Chapitre 7 : Cryptographie classique

SÉcurité et Cryptographie
2013-2014

Sup Galilée INFO3

Introduction

L'objectif fondamental de la cryptographie est de permettre à deux personnes, traditionnellement appelées Alice et Bob, de communiquer au travers d'un canal peu sûr de telle sorte qu'un opposant, Oscar, ne puisse comprendre ce qui est échangé.

Le canal peut être par exemple une ligne de téléphone, Internet, ou autre.

L'information qu'Alice souhaite transmettre à Bob, que l'on appelle texte (ou message) clair, peut être un texte écrit en français ou encore des données numériques.

Alice transforme le texte clair par un procédé de chiffrement, en utilisant une clef prédéterminée, et envoie le texte (ou message) chiffré (ou encore cryptogramme) au travers du canal. Oscar, qui espionne éventuellement le canal, ne peut retrouver le texte clair, mais Bob, qui connaît la clef pour déchiffrer, peut récupére le message clair à partir du cryptogramme.

Définition formelle

Un procédé (ou système ou algorithme) de chiffrement (ou cryptosystème est un quintuplet $A = (\mathcal{P}, \mathcal{C}, \mathcal{K}, E, D)$ où

- $\mathcal{P}$ est un ensemble fini de blocs de textes clairs possibles,
- $\mathcal{C}$ est un ensemble fini de blocs de textes chiffrés possibles,
- $\mathcal{K}$ est un ensemble fini de clefs possibles,
- E est une fonction de chiffrement, $E : \mathcal{P} \times \mathcal{K} \to \mathcal{C}$,
- D est une fonction de déchiffrement $D : \mathcal{C} \times \mathcal{K} \to \mathcal{P}$,
- Pour chaque clef K, il existe au moins une clef K' telle que pour tout $x \in \mathcal{P}$,

$$D(E(x, K), K') = x \ .$$

Définition formelle

La dernière propriété est fondamentale. Elle précise que si un texte clair x est chiffré en un cryptogramme y avec K, alors il existe une clef K' telle que y déchiffré avec K' redonne x.

En termes mathématiques, cela signifie que pour tout K, l'application $E(-, K) : \mathcal{P} \to \mathcal{C}$ est injective, et que $D(-, K') : \mathcal{C} \to \mathcal{P}$ est surjective.

Remarquons que les ensembles $\mathcal{P}$ et $\mathcal{C}$ sont supposés finis donc, dans le cas où $|\mathcal{P}| = |\mathcal{C}|$, les fonctions $E(-, K)$ et $D(-, K')$ sont en fait bijectives, et inverses l'une de l'autre pour toute clef K.

Supposons qu'Alice souhaite transmettre un message à Bob, ce message étant une suite

$$\mathbf{x} = x_1 x_2 \cdots x_n$$

où chaque x_i est un élément de $\mathcal{P}$, en utilisant une clef K. Alors chaque x_i est chiffré avec K. Ainsi Alice calcule $y_i = E(x_i, K)$, $i = 1, \cdots, n$, et transmet à Bob la suite

$$\mathbf{y} = y_1 \cdots y_n \ .$$

Si Bob connaît K', alors il peut déchiffrer le message. Pour ce faire, il calcule pour chaque $i = 1, \cdots, n$, $D(y_i, K') = D(E(x_i, K), K') = x_i$ et récupère donc $\mathbf{x}$.

Le chiffrement par décalage : arithmétique modulaire

Avant d'introduire ce cryptosystème, commençons par la description de l'arithmétique modulaire.

Définition

Si a, b et m sont des entiers, avec $m > 0$, on écrit $a \equiv b \pmod{m}$, et on dit que a est congru à b modulo m, si m divise $a - b$. L'entier m est parfois appelé le modulus.

Remarquons immédiatement que si $m = 1$, alors pour tous a, b, $a \equiv b \mod m$.

Supposons que l'on divise a et b par m (division entière). On obtient alors $a = q_1 m + r_1$ et $b = q_2 m + r_2$ (avec les restes r_1, r_2 satisfaisant $0 \leq r_1, r_2 \leq m - 1$). Il est alors facile de voir que $a \equiv b \pmod{m}$ si, et seulement si, $r_1 = r_2$. Dans la suite, le reste d'un entier a modulo m sera noté $a \mod m$ (sans les parenthèses). Il en résulte que l'on a $a \equiv b \pmod{m}$ si, et seulement si, $a \mod m = b \mod m$. Si on remplace a par $a \mod m$, on dit que l'on réduit a modulo m.

On est maintenant en mesure de définir l'arithmétique modulo m : $\mathbb{Z}/m\mathbb{Z}$ désigne l'ensemble $\{0, \cdots, m\}$ muni de deux opérations $+$ et $\times$.
L'addition et la multiplication dans $\mathbb{Z}/m\mathbb{Z}$ fonctionnent comme l'addition et la multiplication usuelles, excepté le fait que tous les résultats sont réduits modulo m.

Supposons par exemple que l'on veuille calculer 11×13 dans $\mathbb{Z}/16\mathbb{Z}$. Comme entiers ordinaires, on a $11 \times 13 = 143$. Pour réduire 143 modulo 16, on effectue une division euclidienne (ou entière) : $143 = 8 \times 16 + 15$, et donc $143 \mod 16 = 15$, de sorte que $11 \times 13 = 15$ dans $\mathbb{Z}/16\mathbb{Z}$.

Propriétés de +

Avec ces définitions, l'addition et de la multiplication dans $\mathbb{Z}/m\mathbb{Z}$ satisfont la plupart des règles familières en arithmétique. On rappelle la liste de ces propriétés sans les démontrer :

- L'addition est interne : si $a, b \in \mathbb{Z}/m\mathbb{Z}$, alors $a + b \in \mathbb{Z}/m\mathbb{Z}$.
- L'addition est commutative : si $a, b \in \mathbb{Z}/m\mathbb{Z}$, alors $a + b = b + a$.
- L'addition est associative : si $a, b, c \in \mathbb{Z}/m\mathbb{Z}$, alors $(a + b) + c = a + (b + c)$.
- 0 est neutre pour l'addition : si $a \in \mathbb{Z}/m\mathbb{Z}$, alors $0 + a = a = a + 0$.
- Chaque élément $a \in \mathbb{Z}/m\mathbb{Z}$ admet un opposé c'est-à-dire un élément $-a$ (unique) tel que $(-a) + a = 0 = a + (-a)$.

Notons que pour tout $a \in \mathbb{Z}/m\mathbb{Z}$, $a \neq 0$, $-a = m - a$. (En effet, $a + (m - a) = 0$.) Évidemment, $-0 = 0$.

