

ACCQ 207 - Cryptographie

1 Introduction

1.1 Historique

Les premiers usages étaient pour les militaires. Les exemples les plus connus sont :

- chiffrement par décalage, ou de Jules César,
- chiffrement par substitution, on applique une fonction de permutation sur les lettres,
- chiffrement par permutation : on divise le texte en groupes de lettres de même taille et à chaque groupe on applique une permutation entre les lettres.

La plupart des systèmes actuels peuvent se voir comme combinaisons de ces transformations élémentaires.

Principe de Kerckhoff : quasiment toutes les caractéristiques du système doivent être publiques, mais le chiffrement dépend juste d'un paramètre secret, la clé.

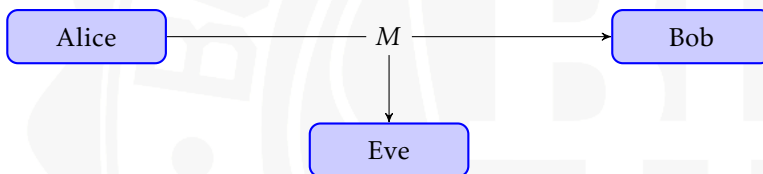
Intérêt :

- déploiement à grande échelle,
- ne compromet pas toute la sécurité du système si une seule clé s'ébruite,
- permet d'analyser la sécurité par des tiers.

Peut-on prouver qu'un système est sûr, mesurer sa sécurité ? Deux approches possibles :

- sécurité informationnelle \rightarrow théorie de l'information (Shannon).
1948 : *A mathematical theory of communications*
1949 : *Communication theory of secrecy systems*
- sécurité computationnelle \rightarrow théorie de la complexité.

1.2 Sécurité informationnelle



Modèle : A veut envoyer un message clair $M \in \mathcal{M}$. M est vu comme une variable aléatoire de loi donnée.

Def. Fonction de chiffrement : $E: \mathcal{M} \times \mathcal{K} \rightarrow \mathcal{C}$ où \mathcal{K} est l'espace des clés et \mathcal{C} est l'espace des chiffrés.

Au préalable Alice et Bob se sont mis d'accord sur une clé $K \in \mathcal{K}$. K est également vue comme une v.a de loi donnée.

Alice envoie à Bob le chiffré $C = E_K(M)$. Eve voit donc passer C . Quelle information peut-elle en déduire sur M ?

1.3 Les systèmes parfaits

Def. Dans un **système parfait au sens de Shannon**, la connaissance de C n'apporte aucune info sur M .

Deux variantes :

- en moyenne : $H(M | C) = H(M)$,
- dans tous les cas : égalité des lois, $\forall M, \forall C, p(M | C) = p(M)$.

Ex. On a les données suivantes :

$$\mathcal{M} = \{\text{oui}, \text{non}\}, \quad p(\text{oui}) = \frac{3}{4}, p(\text{non}) = \frac{1}{4}, \quad \mathcal{K} = \{K_1, K_2, K_3\}, \quad p(K_1) = p(K_2) = p(K_3) = \frac{1}{3}, \quad \mathcal{C} = \{x, y, z\}$$

et le chiffrement donné par

	oui	non
K_1	x	y
K_2	y	x
K_3	z	y

Eve voit C et en déduit que Alice a envoyé M avec probabilité $p(M | C)$. Par formule de Bayes il vient $p(\text{oui} | x) = \frac{p(x|\text{oui})p(\text{oui})}{p(x)}$, avec $p(x) = p(x | \text{oui})p(\text{oui}) + p(x | \text{non})p(\text{non})$ et idem avec y et z ou $M = \text{oui}$.

On trouve $p(x) = \frac{1}{3}$, $p(y) = \frac{5}{12}$ et $p(z) = \frac{1}{4}$.

Supposons $C = x$. Alors $p(\text{oui} | x) = \frac{3}{4}$ et $p(\text{non} | x) = \frac{1}{4}$ donc Eve n'a rien appris de plus que ce qu'elle connaissait déjà.

Supposons $C = y$. Alors $p(\text{oui} | y) = \frac{3}{5}$ et $p(\text{non} | y) = \frac{2}{5}$. L'incertitude est plus grande, en un sens Eve a appris quelque chose de nouveau par rapport à ce qu'elle estimait précédemment.

Supposons $C = z$. Alors $p(\text{oui} | z) = 1$ et $p(\text{non} | z) = 0$. Donc le clair se déduit du chiffré.

En conclusion ce système n'est pas parfait.

