MODÉLISATION MATHÉMATIQUE

Séances 15-16

THÈME : Mathématiques et cryptographie

1er juin 2010

Table des matières

1 Introduction

Cette séance sera consacrée aux différentes techniques cryptage et leur analyse mathématique. La **cryptographie** est une science qui étudie les méthodes permettant de transmettre des messages de façon confidentielle.

On va appeler **message clair** un ensemble de données (texte, image,...) que l'on souhaite transmettre.

Le **chiffrement** est un procédé permettant de transformer le message clair de telle sorte qu'il soit incompréhensible par qui que ce soit d'autre que l'auteur du message et le destinataire.

Le **chiffré** est le résultat du chiffrement.

Le **déchiffrement** est le procédé permettant de retrouver le message clair à partir du chiffré.

La **clé** de chiffrement est un paramètre qui est utilisé dans le procédé de chiffrement ou déchiffrement.

Un **cryptosystème** est un ensemble de clés $\mathcal{K}$, de messages clairs possibles , $\mathcal{M}$, de messages chiffrés possibles $\mathcal{C}$ associée à un algorithme de chiffrage, représenté par une fonction $E : \mathcal{K} \times \mathcal{M} \to \mathcal{C}$, et un procédé de déchiffrage , représenté par une fonction $D : \mathcal{K} \times \mathcal{C} \to \mathcal{M}$. On suppose que pour tout $m \in \mathcal{M}$ il existe une paire de clés de chiffrement et de déchiffrement telles que la relation

$$D(k_D, E(k_E, m)) = m$$

est assurée. On note $(\mathcal{K}, \mathcal{M}, \mathcal{C}, E, D)$ un cryptosystème.

***Remarque* 1.1** (Principe de Kerckhoffs)**.** De nos jours, la plus grande partie de méthodes de cryptographie reposent sur ce principe, énoncé par Auguste Kerckhoffs à la fin XIXe siècle (publié dans un article au Journal des sciences militaires, vol. IX, pp. 5-38, Janvier 1883, pp. 161-191, Février 1883. Source : Wikipédia).

La sécurité d'un cryptosystème ne doit pas reposer sur la non divulgation de la fonction de cryptage mais uniquement sur la non divulgation de la clé.

Un autre énoncé de ce principe, connu sous le nom de maxime de Shannon a été proposé par Claude Shannon, au milieu du XXe sciècle : **L'adversaire connaît le système.**

Deux modèles principaux de chiffrement existent actuellement :

Chiffrement symétrique On appelle aussi ce modèle "chiffrement à clé secrète". C'est l'approche classique. La même clé est utilisée pour chiffrer et déchiffrer un message. Cette clé doit alors rester secrète. Dans les systèmes de communication à grande échelle cela représente la principale faiblesse de ces chiffrements. En effet la clé doit être communiquée au destinataire et dans de nombreux cas elle est envoyée par le même canal de transmission que le message lui même. Si elle est interceptée à ce moment là, toute la communication est découverte !

Chiffrement assymétrique. On l'appelle aussi chiffrement à clé publique. La clé utilisée pour le chiffrement est publique et elle est différente de la clé de déchiffrement, qui est secrète. Ainsi le destinataire du message choisit la clé de déchiffrement et la clé de chiffrement associée. Il peut envoyer à l'émetteur la clé de chiffrement (publique). Et la clé de déchiffrement, secrète, ne circule pas dans les canaux de transmission.

2 Substitutions mono-alphabétiques

Un des systèmes de cryptage les plus anciens, le chiffre de César, appartient à la famille des chiffres de substitution. Le principe est très simple. Il suffit de mélanger les lettres de l'alphabet et remplacer chaque lettre du texte à crypter par la lettre qui lui correspond dans l'alphabet mélangé.

Ainsi la clé de ce chiffrement est la permutation des lettres de l'alphabet. Celui qui connaît l'ordre permuté peut retrouver le message clair à partir du chiffré. Le chiffre de César utilisait une permutation très simple : un décalage (cyclique) de 3 positions. On peut le généraliser en décalant l'alphabet de k positions.

Problème 1. Chiffrement

1. Si on admet que les permutations peuvent être quelconques, combien de clés potentielles y a-t-il ?
2. Si l'on sait que la permutation utilisée est un simple décalage, combien de clés différentes il y a ?
3. Ecrire un programme Scilab qui réalise un chiffrement de César. On supposera que l'alphabet utilisé est composé de 26 lettres majuscules, sans accents. le programme prendra en paramètre la clé, k et le message à crypter sous forme de chaîne de caractères.
4. Ecrire un programme scilab réalise un chiffrement avec une permutation quelconque. Le programme prendra en paramètre le tableau définissant la permutation et le message à crypter

Problème 2. Déchiffrement

1. Dans un chiffrement de César quelle est la transformation inverse ?

2. Ecrire un programme qui déchiffre un message chiffré avec un code de César. Le programme prendra en paramètre la clé et le message crypté.
3. Si la permutation est quelconque, comment déchiffrer quand on connaît la table de permutation ?
4. Ecrire un programme scilab réalise un décodage avec une permutation quelconque. Le programme prendra en paramètre le tableau définissant la permutation et le message crypté.

Problème 3. Piratage

Les méthodes basées sur les substitutions ont une faiblesse majeure. Elle a été découverte par un mathématicien arabe AL-Kindi (IXème siècle). Il a rédigé le premier traité connu sur l'analyse fréquentielle comme technique de cryptanalyse. Il a montré comment retrouver le message en analysant les fréquences d'occurrence de caractères dans une langue donnée. En effet, dans chaque langue les différents caractères ont des fréquences d'utilisation différentes. par exemple, en français et en anglais, la lettre la plus fréquente est "e". Lorsque l'on sait en quelle langue est écrit le message, l'analyse fréquentielle consiste à identifier les lettres par leurs fréquences d'occurrence. On arrive ainsi à déterminer le décalage ou même la table de permutation d'un chiffrement par substitution.