Propriétés de $\times$

- La multiplication est interne : si $a, b \in \mathbb{Z}/m\mathbb{Z}$, alors $a \times b \in \mathbb{Z}/m\mathbb{Z}$.
- La multiplication est commutative : si $a, b \in \mathbb{Z}/m\mathbb{Z}$, alors $a \times b = b \times a$.
- La multiplication est associative : si $a, b, c \in \mathbb{Z}/m\mathbb{Z}$, alors $(a \times b) \times c = a \times (b \times c)$.
- 1 est neutre pour La multiplication : si $a \in \mathbb{Z}/m\mathbb{Z}$, alors $1 \times a = a = a \times 1$.
- La multiplication est distributive sur l'addition : si $a, b, c \in \mathbb{Z}/m\mathbb{Z}$, alors $a \times (b + c) = a \times b + a \times c$.

Les propriétés de $+$ font de $\mathbb{Z}/m\mathbb{Z}$ un groupe abélien, celles de $\times$ font de $\mathbb{Z}/m\mathbb{Z}$ un monoïde commutatif, et la totalité de ces propriétés en fait un anneau commutatif.

Puisque les opposés existent dans $\mathbb{Z}/m\mathbb{Z}$, on peut également réaliser des soustractions. Pour $a, b \in \mathbb{Z}/m\mathbb{Z}$, $a - b = a + (-b)$.

On obtient $a - b$ comme suit : On calcule $a - b$ comme des entiers usuels, puis on réduit modulo m.

Exemple : Calculer $11 - 18$ dans $\mathbb{Z}/31\mathbb{Z}$ (de deux façons différentes). On peut calculer -18 dans $\mathbb{Z}/31\mathbb{Z}$, soit $-18 = 13$, puis calculer $11 + 13$ mod m, soit 24. On peut aussi calculer directement $11 - 18 = -7$, puis réduire -7 modulo 31, soit $31 - 7 = 24$.

Chiffrement par décalage

Il est défini par les données suivantes : $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}/26\mathbb{Z}$.

Pour $0 \leq K \leq 25$ et $0 \leq x \leq 25$, on définit

$$E(x, K) = x + K \mod 26$$

et

$$D(x, K) = y - K \mod 26 \ .$$

Remarque

Lorsque $K = 3$, le système par décalage s'appelle le chiffrement de César car il était utilisé par un certain Caius Iulius Caesar IV (-100 à -44).

On peut utiliser le chiffrement par décalage pour chiffrer un texte ordinaire en décidant d'une correspondance entre les caractères alphabétiques et les résidus modulo 26 comme donné dans la table suivante :

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Exercice

Déchiffrer le message suivant en utilisant la clef $K = 11$.
HPHTWWXPPELEXTOYTRSE

Pour qu'un système de chiffrement soit utilisable en pratique, il doit satisfaire certaines propriétés :

- La fonction de chiffrement E et la fonction de déchiffrement D doivent être calculables efficacement.
- Un opposant observant le texte chiffré $\mathbf{y}$ doit être incapable de déterminer ni la clef K utilisée, ni le texte clair $\mathbf{x}$.

La seconde propriété définit de manière informelle la notion de "sécurité". L'opération consistant à rechercher la clef K à partir du texte chiffré $\mathbf{y}$ est appelée cryptanalyse. On observe que si Oscar peut retrouver K, alors il peut retrouver $\mathbf{x}$ comme le fait Bob en utilisant D. Donc retrouver la clef K est au moins aussi difficile que de retrouver le texte clair.

On remarque que le chiffrement par décalage n'est pas sûr, car il peut être cryptanalysé par la méthode de recherche exhaustive ou force brute. Comme il n'y a que 26 clefs possibles, essayer le déchiffrement avec toutes les clefs jusqu'à trouver un texte clair compréhensible est aisé.

Exercice

Déchiffrer le message (chiffré avec le système par décalage) par force brute.
JBCRCLQRWCRVNBJENBWRWN.

Solution : clef=9 et message="A stitch in time saves nine" ("Un point à temps en vaut cent" autrement dit l'entretien consciencieux permet d'éviter le gaspillage).

Chiffrement par substitution

Un autre procédé de chiffrement bien connu est le chiffrement par substitution.

Il est défini comme suit : $\mathcal{P} = \mathcal{C} = A$, où A désigne l'alphabet des 26 lettres usuelles. L'ensemble des clefs $\mathcal{K}$ est l'ensemble $\mathfrak{S}_A$ des permutations de A. Rappelons ici qu'une permutation sur un ensemble X est une bijection de X sur lui-même.

Pour chaque permutation $\pi \in \mathfrak{S}_A$, on définit

$$E(\alpha, \pi) = \pi(\alpha)$$

et

$$D(\alpha, \pi) = \pi^{-1}(\alpha)$$

pour $\alpha \in A$.

Exercice

Soit la clef suivante :

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

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

- Chiffrer le message suivant “On ne peut rien apprendre aux gens. On peut seulement les aider à découvrir qu’ils possèdent déjà en eux tout ce qui est à apprendre” (Galilée).
- Déchiffrer le message suivant “MGZVYZLGHCMHJMYXSSFMNHAHYCDLMHA”.
- Quel est le nombre total de clefs possibles dans le chiffrement par substitution ? (26!)

Chiffrement affine

Le chiffrement par décalage est un cas particulier du chiffrement par substitution (qui n'utilise que 26 des 26! clefs possibles) dans la mesure où l'on peut voir le premier comme un système de chiffrement sur l'alphabet A (des 26 lettres usuelles, par un codage des lettres en des nombres) et que chaque fonction de chiffrement/déchiffrement est une permutation.

Un autre cas particulier du chiffrement par substitution est le chiffrement affine. Dans ce procédé, on limite les fonctions de chiffrement à celles de la forme

$$E(x, K) = ax + b \mod 26$$

où $a, b \in \mathbb{Z}/26\mathbb{Z}$ et $K = (a, b)$.

Les fonctions de la forme $E(-, K)$ sont appelées des fonctions affines (par analogie avec les droites affines de la géométrie du plan réel). Observons que l'on retrouve le chiffrement par décalage pour $a = 1$.

Pour que l'opération de chiffrement soit possible, il est nécessaire que la fonction affine soit injective donc bijective. Autrement dit, pour chaque $y \in \mathbb{Z}/26\mathbb{Z}$, l'équation

$$ax + b \equiv y \pmod{26}$$

doit avoir une unique solution x (pour a, b fixés). Cette équation est équivalente à $ax \equiv y - b \pmod{26}$. Lorsque y parcourt l'ensemble $\mathbb{Z}/26\mathbb{Z}$, $y - b$ parcourt également $\mathbb{Z}/26\mathbb{Z}$ (puisque $y \mapsto y - b$ est une bijection). Il suffit donc d'étudier l'équation $ax \equiv y \pmod{26}$ pour chaque $y \in \mathbb{Z}/26\mathbb{Z}$.

