l'espace.

Généralités - Premiers principes

Principe : Chiffrement symétrique



- $K_D = K_E = K$ (clé privé convenue secrètement entre Alice et Bob)
  - Efficace point de vue temps de calcul.
  - Inconvénient : partage du secret
- Premier type de chiffrement utilisé
- Fournit le seul chiffrement théoriquement indéchiffrable
  - Chiffrement de Vernam (one time pad) : utilisé entre les US et la Russie pendant la guerre froide (Téléphone Rouge)
  - Démonstration du mathématicien Claude Shannon (1949)

Outils utilisés pour le chiffrement symétrique

- 2 types de méthodes à la base du chiffrement symétrique
  1. Transposition
  2. Substitution
- Calculs :
  1. Calculs modulaires dans $\mathbb{Z}/n\mathbb{Z}$:

     $$a \equiv b \mod n \Leftrightarrow n \text{ divise } a - b$$

     En pratique, b est le reste de la division de a par n
  2. Opération XOR (ou exclusif : $\oplus$)

$\oplus$	0	1
0	0	1
1	1	0

     - Opération involutive (bijection égale à sa bijection réciproque)
     - Addition bit à bit modulo 2
     - Rapidité de calcul

Les premiers procédés

- Chiffrement par permutation :
  - Alphabet : de n lettres
  - Une fonction de chiffrement est une permutation des n lettres de l'alphabet.
  - $n!$ possibilités
- Chiffrement par décalage : utilise n permutations des $n!$ possibles

$$\begin{cases} E_K(M) & = & M + K \mod n \\ D_K(C) & = & C - K \mod n \end{cases}$$

  - Faiblesses :
    - Seulement n clés possibles
    - Recherche exhaustive facile
  - Cas particulier : chiffrement de César $K = 3$

Généralisation du chiffrement par décalage

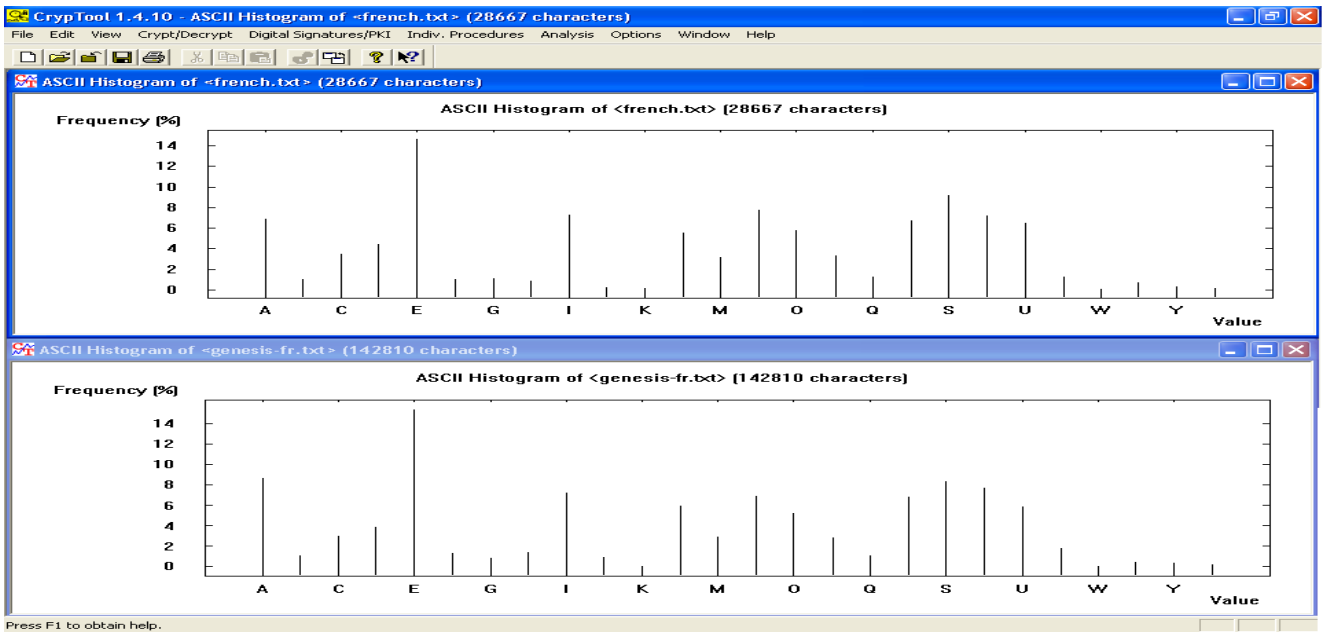
- chiffrement affine
  - $E_K(M) = a \times M + b \mod n$
  - $K = (a, b)$
  - Nombre de clés possibles : ?

$$n \times \phi(n) \quad \begin{cases} n & = & \prod_{i=1}^{m} p_i^{e_i} \\ \phi(n) & = & \prod_{i=1}^{m} (p_i^{e_i} - p_i^{e_i - 1}) \end{cases}$$

  - Pour $n = 26$, $n \times \phi(n) = 312$ clés.