Ex (Un système parfait : le one-time pad, ou masque jetable). On prend $\mathcal{M} = \{0,1\}^n$, $\mathcal{K} = \{0,1\}^n$ avec une distribution uniforme des clés et $\mathcal{C} = \{0,1\}^n$. On prend $E_K(M) := M \oplus K$. Alors $\forall M, \forall C, p(M | C) = p(M)$. Cette méthode est lourde : clé de même longueur que le message.

Th. Dans un système parfait on a nécessairement $\text{Card}(\mathcal{K}) \geq \text{Card}(\mathcal{C})$.

Démonstration. Pour un M donné on a

$$\forall C, p(C | M) = \frac{p(C, M)}{p(M)} = \frac{p(C | M)}{p(M | C)} = p(C) > 0$$

donc il existe une clé K telle que $C = E_K(M)$, d'où le résultat. \square

Rem. On a toujours $\text{Card}(\mathcal{C}) \geq \text{Card}(\mathcal{M})$ pour pouvoir opérer un décodage (fonction E injective). Le cas le plus économique en système parfait est donc $\text{Card}(\mathcal{K}) = \text{Card}(\mathcal{C}) = \text{Card}(\mathcal{M})$ et alors $\forall M, \forall C, \exists ! K, C = E_K(M)$.

On peut ensuite en déduire que la distribution de K doit être uniforme \rightarrow c'est le one-time pad.

1.4 Distance d'unicité

Scénario : réutilisation d'une même clé pour chiffrer plusieurs messages.

Ex (Chiffrement par substitution). On prend $\mathcal{M} = \mathcal{C} = \{A, B, C, \dots, Z\}$ et \mathcal{K} l'ensemble des permutations de \mathcal{M} . La réutilisation d'une clé permet d'étendre $E: \mathcal{M} \times \mathcal{K} \rightarrow \mathcal{C}$ en $E: \mathcal{M}^n \times \mathcal{K} \rightarrow \mathcal{C}^n$.

Ce système s'attaque par analyse des fréquences dès lors qu'on sait que le clair est restreint à un certain sous-ensemble de \mathcal{M}^n , e.g. texte en français qui n'est pas une suite de lettres "très aléatoire".

Question : à partir de quelle longueur de chiffré peut-on retrouver la permutation clé ?

Prenons un alphabet \mathcal{A} . Alors $\mathcal{M} = \mathcal{A}^n$ avec une certaine redondance (non uniformité dans \mathcal{A}).

Def. Entropie du langage : $h = \lim_{n \rightarrow \infty} \frac{1}{n} H(M)$.

Def. Redondance du langage : $r = 1 - \frac{h}{\log_2(\text{Card}(\mathcal{A}))}$.

Intuitivement un texte de n symboles dans le langage peut se compresser en nh bits ($n \rightarrow \infty$).

En longueur n , parmi les $\text{Card}(\mathcal{A})^n$ suites de n symboles possibles, il y en a $\approx 2^{nh}$ qui sont dans le langage. Les chiffrés doivent sembler aléatoires, on peut espérer en avoir $\text{Card}(\mathcal{A})^n$.

À chaque chiffré correspond $\frac{2^{nh + \log_2(\#\mathcal{K})}}{(\#\mathcal{A})^n}$ couples (M, K) possibles.

Une recherche exhaustive donne un déchiffrement unique dès lors que $\frac{2^{nh + \log_2(\#\mathcal{K})}}{(\#\mathcal{A})^n} \leq 1$, où $\frac{2^{nh + \log_2(\#\mathcal{K})}}{(\#\mathcal{A})^n} = 2^{n(h - \log_2 \#\mathcal{A} + \log_2 \#\mathcal{K})} = 2^{-rn \log_2 \#\mathcal{A} + \log_2 \#\mathcal{K}}$.

Def. Distance critique : $n = \frac{\log_2 \#\mathcal{K}}{r \log_2 \#\mathcal{A}}$.

Ex. Pour une substitution simple, en français, $\#\mathcal{K} = 26!$, donc $\log_2 \#\mathcal{K} \approx 88$, $r \in [0,75; 0,8]$ et $\#\mathcal{A} = 26$ donc $\log_2 \#\mathcal{A} \approx 4,7$. Alors $n \approx 25$. En pratique, à la main, on peut casser le chiffrement pour n jusqu'à 200 ou 250.

Autre exemple avec sécurité informationnelle : partage de secret à seuil.

Ex (Le système de Shamir). Un "distributeur" dispose d'un secret $S \in \mathbb{F}_q$ et veut donner une part de ce secret à n participants de sorte que :

- t participants qui se concertent peuvent reconstruire le secret avec leurs parts,
- $t - 1$ participants n'ont aucune info sur S .