Solutions de l'équation $ax \equiv y \pmod{26}$

Théorème

Pour chaque $y \in \mathbb{Z}/26\mathbb{Z}$, l'équation $ax \equiv y \pmod{26}$ admet une unique solution si, et seulement si, $pgcd(a, 26) = 1$.

Preuve : Supposons pour commencer que $pgcd(a, 26) = d > 1$. Alors l'équation $ax \equiv 0 \pmod{26}$ admet au moins deux solutions distinctes dans $\mathbb{Z}/26\mathbb{Z}$, à savoir $x = 0$ et $x = 26/d$. La fonction $E(-, (a, 0)) : x \mapsto ax \bmod 26$ n'est donc pas injective.
Supposons maintenant que $pgcd(a, 26) = 1$. Soient x_1, x_2 tels que $ax_1 \equiv ax_2 \pmod{26}$. Alors $a(x_1 - x_2) \equiv 0 \pmod{26}$ et donc 26 divise $a(x_1 - x_2)$. On utilise une propriété de la division : si $pgcd(a, b) = 1$ et si a divise bc, alors a divise c. Comme 26 divise $a(x_1 - x_2)$ et $pgcd(a, 26) = 1$, il en résulte que 26 divise $x_1 - x_2$, de sorte que $x_1 \equiv x_2 \pmod{26}$. On vient de montrer que si $pgcd(a, 26) = 1$, alors l'équation $ax \equiv y \pmod{26}$ admet au plus une solution dans $\mathbb{Z}/26\mathbb{Z}$. Donc quand on fait varier x dans $\mathbb{Z}/26\mathbb{Z}$, $ax \pmod{26}$ ne peut pas prendre deux fois la même valeurs, et donc parcourt également $\mathbb{Z}/26\mathbb{Z}$. Donc, pour chaque $y \in \mathbb{Z}/26\mathbb{Z}$, l'équation $ax \equiv y \pmod{26}$ admet une solution unique.

Remarque

Le résultat précédent reste vrai si 26 est remplacé par un entier $m > 0$ quelconque.

Soit $a \in \mathbb{Z}/m\mathbb{Z}$. Si $pgcd(a, m) = 1$, alors il existe en particulier une unique solution à l'équation $ax \equiv 1 \pmod{m}$. Dans ce cas on dit que a est inversible modulo m.

Exercice

- Calculer $pgcd(4, 26)$. $(= 2)$.
- Trouver un entier $n \in \mathbb{Z}/26\mathbb{Z}$ tel que $E(x, (4, 0)) = E(x + n, (4, 0))$ pour tout $x \in \mathbb{Z}/26\mathbb{Z}$. $(n = 13)$.
- Donner la liste des éléments x de $\mathbb{Z}/26\mathbb{Z}$ tels que $pgcd(x, 26) = 1$.(Puisque $26 = 2 \times 13$, les éléments x tels que $pgcd(x, 26) = 1$ sont $1, 3, 5, 9, 11, 15, 17, 19, 21, 23, 25$.)
- Soient m un entier > 0 et $a \in \mathbb{Z}/m\mathbb{Z}$. Montrer que si a est inversible modulo m, alors $pgcd(a, m) = 1$. (L'équation $ax \equiv y \pmod{26}$ admet une unique solution !) En déduire que a est inversible modulo m si, et seulement si, $pgcd(a, m) = 1$.
- Expliquer pourquoi tout élément $x \in \mathbb{Z}/m\mathbb{Z}$, qui est un entier premier, est nécessairement inversible modulo m.

Comme $26 = 2 \times 13$, les valeurs a de $\mathbb{Z}/26\mathbb{Z}$ telles que $pgcd(x, 26) = 1$ sont $1, 3, 5, 9, 11, 15, 17, 19, 21, 23, 25$. Le paramètre peut quant à lui être quelconque dans $\mathbb{Z}/26\mathbb{Z}$. Le chiffrement affine admet donc $12 \times 26 = 312$ clefs possibles, ce qui est évidemment trop petit pour être sûr.

Considérons maintenant le cas général où le modulus est $m > 0$. On utiliser une définition issue de la théorie des nombres.

Définition

Soient des entiers $a \geq 1$ et $m \geq 2$. Si $pgcd(a, m) = 1$, on dit que a et m sont premiers entre eux. Le nombre d'entiers de $\mathbb{Z}/m\mathbb{Z}$ qui sont premiers avec m est noté $\phi(m)$, et ϕ est appelé la fonction indicatrice d'Euler.

Un résultat bien connu de la théorie des nombres donne la valeur de $\phi(m)$ à partir de sa décomposition en puissances de facteurs premiers. (Rappelons ici qu'un entier $p > 0$ est dit premier s'il est distinct de 1, et n'est divisible que 1 et par lui-même. Tout entier $m > 1$ se factorise en produit de puissances de nombres premiers de manière unique, à l'ordre près des facteurs. Par exemple, $60 = 2^2 \times 3 \times 5$ et $98 = 2 \times 7^2$.)

On rappelle la formule de $\phi(m)$ dans le théorème suivant (non démontré ici).

Définition

Supposons que $m = \prod_{i=1}^{n} p_i^{e_i}$ où les p_i sont des nombres premiers deux-à-deux distincts, et les $e_i > 0$ pour $1 \leq i \leq m$. On a $\phi(m) = \prod_{i=1}^{n} (p_i^{e_i} - p_i^{e_i - 1})$.

Si on considère le chiffrement affine sur $\mathbb{Z}/m\mathbb{Z}$ (le même que celui que l'on a vu en remplaçant 26 par m), alors le nombre de clefs est $m\phi(m)$. (Dans la fonction de chiffrement $E(-,(a,b))$, m compte le nombre de b possibles, et $\phi(m)$ le nombre de a possibles.) Par exemple, si $m = 60 = 2^2 \times 3 \times 5$, $\phi(60) = (2^2 - 2)(3-1)(5-1) = 2 \times 2 \times 4 = 16$, et le nombre total des clefs est $16 \times 60 = 960$.

Considérons maintenant la fonction de déchiffrement dans le chiffrement affine avec $m = 26$. Supposons que $pgcd(a, 26) = 1$. Pour déchiffrer on a besoin de résoudre $y \equiv ax + b \pmod{26}$ pour chaque x. Le raisonnement précédent donne l'existence d'une unique solution (car $pgcd(a,b) = 1$) mais ne nous donne pas de moyen de la calculer (et encore moins pour la calculer de façon efficace). On utilise maintenant la notion d'inverse.

Définition

Soit $a \in \mathbb{Z}/m\mathbb{Z}$. L'inverse de a est un élément $a^{-1} \in \mathbb{Z}/m\mathbb{Z}$ tel que $aa^{-1} \equiv 1 \pmod{m}$.

Par des arguments semblables à ceux déjà utilisés, on peut montrer que a admet un inverse modulo m si, et seulement si, $pgcd(a, m) = 1$.