Cryptanalyse des chiffrement par substitution mono-alphabétique

- Nombre de possibilités : $n!$ soit $\geq 4.10^{26}$ trop grand pour une recherche exhaustive même pour un ordinateur.
- Faiblesse mono-alphabétique : pour une permutation donnée, une lettre est toujours remplacée par la même lettre.
- Analyse fréquentielle possible ! (consultation de tables de fréquence par langue)



Chiffrement poly-alphabétique

- Principe : une lettre du texte clair n'admet pas une unique lettre chiffrée.
- Exemple célèbre : Chiffre de Vigenère (Blaise Vigenère, *XVI*^{e} siècle)

Chiffrement de Vigenère

- Représentation des lettres de l'alphabet : $A \leftrightarrow 0, \ldots, Z \leftrightarrow 25$
- Choix d'un mot clé de longueur m : *CLESEC* (ici $m = 6$)
- Texte clair : *CECRYPTOSYSTEMEESTCASSE*
- Découpage du texte en blocs de m lettres et ajout de la clé (additions dans $\mathbb{Z}/26\mathbb{Z}$)

	C	E	C	R	Y	P		T	O	S	Y	S	T		E	M	E	E	S	T		C	A	S	S	E
+	C	L	E	S	E	C		C	L	E	S	E	C		C	L	E	S	E	C		C	L	E	S	E
=	E	P	G	J	C	R		V	Z	W	Q	W	V		G	X	I	W	W	V		E	L	W	K	I

Analyse du chiffre de Vigenère

- Nombre de mots clés possibles de longueur m : 26^m
- Recherche exhaustive coûteuse, même pour de petites valeurs de m
- Cryptanalyse plus difficile que les substitutions mono-alphabétique
- Système non cassé pendant $\approx$ 2 siècles
- Cassé fin XIX^e par Babbage (1854) et Kasiski (1863)
- Idée principale : se ramener à une cryptanalyse mono-alphabétique en déterminant dans un premier temps la taille de la clé.
- Comment trouver la taille de la clé ? (TD)

Autre chiffrement poly-alphabétique

- Chiffrement de Hill (1929)
- Idée : transforme m caractères d'un bloc de texte clair en m caractères d'un bloc de texte chiffré par des combinaisons linéaires
- Lien avec les matrices et l'algèbre linéaire
- Clé de chiffrement = K matrice $m \times m$ inversible
- Cas particulier : chiffrement par permutation
  - Rappel du chiffrement par permutation : conservation des lettres, changement de l'ordre (π permutation de 26 lettres)
  - Si $K_\pi = (k_{i,j})$ avec $k_{i,j} = \begin{cases} 1 & \text{si } i = \pi(j) \\ 0 & \text{sinon} \end{cases}$
- Cryptanalyse en deux temps :
  1. Calcul de m
  2. Attaque à clair choisi

Sécurité inconditionnelle

- Cryptosystèmes précédents utilisent répétition de la clé

Définition (Sécurité inconditionnelle)

Un système est dit inconditionnellement sûr lorsque la connaissance du message chiffré n'apporte aucune information sur le message clair.
Conséquences :

- Seule attaque possible : attaque exhaustive clé secrète
- Clé secrète au moins aussi longue que le texte clair

Un tel système existe-il ?