On supposera $q > n$. Le distributeur choisit $t - 1$ éléments $a_1, \dots, a_{t-1} \in \mathbb{F}_q$ uniformément indépendants. Il prend le polynôme $P(X) = S + a_1 X + a_2 X^2 + \dots + a_{t-1} X^{t-1}$.

Étant choisis $x_1, \dots, x_n \in \mathbb{F}_q^\times$ distincts, on calcule $y_i = P(x_i)$. Au participant n° i est envoyé le couple (x_i, y_i) . L'assignation des x_i peut être publique, la part de secret est contenue dans y_i .

Cela vérifie bien les propriétés voulues :

- Si t participants se concertent P est déterminé de façon unique par interpolation de Lagrange : $P(X) = \sum_{i=1}^t y_i \prod_{j \neq i} \frac{X - x_j}{x_i - x_j}$ et alors $S = P(0)$.
- Si $t - 1$ participants se concertent ils ne peuvent en déduire aucune information car toutes les valeurs de S sont compatibles par interpolation de Lagrange.

1.5 Sécurité computationnelle

En cryptographie symétrique, A et B possèdent une clé secrète commune. En cryptographie asymétrique on veut communiquer de façon sécurisée sans avoir eu la possibilité au préalable de se mettre d'accord sur un secret commun.

Historique : 1974, Merkle, projet de fin d'études *Énigmes*.

Bob prépare un grand nombre N de messages clairs qui disent « la clé $n^o i$ est k_i ». Il les chiffre de façon pas très sûre, cassable en temps T . Il publie ces chiffres (dans le désordre).

Alice choisit une devinette, elle la résout et apprend un couple (i, k_i) . Elle dit à Bob « je connais la clé $n^o i$ ». Eve ne sait pas quelle devinette correspond à la clé i , elle doit les résoudre toutes \rightarrow en moyenne temps $\frac{N}{2}T$.

Ex (Diffie-Hellman (1976)). Système d'échange de clé (key agreement).

Alice	Public	Bob
	\mathbf{F}_q	
	$\alpha \in \mathbf{F}_q^\times$	
$r \in \mathbf{Z}$ aléatoire		$s \in \mathbf{Z}$ aléatoire
$\beta = \alpha^r$	$\xrightarrow{\beta}$	
	$\xleftarrow{\gamma}$	
$\gamma^r = \alpha^{rs}$		$\gamma = \alpha^s$ $\beta^s = \alpha^{rs}$

À la fin α^{rs} est leur secret commun. Eve voit passer q, α, β, γ . Elle doit en déduire un certain δ tel que $\exists r, s, \delta = \alpha^{rs}, \beta = \alpha^r, \gamma^s$. C'est le *problème de Diffie-Hellman*. Cela ressemble (mais n'est pas équivalent) au problème du log discret : connaissant q, α, β , trouver r tel que $\beta = \alpha^r$.

Ex (RSA, 1978). Méthode de chiffrement et de signature.

Bob choisit p et q premiers secrets. Il calcule $N = pq$, alors $\varphi(N) = (p-1)(q-1)$. Il choisit $e \in \mathbf{N}$ premier à $\varphi(N)$ et d son inverse, i.e $ed = 1 \pmod{\varphi(N)}$.

La clé publique de Bob est (N, e) , où N est appelé *module RSA* et e est l'*exposant de chiffrement*. L'entier d sera appelé *exposant de déchiffrement* et constitue la clé secrète.

Alice veut envoyer un message clair à Bob encodé comme un élément $m \in (\mathbf{Z}/N\mathbf{Z})^\times$. Elle calcule $c = m^e$, le chiffré, et l'envoie à Bob. Bob déchiffre $c^d = m^{ed} = m$.

Eve peut vouloir :

- 1 - retrouver le secret primitif de Bob, i.e p et q , ensuite elle trouve $\varphi(N)$ puis $d = e^{-1}$ par Euclide,
- 2 - retrouver juste d à partir de (N, e) ,
- 3 - être capable de retrouver un m correspondant à un c ,
- 4 - retrouver un bit d'information sur m à partir de c où m est vu comme un entier dans $\llbracket 1; N-1 \rrbracket$.

Remarque : en théorie de la complexité on appelle "faciles" les problèmes résolubles en temps polynomial en la taille des entrées. On appelle difficile les problèmes résolubles en tant exponentiel \rightarrow sous classe NP.

En crypto on veut typiquement que le déchiffrement soit NP : facile si on connaît la clé et difficile si on ne la connaît pas.