Si un inverse existe, alors il est unique. (Supposons que a admette deux inverses x et y, alors $y \equiv (xa)y = xay = x(ay) \equiv x \pmod{m}$.) Et si $a = b^{-1}$, alors $b = a^{-1}$. (On a $ba = bb^{-1} = 1$, donc $b = a^{-1}$ par unicité de l'inverse de a.)

On remarque que si m est un entier premier p, alors tous les éléments non nuls de $\mathbb{Z}/p\mathbb{Z}$ admettent un inverse. On dit dans ce cas que $\mathbb{Z}/p\mathbb{Z}$ est un corps.

Nous ne connaissons pas (encore) d'algorithme efficace pour le calcul des inverses. Par exemple, pour trouver l'inverse de 7 modulo 26 il faut résoudre $7 \times x \equiv 1 \pmod{26}$, soit trouver deux entiers x et q, $1 \leq x \leq 25$, tels que $7x = q26 + 1$. Dans ce cas $x = 7^{-1}$. En testant on trouve $7 \times 15 = 105 \equiv 1 \pmod{26}$, de sorte que $15 = 7^{-1}$. Il en résulte immédiatement que $15^{-1} = 7$.

L'équation $ax + b \equiv y \pmod{26}$ est équivalente à $ax \equiv y - b \pmod{26}$. Comme $pgcd(a, 26) = 1$, a admet un inverse modulo 26. En multipliant les deux membres de l'équation par a^{-1}, on obtient $a^{-1}(ax) \equiv a^{-1}(y - b) \pmod{26}$. Par associativité de la multiplication dans $\mathbb{Z}/26\mathbb{Z}$, le membre de gauche devient $a^{-1}(ax) \equiv (a^{-1}a)x \equiv 1x \equiv x \pmod{26}$. En conséquence de quoi, $x \equiv a^{-1}(y - b) \pmod{26}$.

Le procédé de chiffrement affine est donc défini comme suit :
$\mathcal{P} = \mathcal{C} = \mathbb{Z}/26\mathbb{Z}$, et $\mathcal{K} = \{ (a, b) \in \mathbb{Z}/26\mathbb{Z} \times \mathbb{Z}/26\mathbb{Z} : pgcd(a, 26) = 1 \}$.
Pour $K = (a, b) \in \mathcal{K}$, on définit :

$$E(x, (a, b)) = ax + b \mod 26$$

et

$$D(y, (a, b)) = a^{-1}(y - b) \mod 26$$

pour $x, y \in \mathbb{Z}/26\mathbb{Z}$.

Exemple d'utilisation

Supposons que $K = (7, 3)$. Rappelons que $7^{-1} \mod 26 = 15$. La fonction de chiffrement est

$$E(x, K) = 7x + 3$$

et celle de déchiffrement est $D(y, K) = 15(y - 3) = 15y - 19$.

Vérifions que $D(E(x, K), K) = x$ pour tout $x \in \mathbb{Z}/26\mathbb{Z}$. On a

$$\begin{aligned} D(E(x, K), K) &= D(7x + 3, K) \\ &= 15(7x + 3) - 19 \\ &= x + 45 - 19 \\ &= x \ . \end{aligned} \tag{1}$$

Chiffrer le mot **affine** avec la même clef $K = (7, 3)$ (le résultat doit être donné sous la forme d'un message formé avec des lettres). On convertit tout d'abord les lettres en nombres ($0 \leftrightarrow a$, $1 \leftrightarrow b$, ...), puis on calcule $E(x, (7, 3))$ pour chacun des nombres x obtenus. Cela donne une suite de 6 nombres que l'on re-transforme en lettre (en utilisant le même codage en sens inverse).

Chiffrement de Vigenère

Dans le cas du chiffrement par décalage ou par substitution, dès qu'une clef est fixée, chaque caractère alphabétique, partout où il apparaît dans le texte, est transformé en un même caractère. Autrement dit, pour toute lettre α, chaque occurrence de α dans le texte clair est transformée en $E(\alpha, K)$ (pour être précis il faudrait remplacer α dans le chiffrement par son codage en un nombre...). Pour cette raison, le procédé est dit monoalphabétique.

On présente maintenant un chiffrement qui n'est pas monoalphabétique : le chiffrement de Vigenère.
Soit m un entier strictement positif. Soit $\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbb{Z}/26\mathbb{Z})^m$. Pour toute clef $K = (k_1, \cdots, k_m)$ (où $k_i \in \mathbb{Z}/26\mathbb{Z}$ pour chaque $i = 1, \cdots, m$), on définit

$$E(x_1, x_2, \cdots, x_m, K) = (x_1 + k_1, x_2 + k_2, \cdots, x_m + k_m)$$

et

$$D(y_1, y_2, \cdots, y_m, K) = (y_1 - k_1, y_2 - k_2, \cdots, y_m - k_m)$$

où les opérations sont effectuées dans $\mathbb{Z}/26\mathbb{Z}$.

En utilisant la correspondance $0 \leftrightarrow a$, $1 \leftrightarrow b$, ..., $25 \leftrightarrow z$, on décrit chaque clef K du chiffrement de Vigenère par une chaîne de caractères de longueur m appelée mot-clef.

Le chiffrement de Vigenère traite m caractères alphabétiques à la fois : chaque bloc du texte clair est équivalent à m caractères alphabétiques.

Exercice

Déchiffrer le texte suivant (chiffré avec la méthode de Vigenère et le mot-clef “CIPHER") :
VPXZGIAXIVWPUBTTMJPWIZITWZT.

Solution : La clef correspondant au mot-clef est $(2, 8, 15, 7, 4, 17)$. Et le texte déchiffré est
THISCRYPTOSYSTEMISNOTSECURE.

Quel est le nombre de clefs possibles dans le chiffrement de Vigenère ?
$\mathbf{26^m}$.

Déjà pour $m = 5$, cela représente plus de dix millions de possibilités (11 881 673). C'est donc assez grand pour exclure une recherche exhaustive "à la main" (mais pas avec un ordinateur).

Dans le chiffrement de Vigenère avec un mot-clef de longueur m, un caractère alphabétique peut être transformé en m caractères différents (au plus), si on suppose que le mot-clef contient m caractères deux-à-deux distincts. Un tel procédé est dit polyalphabétique. En général, la cryptanalyse est plus difficile dans de tels systèmes.

Chiffrement de Hill

On décrit maintenant un autre système cryptographique polyalphabétique appelé chiffrement de Hill. Soit m un entier strictement positif, et soit $\mathcal{P} = \mathcal{C} = (\mathbb{Z}/26\mathbb{Z})^m$. L'idée consiste à transformer m caractères d'un bloc de texte clair en m caractères d'un bloc de texte chiffré par des combinaisons linéaires.