1. Ecrire une fonction qui calcule la table de fréquences des caractères présents dans une chaîne de caractères
2. Utiliser ce programme pour décrypter un chiffré utilisant le chiffrement de César.
3. Tentez de décrypter ceci par analyse fréquentielle :

   JLI CR KIZSLEV SLJKVIZVLIV U LE SLJ HLZ KIREJYRSLKRZK MVIJ LE SLK GVL SLTFCZHLV UVJ SLIVRLTIRKVJ RSLKZJ LE SLICVJHLV WLERDSLCV R CR SLTTLCV CFZE UL SLJKV VK RL XZSLJ JREJ SLIRE WZK SILJHLVDVEK UL XIRSLXV TFEKIV LE SLIXIRMV HLZ CV SFLJTLCRZK SLKFI P R UV C'RSLJ J RKKIZSLREK LE KRSLIVK ZC J P TLCSLKR KVC LE FLIJ UREJ LEV TRDSLJV SLCKVIZVLIVDVEK VE LE TFETZCZRSLCV ZC SLKZERZK TVKKV JKZSLCRKZFE SLJV TV XCFSLCVLO SLKFE SLTYV DRC KFE SLIEFLJ

4. Si la permutation est quelconque, comment déchiffrer quand on connaît la table de permutation ?
5. Ecrire un programme scilab réalise un décodage avec une permutation quelconque. Le programme prendra en paramètre le tableau définissant la permutation et le message crypté.

3 Chiffres algébriques

Nous allons étudier également une autre technique de chiffrement, proposée par un mathématicien britanique, Hill, en 1929. Sa méthode repose sur un codage no pas lettre par lettre mais par blocs de m lettres. Cela rend les attaques plus difficiles. Le chiffre de Hill utilise l'algèbre linéaire et l'arithmétique.

On remplace les lettres par des nombres selon leur place dans l'alphabet : A sera représenté par 0 et Z par 25. On décompose le message à chiffrer en blocs de m lettres. Soit un tel bloc, $X = (x_1, \ldots, x_m)$. On obtient le bloc chiffré, $Y = (y_1, \ldots, y_m)$ en faisant des combinaisons linéaires des nombres du bloc X :

$$y_i = a_{i,1}x_1 + \cdots + a_{i,m}x_m, \;\; i = 1, \ldots, m$$

Fréquences d'apparition des lettres			
Lettre	Fréquence	Lettre	Fréquence
A	8.40%	N	7.13%
B	1.06%	O	5.26%
C	3.03%	P	3.01%
D	4.18%	Q	0.99%
E	17.26%	R	6.55%
F	1.12%	S	8.08%
G	1.27%	T	7.07%
H	0.92%	U	5.74%
I	7.34%	V	1.32%
J	0.31%	W	0.04%
K	0.05%	X	0.45%
L	6.01%	Y	0.30%
M	2.96%	Z	0.12%

FIGURE 1 – fréquences des caractères en français

Pour que les nombres y_i correspondent aux lettres on effectue les opérations modulo 26 :

$$y_i = (a_{i,1}x_1 + \cdots + a_{i,m}x_m) \bmod 26, \quad i = 1, \ldots, m$$

Ce sont les mêmes coefficients $\{a_{i,j}\}_{i,j=1}^{m}$ qui sont utilisés pour chiffrer tous les blocs. Ils sont choisis également parmi les nombres entiers $\{0, 1, \ldots, 25\}$.

Pour simplifier, nous allons dans la suite considérer en détails le cas $m = 2$. L'opération de chiffrement consiste alors à choisir une matrice

$$C = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

d'entiers entre 0 et 25 et à calculer chaque bloc $Y = (y_1, y_2)$ du chiffré par multiplication

$$Y = CX \bmod 26$$

Toutes les considérations se font dans $\mathbb{Z}|26\mathbb{Z}$.

Problème 4. Chiffrement

1. Qu'est ce qui représente la clé de chiffrement ?
2. Soit
$$C = \begin{pmatrix} 1 & 2 \\ 3 & 7 \end{pmatrix}$$
Chiffrer le message
"bla bli blu"
3. Ecrire un programme qui chiffre un message par codage de Hill par blocs de 2 caractères. Il prend en entrée la clé et le message à chiffrer.

Problème 5. Déchiffrement

1. Quelles doivent être les propriétés de la matrice de chiffrement pour que le chiffre soit inversible ? Donner une expression algébrique de ces conditions.
2. Ecrire algébriquement l'opération de déchiffrement, inverse du chiffrement
3. Ecrire un programme scilab qui réalise un décodage Le programme prendra en paramètre la clé de chiffrement et le message crypté.

Problème 6. Piratage

On va étudier ici l'attaque à texte clair connu. Supposons que l'on dispose d'un morceau de chiffré et d'une partie de texte clair qui lui correspond. Supposons également que l'on connaît la longueur des blocs utilisés, ici $m = 2$. On cherche alors à déduire la matrice de chiffrement C.

Prenons l'exemple suivant. le texte claire est
"MON TP DE MATHS"
Le chiffré est
"AOGZSVQBTMZR"

$$C = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

1. Prenons le premier couple de caractères du chiffré AO. Traduire cela en couple d'entiers et poser les équations de chiffrage avec la matrice inconnue C. Faire la même chose avec le second couple. Ecrire les équations sous forme matricielle
$$A = CB$$
en précisant les matrices A et B.
2. Peut on en déduire C ? Expliquer pourquoi.
3. Continuer la procédure jusqu'à ce que la matrice C puisse être calculée.