Dans le cas de RSA :

- 1 \rightarrow problème de factorisation donné N , trouver ses facteurs premiers p et q . Le problème est réputé NP, mais ni P ni NP-complet.
- 2 \rightarrow on peut montrer que ça équivaut à 1.
- 3 \rightarrow extraction de racine $e^{\text{ième}}$ dans $\mathbf{Z}/N\mathbf{Z}$, $m = \sqrt[e]{c} \pmod{N}$. Si $e = 2$ on a vu que c'était équivalent à la factorisation de N . Mais lorsque e est impair (souvent $e = 3$), cela pourrait être plus facile que la factorisation mais estimé non polynomial.
- 4 \rightarrow équivalent à 3. Plus précisément on a une réduction polynomiale de l'un à l'autre. Soit A une "boîte noire" qui nous dit si $m < \frac{N}{2}$ ou non à partir du chiffré c . Or on a $2^e c = (2m)^e$ le chiffré correspondant au clair $2m \pmod{N}$. Si $m < \frac{N}{2}$, $2m \pmod{N} = 2m$ pair et si $m > \frac{N}{2}$, $2m \pmod{N} = 2m - N$ impair. Ensuite, par dichotomie, en $\log_2(N)$ appels à la boîte A on a complètement localisé m .

Ex. Public : \mathbf{F}_q et $\alpha \in \mathbf{F}_q^\times$. Clé publique de Bob : $\beta \in \mathbf{F}_q^\times$. Clé secrète : $r \in \mathbf{Z}$ tel que $\beta = \alpha^r$.

Alice veut envoyer le message en clair $m \in \mathbf{F}_q^\times$. Elle choisit $s \in \mathbf{Z}$ aléatoire et calcule $c_1 = \alpha^s$ et $c_2 = \beta^s m$. Elle envoie (c_1, c_2) à Bob qui déchiffre par $c_2 c_1^r = \beta^s m \alpha^{-rs} = m$.

2 Courbes elliptiques

2.1 Définitions

Def. Une **courbe elliptique** sur un corps K est

- soit la donnée d'une courbe algébrique E projective lisse de genre 1 sur K et d'un point $O_E \in E(K)$,

- soit la donnée d'une équation "de Weierstrass" de la forme $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ qui définit une courbe plane où les coefficients $a_1, a_2, a_3, a_4, a_6 \in K$ sont choisis pour que E soit lisse. La courbe admet alors un unique point à l'infini, noté O_E .

Rem. Lorsque K est de caractéristique différente de 2 ou 3, on se ramène par changement de variable à une équation de la forme $y^2 = x^3 + ax + b$. La lissité équivaut donc à ce que $x^3 + ax + b$ soit sans racine double, i.e. $\Delta = 4a^3 + 27b^2 \neq 0$ dans K .

2.2 Loi de groupe

Lem. Soit $D \in \text{Div}^0(E)$, alors $\exists! P \in E(K), D \sim (P) - (O_E)$.

On a donc une bijection
$$\begin{array}{ccc} E(K) & \xrightarrow{\sim} & Cl^0(E)_K \\ P & \mapsto & (P) - (O_E) \end{array}$$
 avec $Cl^0(E)_K$ le groupe des classes d'équivalence linéaire de diviseurs de degré 0 définis sur K .

Def. On munit $E(K)$ d'une loi de groupe $+$ en transportant la loi de $Cl^0(E)_K$ par cette bijection.

Prop. (i) L'élément neutre de $E(K)$ est O_E .

- (ii) $\forall P, Q \in E(K)$, $P+Q$ dans $E(K)$ est l'unique point tel que $(P) - (O_E) + (Q) - (O_E) \sim (P+Q) - (O_E)$ dans $\text{Div}^0(E)$, i.e. tel que $\exists f \in E(K), \text{div}(f) = (P) + (Q) - (P+Q) - (O_E)$.
- (iii) Soit $D = \sum_{P \in E(K)} n_P \cdot (P)$ un diviseur sur E . Alors D est principal si et seulement si $\deg(D) = \sum_P n_P = 0$ et $\sum_P n_P P = O_E$ dans $E(K)$.
- (iv) En particulier $P + Q + R = O_E \iff (P) - (O_E) + (Q) - (O_E) + (R) - (O_E) \sim 0$ dans $\text{Div}^0(E)$.
- (v) $\forall P \in E(K)$, $-P \in E(K)$ est l'unique point tel que $\exists f, \text{div}(f) = (P) + (-P) - 2(O_E)$.

Rem. On a $P + Q + R = O_E$ si et seulement si P, Q, R sont les trois points d'intersection de E et d'une droite.

Rem. Si P est un point de coordonnées affines (x_P, y_P) alors $-P$ a pour coordonnées $(x_P, -y_P)$.