Si $m = 2$, alors pour un bloc de texte clair $x = (x_1, x_2)$, on obtient un bloc de texte chiffré $y = (y_1, y_2)$ où y_1 et y_2 sont obtenus comme combinaisons linéaires de x_1 et x_2. Par exemple, $y_1 = 11x_1 + 3x_2$, $y_2 = 8x_1 + 7x_2$.
(L'addition et la multiplication sont réalisées dans $\mathbb{Z}/26\mathbb{Z}$.)
On peut bien entend écrire cela en sous la forme d'un produit matriciel :
$(y_1, y_2) = (x_1, x_2)\begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix}$.

En général, on prend une matrice carrée de taille $m \times m$ pour clef K. Si le coefficient (i, j) de la matrice est $k_{i,j}$, on écrit $K = (k_{i,j})$. Pour $x = (x_1, \cdots, x_m) \in \mathcal{P}$, et $K \in \mathcal{K}$, on calcule $y = E(x, K) = (y_1, \cdots, y_m)$ ainsi

$$(y_1, \cdots, y_m) = (x_1, \cdots, x_m) \begin{pmatrix} k_{1,1} & \cdots & k_{1,m} \\ \vdots & \ldots & \vdots \\ k_{m,1} & \cdots & k_{m,m} \end{pmatrix}.$$

C'est-à-dire que l'on calcule $y = xK$.

On dit que le texte chiffré est obtenu par une transformation linéaire. Pour voir comment le procédé de chiffrement fonctionne, on doit trouver comment calculer x à partir de y. Si vous vous souvenez de vos cours d'algèbres linéaires, alors vous savez que l'on utilise la matrice inverse K^{-1}. En effet, le texte clair est calculé par la formule $x = yK^{-1}$.

Un bref rappel d'algèbre linéaire élémentaire

Si $A = (a_{i,k})$ est une matrice de taille $m \times \ell$ et $B = (b_{k,j})$ est une matrice de taille $\ell \times n$, alors on définit le produit matriciel $AB = (c_{i,j})$ par la formule

$$c_{i,j} = \sum_{k=1}^{\ell} a_{i,k} b_{k,j}$$

pour $1 \leq i \leq m$, $1 \leq j \leq n$. Ainsi AB est une matrice de taille $m \times n$.

Ce produit matriciel est associatif (c'est-à-dire $(AB)C = A(BC)$, avec A de taille $m \times \ell$, B de taille $\ell \times n$ et C de taille $n \times p$; le résultat étant de taille $m \times p$), mais n'est pas commutatif en général (on a rarement $AB = BA$).

La matrice identité $I_m = (a_{i,j})$ de taille $m \times m$ est constituée de uns sur sa diagonale ($a_{i,i} = 1$) et de zéros partout ailleurs ($a_{i,j} = 1$ pour tous $i \neq j$).

Par exemple pour $m = 3$, on a $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Cette matrice I_m s'appelle matrice identité car elle satisfait $AI_m = A$ et $I_m B = B$ pour toutes matrices A de taille $\ell \times m$, ℓ quelconque, et B de taille $m \times k$, k quelconque.

La matrice inverse d'une matrice A carrée de taille $m \times m$ est (lorsqu'elle existe) une matrice notée A^{-1} telle que $AA^{-1} = I_m = A^{-1}A$. Certaines matrices ne possèdent pas d'inverse, mais si l'inverse existe, il est unique. L'inverse de I_m est I_m. Si $A = B^{-1}$, alors $B = A^{-1}$.

Retour un instant au chiffrement de Hill

Si la clef K choisie est une matrice inversible, alors il est facile de trouver la formule de déchiffrement : comme $y = xK$, on peut multiplier les deux membres de l'égalité par K^{-1}, et on obtient
$yK^{-1} = (xK)K^{-1} = x(KK^{-1}) = xI_m = x$.

On peut démontrer que l'on peut déchiffrer un message chiffré avec la méthode de Hill si, et seulement si, la clef K choisie est une matrice inversible.

L'inversibilité d'une matrice (carrée) dépend de la valeur de son déterminant. Contentons-nous détudier le cas des matrices 2×2.

Définition

Le déterminant d'une matrice $A = (a_{i,j}) = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$ de taille 2×2 est la valeur

$$\det A = a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \ .$$

Deux propriétés essentielles du déterminant sont $\det I_m = 1$ et $\det(AB) = \det A \times \det B$.

Par ailleurs il est bien connu qu'une matrice à coefficients complexes (ou réels) est inversible si, et seulement si, son déterminant est non nul.

Cependant ce n'est pas vrai dans $\mathbb{Z}/26\mathbb{Z}$ ni dans $\mathbb{Z}/m\mathbb{Z}$ en général.

Le résultat analogue dans le cas général est celui-ci : une matrice A à coefficients dans $\mathbb{Z}/m\mathbb{Z}$ est inversible si, et seulement si, son déterminant est inversible modulo m (c'est-à-dire $pgcd(\det A, m) = 1$).

Voyons une ébauche de la démonstration de ce résultat. On suppose tout d'abord que l'on a $pgcd(\det A, m) = 1$. Le nombre $\det A$ est donc inversible dans $\mathbb{Z}/m\mathbb{Z}$. Pour $1 \leq i \leq m$ et $1 \leq j \leq m$, soit $A(i,j)$ la matrice obtenue à partir de A en supprimant la $i$ème ligne et la $j$ème colonne. On définit ensuite $co(A)$ dont le coefficient i,j est $(-1)^{i+j} \det A(j,i)$. Cette matrice $co(A)$ est appelée la comatrice de A. On peut alors montrer que $A^{-1} = (\det A)^{-1} co(A)$, et donc que A est inversible. Inversement , supposons que A ait un inverse A^{-1}. D'après la règle de calcul des déterminant, $1 = \det I_m = \det(AA^{-1}) = \det(A)\det(A^{-1})$, de sorte que $\det A^{-1}$ est inversible dans $\mathbb{Z}/m\mathbb{Z}$.

Dans le cas des matrices 2×2, on a la formule suivante :

Théorème

Si $A = (a_{i,j})$ est une matrice 2×2 à coefficients dans $\mathbb{Z}/26\mathbb{Z}$ telle que $\det A$ est inversible modulo 26. Alors on a $A^{-1} = (\det A)^{-1} \begin{pmatrix} a_{2,2} & -a_{1,2} \\ -a_{2,1} & a_{1,1} \end{pmatrix}$

Par exemple, considérons la matrice $A = \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix}$.

On a

$$\begin{aligned} \det A &\equiv 11 \times 7 - 3 \times 8 \pmod{26} \\ &\equiv 77 - 24 \pmod{26} \\ &\equiv 53 \pmod{26} \\ &\equiv 1 \pmod{26}. \end{aligned} \tag{2}$$

Comme $1^{-1} \mod 26 = 1$, on a $\begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix}^{-1} = \begin{pmatrix} 7 & 18 \\ 23 & 11 \end{pmatrix}$.