Chiffre de Vernam (One Time Pad)

- Relation fondamentale :

$$\forall M, K \text{ tel que } \mid M \mid = \mid K \mid, (M \oplus K) \oplus K = M$$

Fonctions de chiffrement/déchiffrement : $\begin{cases} E_K(M) = M \oplus K \\ D_K(C) = C \oplus K \end{cases}$

- Vigenère avec |mot clé|=|Texte clair| !
- Ex :

$$\begin{array}{rcl} M & = & 10010111 \\ K & = & 01011010 \\ \hline C = M \oplus K & = & 11001101 \end{array}$$

Système inconditionnellement sûr, Claude Shannon (1949)

Si K est **totalement aléatoire** et n'est utilisée **qu'une seule fois**, alors il est impossible d'obtenir une information sur M à partir de C

Système pratiquement sûr

- Vernam seul système prouvé inconditionnellement sûr
  - Sécurité basée sur la génération aléatoire de la clé
  - Nécessite un « bon »générateur aléatoire
  - Problème du stockage de la clé
  - Tous les autres systèmes sont cassables !
- Système pratiquement sûr

Définition : Système pratiquement sûr

Un système de chiffrement est dit pratiquement sûr s'il n'est pas possible de retrouver à partir du texte chiffré :

- La clé
  ou/et
- Le texte clair

en un temps humainement raisonnable

- Permet d'utiliser des clés plus petites : 56 bits, 128 bits
- Pourquoi ne pas toutes les tester ?
- Puissance de calcul limitée
- Grands nombres, limitations thermodynamiques !

Cryptographie moderne

Cryptographie moderne (1)

Principe de Kerckhoff, reformulation

1. La sécurité repose sur la clé et non sur le secret de la méthode employée
2. Le déchiffrement de la clé doit être pratiquement impossible
3. Trouver la clé à partir du clair et du chiffré doit être pratiquement impossible

Cryptographie moderne (2)

Deux grands principes

Un algorithme de chiffrement moderne doit vérifier deux principes fondamentaux :

1. Principe de **Confusion :** rendre la relation entre le texte chiffré et la clé secrète la plus complexe possible
2. Principe de **Diffusion :** toute modification (comprendre la plus faible possible) faite sur le texte clair doit se répercuter sur tout le texte chiffré
   Critère d'avalanche strict : L'inversion d'un bit en entrée doit changer tous les bits en sortie avec une probabilité de $1/2$.

Cryptographie moderne (3)

Théorie de l'information (Shannon, 1948)

- Source d'information $(\mathcal{S}, \mathcal{P})$ sans mémoire (loi de proba ne varie pas au cours du temps)
  - $\mathcal{S} = (s_1, \ldots, s_n)$, $\mathcal{P} = (p_1, \ldots, p_n)$
  - p_i probabilité de s_i
- Entropie :
  - Quantité d'information de s_i : $I(s_i) = \log_2\left(\frac{1}{p_i}\right)$
  - Quantité d'information d'une source $(\mathcal{S}, \mathcal{P})$:
    - $H(\mathcal{S}) = \sum_{i=1}^{n} p_i \log_2\left(\frac{1}{p_i}\right)$
    - Nombre moyen de questions à poser pour déterminer la valeur obtenue
    - Ex. Dé non pipé : $H(\mathcal{S}) = \sum_{i=1}^{6} \frac{1}{6} \log_2(6) \approx 2,58$
      Dé pipé ($p_1 = \frac{1}{2}$, $p_{i\neq 1} = \frac{1}{10}$) : $H(\mathcal{S}) = 2,161$
    - Entropie maximale lorsque équiprobabilité
      Ex : Entropie d'un texte maximale$\Rightarrow$ impossible d'appliquer l'attaque fréquentielle.

Cryptographie moderne (4)

Théorie de la complexité

- Méthodologie pour analyser la complexité de calcul des algorithmes
  - Complexité en temps (ou nombre d'opérations élémentaires)
  - Complexité en mémoire (espace de stockage)
- S'exprime comme fonction de la taille du paramètre d'entrée
- Complexité d'un problème = complexité de l'algorithme permettant de résoudre l'instance la plus difficile
- Classes de complexité (très simplifié) :
  - Problèmes solubles (en temps polynômial) : Classe *P*
  - Problèmes difficiles (solubles en temps exponentiel) : Classe *NP*
  - Problèmes indécidables : Il n'existe pas d'algo permettant de répondre par « oui »ou par « non »à un problème posé

Complexité et cryptographie

- Niveau de complexité d'une attaque
  - A comparer avec le coût d'une attaque exhaustive
- Chiffrement idéal ($\leq$ parfait !) - pratiquement sûr
  - Toutes les attaques sont au mieux exponentielles
- Chiffrement sûr :
  - Toutes les attaques **connues** sont de complexité exponentielle.
- Chiffrement pratique :
  - Attaquer coûte plus cher que la valeur du secret
  - Attaquer prend plus de temps que la validité du secret