Formellement, le procédé de Hill est défini comme suit :
$\mathcal{P} = \mathcal{C} = (\mathbb{Z}/26\mathbb{Z})^m$ et
$\mathcal{K} = \mathbf{GL}_m(\mathbb{Z}/26\mathbb{Z}) = \{\text{ matrices } m \times m \text{ inversibles dans } \mathbb{Z}/26\mathbb{Z}\}$.

Pour toute clef K, on définit

$$E(x, K) = xK$$

et

$$D(y, K) = yK^{-1}$$

où toutes les opérations sont effectuées dans $\mathbb{Z}/26\mathbb{Z}$, et $x, y \in (\mathbb{Z}/26\mathbb{Z})^m$.

Chiffrement par permutation

Tous les systèmes de chiffrement rencontrés précédemment dans ce cours reposent sur une substitution : tout caractère du texte clair est remplacé par un autre dans le texte chiffré. L'idée du chiffrement par permutation est de conserver les mêmes caractères en les réordonnant.

Soit un entier m strictement positif. Soit $\mathcal{P} = \mathcal{C} = \{0, 1, \cdots, 25\}^m$, et soit $\mathcal{K}$ l'ensemble $\mathfrak{S}_{\{1,\cdots,m\}}$ des permutations de $\{1, \cdots, m\}$. Pour toute clef π, on définit

$$E((x_1, \cdots, x_m), \pi) = (x_{\pi(1)}, \cdots, x_{\pi(m)})$$

et

$$D((y_1, \cdots, y_m), \pi) = (y_{\pi^{-1}(1)}, \cdots, y_{\pi^{-1}(m)})$$

où π^{-1} désigne la permutation inverse de π.

Comme pour le chiffrement par substitution, il est plus commode d'utiliser l'ensemble des caractères alphabétique A plutôt que $\mathbb{Z}/26\mathbb{Z}$, car il n'y a pas ici d'opérations algébriques à effectuer.

Voici un exemple. Supposons que $m = 6$ et que la clef soit la permutation suivante :

$$\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 3 & 5 & 1 & 6 & 4 & 2 \end{array} \tag{3}$$

1. Quelle est la permutation inverse π^{-1} ? Il s'agit de

$$\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 3 & 6 & 1 & 5 & 2 & 4 \end{array} \tag{4}$$

2. Déchiffrer le message "EESLSHSALSESLSHBLEHSYEETHRAEOS". Le texte clair est "she sells sea shells by the sea shore" ("elle vend des coquillages sur la plage").

Le chiffrement par permutation est un cas particulier du chiffrement de Hill. En effet à toute permutation π de l'ensemble $\{1, \cdots, m\}$, on peut associer une matrice $P_\pi = (p_{i,j})$ carrée de taille $m \times m$ par la formule

$$p_{i,j} = \begin{cases} 1 & \text{si } i = \pi(j) \\ 0 & \text{sinon.} \end{cases}$$

Par ailleurs, on appelle matrice de permutation toute matrice carrée dont les coefficients sont 0 ou 1, et telle que dans chaque ligne et dans chaque colonne il y a exactement un coefficient égal à 1.

Il est facile de voir que pour toute permutation π de $\{1, \cdots, m\}$, la matrice P_π est une matrice de permutation et qu'inversement pour toute matrice de permutation P de taille $m \times m$, il existe une permutation π_P de $\{1, \cdots, m\}$ telle que $P_{\pi_P} = P$.

Le chiffrement de Hill avec une matrice de permutation P_π est équivalent au chiffrement par permutation avec comme clef π. De plus $P_\pi^{-1} = P_{\pi^{-1}}$ de sorte que le déchiffrement de Hill est équivalent au déchiffrement par permutation.

Exercice

Soit la matrice suivante : $P = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$

1. La matrice P est-elle une matrice de permutation ? Et si oui, de quelle permutation ? (C'est-à-dire quelle permutation π vérifie $P_\pi = P$?)
2. Calculer P^{-1}.
3. Calculer l'inverse de π.

Chiffrement en chaîne

Dans tous les procédés de chiffrement précédemment rencontrés, les éléments de texte clair sont chiffrés de la même manière à partir de la clef K. En fait, la chaîne du texte clair $\mathbf{y}$ est obtenue ainsi

$$\mathbf{y} = y_1 y_2 \cdots = E(x_1, K)E(x_2, K)\cdots$$

Les procédés de ce type sont appelés chiffrement par bloc.

Une autre approche consiste à utiliser la notion de chiffrement en chaîne. L'idée de base consiste à engendrer une suite de clefs $\mathbf{z} = z_1 z_2 \cdots$, et à l'utiliser pour chiffrer la chaîne $\mathbf{x} = x_1 x_2 \cdots$ suivant la règle

$$\mathbf{y} = y_1 y_2 \cdots = E(x_1, z_1)E(x_2, z_2)\cdots$$

Un chiffrement en chaîne fonctionne ainsi. Supposons que l'on ait une clef $K \in \mathcal{K}$ et un texte clair $x_1 x_2 \cdots$ On utilise une fonction f_i pour créer z_i (le $i$ème élément de la suite de clefs), à l'aide de K et des $i-1$ premiers caractères du texte clair :

$$z_i = f_i(K, x_1, \cdots, x_{i-1}) \ .$$

On remarque que $z_1 = f_1(K)$. L'élément z_i est utilisé pour chiffrer x_i par $y_i = E(x_i, z_i)$. Donc pour chiffrer le texte clair $x_1 x_2 \cdots$, on calcule successivement

$$z_1, y_1, z_2, y_2, \cdots$$

Le déchiffrement du texte chiffré s'effectue en calculant successivement

$$z_1, x_1, z_2, x_2, \cdots$$

Définition formelle

Définition

Un chiffrement en chaîne est la donnée d'un 7-uplet $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{L}, \mathcal{F}, E, D)$ où

- $\mathcal{P}$ est un ensemble fini de textes clairs possibles.
- $\mathcal{C}$ est un ensemble fini de textes chiffrés possibles.
- $\mathcal{K}$ est un ensemble fini de clefs principales possibles.
- $\mathcal{L}$ est un ensemble fini de sous-clefs possibles.
- $\mathcal{F} = (f_1, f_2, \cdots)$ est le générateur de sous-clefs, où pour tout $i \geq 1$, $f_i : \mathcal{K} \times \mathcal{P}^{i-1} \to \mathcal{L}$ (en particulier, $f_1 : \mathcal{K} \to \mathcal{L}$).
- E est une fonction de chiffrement, $E : \mathcal{P} \times \mathcal{L} \to \mathcal{C}$.
- D est une fonction de déchiffrement $D : \mathcal{C} \times \mathcal{L} \to \mathcal{P}$,
- Pour chaque sous-clef z, il existe au moins une sous-clef z' telle que pour tout $x \in \mathcal{P}$,

$$D(E(x, z), z') = x \ .$$

Un chiffrement par bloc est un cas particulier de chiffrement en chaîne. **Pourquoi ?** La suite des sous-clefs est constante : $z_i = K$ pour tout i (autrement dit, $\mathcal{L} = \mathcal{K}$ et $f_i : \mathcal{K} \times \mathcal{P}^{i-1} \to \mathcal{L}$ est donnée par $f_i(K, x_1, \cdots, x_{i-1}) = K$ pour chaque $i \geq 1$).

Un chiffrement en chaîne est dit synchrone si sa suite de sous-clefs est indépendante du texte clair, c'est-à-dire si elle est fonction de la clef K uniquement. Autrement dit, pour tout $i \geq 1$, $f_i(K, x_1, \cdots, x_{i-1}) = g_i(K)$ où $g_i : \mathcal{K} \to \mathcal{L}$.

Un chiffrement en chaîne est dit périodique de période d si $z_{i+d} = z_i$ pour tout $i \geq 1$. Le chiffrement de Vigenère avec un mot-clef de longueur m peut être vu comme un chiffrement en chaîne périodique de période m. **Pourquoi ?** En effet, soit la clef $K = (k_1, \cdots, k_m)$. Tout entier $n \geq 1$ s'écrit sous la forme $n = i + dm$, $1 \leq i \leq m$. (Cette écriture est unique. Supposons en effet que $i + dm = n = j + em$, $1 \leq i \leq m$, $1 \leq j \leq m$. Si $i = j$, alors il en résulte que $d = e$. Supposons donc que $i > j$ par exemple. On a $i - j = (e - d)m$. Mais par hypothèse, $0 < i - j < i < m$, donc l'égalité précédente est impossible.) On définit maintenant $f_{i+dm}(K, x_1, \cdots, x_{i-1}, \cdots, x_{i+dm-1}) = x_i$ pour tout $1 \leq i \leq m$ et $d \geq 0$.

Les chiffrements en chaîne sont souvent décrits sur un alphabet binaire, c'est-à-dire $\mathcal{P} = \mathcal{C} = \mathbb{Z}/2\mathbb{Z} = \{0, 1\}$.

Supposons que l'on se donne $k_1, \cdots, k_m \in \mathbb{Z}/2\mathbb{Z}$. On pose $z_i = k_i$ pour chaque $i = 1, \cdots, m$. On engendre les autres sous-clefs par une relation de récurrence linéaire de degré m

$$z_{i+m} = \sum_{j=0}^{m-1} c_j z_{i+j} \mod 2$$

pour $i \geq 1$, et où $c_0, \cdots, c_{m-1} \in \mathbb{Z}/2\mathbb{Z}$ sont des constantes fixées une fois pour toute.

Il est possible de démontrer que si $c_0, \cdots, c_{m-1}$ sont judicieusement choisis, pour toute valeur de $(k_1, \cdots, k_m)$, la suite de sous-clefs engendrées a pour (plus petite) période $2^m - 1$. Donc une "petite" clef peut engendrer une suite de très longue période.

Exercice

Donner les 20 premières clefs engendrées par la relation de récurrence linéaire d'ordre 4 suivante

$$z_{i+4} = z_i + z_{i+1} \mod 2,\ i \geq 1$$

avec $z_1 = 1$, $z_2 = z_3 = z_4 = 0$. Quelle est la période ?

Solution : $z_5 = z_1 + z_2 = 1$, $z_6 = z_2 + z_3 = 0$, $z_7 = z_3 + z_4 = 0$, $z_8 = z_4 + z_5 = 1$, $z_9 = z_5 + z_6 = 1$, $z_{10} = z_6 + z_7 = 0$, $z_{11} = z_7 + z_8 = 1$, $z_{12} = z_8 + z_9 = 0$, $z_{13} = z_9 + z_{10} = 1$, $z_{14} = z_{10} + z_{11} = 1$, $z_{15} = z_{11} + z_{12} = 1$, $z_{16} = z_{12} + z_{13} = 1$, $z_{17} = z_{13} + z_{14} = 0$, $z_{18} = z_{14} + z_{15} = 0$, $z_{19} = z_{15} + z_{16} = 0$, $z_{20} = z_{16} + z_{17} = 1$, ... La période est $15 = 2^4 - 1$.

Registre à décalage

Un intérêt de cette méthode est que la suite des clefs peut être produite par un registre à décalage linéaire ou LFSR (pour "Linear Feedback Shift Register") de façon efficace. On utilise pour cela un registre à décalage à m étages. Le vecteur $(k_1, \cdots, k_m)$ est utilisé pour initialiser le registre à décalage. À chaque unité de temps, les opérations suivantes sont effectuées simultanément :

1. k_1 est copié sur le terme suivant de la séquence de clefs.
2. $k_2, \cdots, k_m$ sont décalés d'un rang ("étage") vers la gauche.
3. La nouvelle valeur de k_m est calculée par $\sum_{j=0}^{m-1} c_j k_{j+1}$.

On note que le calcul du nouveau k_m est effectué en prenant certaines valeurs du registre à certains étages (ceux pour lesquels $c_j = 1$) et en calculant leur somme modulo 2 (qui est le ou-exclusif).

Exemple/exercice

Soit le chiffrement par chaîne asynchrone suivant : $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathcal{L} = \mathbb{Z}/26\mathbb{Z}$. Soit $z_1 = K$ et $z_i = x_{i-1}$ $(i \geq 2)$. Pour $0 \leq z \leq 25$, on définit

$$e_z(x) = x + z \mod 26$$

et

$$d_z(y) = y - z \mod 26$$

pour $x, y \in \mathbb{Z}/26\mathbb{Z}$. Supposons que $K = 8$. Chiffrer le texte rendezvous.

Solution : On convertit tout d'abord ce texte en une suite d'entiers : $17, 4, 13, 3, 4, 25, 21, 14, 20, 18$. On calcule la suite des clefs : $8, 17, 4, 13, 3, 4, 25, 21, 14, 20$. En ajoutant les éléments correspondants modulo 26 on obtient $25, 21, 17, 16, 7, 3, 20, 9, 8, 12$. En caractères alphabétiques le texte chiffré est ZVRQHDUJIM.

Cryptanalyse : principe de Kerckhoff

On présente maintenant quelques techniques de cryptanalyse. L'hypothèse généralement faite est que l'opposant Oscar connaît le système de chiffrement utilisé. Il s'agit du principe de Kerckhoff. Bien sûr, si Oscar ne connaît pas le procédé employé, sa tâche sera plus compliquée, mais on ne souhaite pas baser la sécurité du système sur la protection de la description des fonctions cryptographiques. Le but est donc d'étudier les procédés de chiffrement suivant le principe de Kerckhoff.

Niveaux d'attaques

Tout d'abord, différencions les niveaux d'attaques possibles. Les modèles les plus courants sont énumérés ici.

1. Texte chiffré connu : L'opposant ne connaît que la chaîne du message chiffré $\mathbf{y}$.
2. Texte clair connu : L'opposant dispose d'un texte clair $\mathbf{x}$ et du texte chiffré $\mathbf{y}$ correspondant.
3. Texte clair choisi : L'opposant a accès à une machine chiffrante. Ainsi il peut choisir un texte clair $\mathbf{x}$ et obtenir son texte chiffré $\mathbf{y}$.
4. Texte chiffré choisi : L'opposant a temporairement accès à une machine déchiffrante. Ainsi il peut choisir un texte chiffré $\mathbf{y}$ et obtenir son texte clair $\mathbf{x}$ correspondant.

Dans tous les cas, le but de l'opposant est de déterminer quelle clef est utilisée.

Ces quatre niveaux sont clairement énumérés par ordre croissant de pouvoir dont dispose l'opposant.

Attaque à texte clair connu du chiffrement de Hill

Le chiffrement de Hill est plus difficile à casser par une attaque à texte chiffré connu, mais succombe facilement à une attaque à texte clair choisi. Supposons tout d'abord que l'opposant ait trouvé la valeur m utilisée. Supposons ensuite qu'il dispose d'au moins m paires de tuples $x_j = (x_{1,j}, x_{2,j}, \cdots, x_{m,j})$ et $y_j = (y_{1,j}, y_{2,j}, \cdots, y_{m,j})$ $(1 \leq j \leq m)$ telles que $y_j = e_K(x_j)$, $1 \leq j \leq m$. Si on définit deux matrices $m \times m$ $X = (x_{i,j})$, $Y = (y_{i,j})$, alors on obtient l'équation matricielle $Y = XK$, où K est la matrice (inconnue) $m \times m$ définissant la clef. Si X est inversible, alors Oscar peut calculer X^{-1} et obtenir $K = X^{-1}Y$. Si X n'est pas inversible, alors il faut un autre ensemble de m paires de vecteurs.

Exercice : Supposons que le texte clair "friday" soit chiffré en utilisant le chiffrement de Hill avec $m = 2$, en un texte chiffré "PQCFKU". Appliquer l'attaque précédente pour retrouver la clef K utilisée.

Solution : On a $E((5,17),K) = (15,16)$, $E((8,3),K) = (2,5)$ et $E((0,24),K) = (10,20)$. À partir des deux premières paires de textes clairs et de textes chiffrés, on obtient l'équation matricielle

$$\begin{pmatrix} 15 & 16 \\ 2 & 5 \end{pmatrix} = \begin{pmatrix} 5 & 17 \\ 8 & 3 \end{pmatrix} K \ . \tag{5}$$

On calcule facilement

$$\begin{pmatrix} 5 & 17 \\ 8 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} 9 & 1 \\ 2 & 15 \end{pmatrix} \tag{6}$$

donc

$$K = \begin{pmatrix} 9 & 1 \\ 2 & 15 \end{pmatrix} \begin{pmatrix} 15 & 16 \\ 2 & 5 \end{pmatrix} = \begin{pmatrix} 7 & 19 \\ 8 & 3 \end{pmatrix} . \tag{7}$$

Cryptanalyse d'un chiffrement en chaîne basé sur un LFSR

Le texte chiffré est la somme modulo 2 du texte clair et de la clef soit $y_i = x_i + z_i \mod 2$. La suite des clefs est prdouite à partir de $z_1, \cdots, z_m$ en utilisant une relation de récurrence linéaire

$$z_{m+i} = \sum_{j=0}^{m-1} c_j z_{i+j} \mod 2$$

où $c_0, \cdots, c_{m-1} \in \mathbb{Z}/2\mathbb{Z}$ (et on peut supposer en outre que $c_0 = 1$).

Comme toutes les opérations sont linéaires, il semble que le procédé soit aussi vulnérable que le chiffrement de Hill face à une attaque à texte clair connu. Supposons qu'Oscar dispose d'un texte clair $x_1 x_2 \cdots x_n$ et de son texte chiffré $y_1 y_2 \cdots y_n$. Il peut calculer la suite de clefs $z_i = x_i + y_i \mod 2$, $1 \leq i \leq n$. En supposant qu'il connaisse également m, il a seulement besoin de calculer $c_0, \cdots, c_{m-1}$ pour reconstituer toute la clef. Donc il a seulement besoin de retrouver m valeurs inconnues.

Pour tout $i \geq 1$, $z_{m+i} = \sum_{j=0}^{m-1} c_j z_{i+j} \mod 2$ est une équation linéaire avec m inconnues. Si $n \geq 2m$, on peut sélectionner m équations et les résoudre.

Le système des m équations linéaires peut sécrire sous une forme matricielle

$$(z_{m+1}, z_{m+2}, \cdots, z_{2m}) = (c_0, \cdots, c_{m-1}) \begin{pmatrix} z_1 & z_2 & \cdots & z_m \\ z_2 & z_3 & \cdots & z_{m+1} \\ \vdots & \vdots & \cdots & \vdots \\ z_m & z_{m+1} & \cdots & z_{2m-1} \end{pmatrix}. \quad (8)$$

Si la matrice est inversible modulo 2, alors on obtient la solution

$$(c_0, c_1, \cdots, c_{m-1}) = (z_{m+1}, z_{m+2}, \cdots, z_{2m}) \begin{pmatrix} z_1 & z_2 & \cdots & z_m \\ z_2 & z_3 & \cdots & z_{m+1} \\ \vdots & \vdots & \cdots & \vdots \\ z_m & z_{m+1} & \cdots & z_{2m-1} \end{pmatrix}^{-1}. \quad (9)$$

Exercice : Supposons qu'Oscar obtienne le texte chiffré 101101011110010 correspondant au texte clair 011001111111000.

1. Calculer la suite des clefs. (**Solution:** 11010010000101010.)
2. Supposons par ailleurs qu'Oscar sait que cette suite a été produite par un LFSR à 5 étages. Retrouver les coefficients c_0, c_1, c_2, c_3, c_4 qui permettent de produire cette suite. Vous utiliserez le fait que

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}. \tag{10}$$

Solution : On résoud l'équation matricielle à partir des 10 premiers bits de clefs.

$$(0,1,0,0,0) = (c_0, c_1, c_2, c_3, c_4) \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{11}$$

Cela donne

$$(c_0, c_1, c_2, c_3, c_4) = (0, 1, 0, 0, 0) \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix} = (1, 0, 0, 1, 0) \ . \quad (12)$$

Donc la relation de récurrence utilisée par le générateur est
$z_{i+5} = z_i + z_{i+3} \mod 2$.