CH.3 CRYPTOSYSTÈMES

- 3.1 L'environnement des cryptosystèmes
- 3.2 Les cryptosystèmes historiques
- 3.3 Le cryptosystème parfait
- 3.4 Les registres à décalage

3.1 L'environnement des cryptosystèmes

Une excellente source bibliographique se trouve sur le site de Didier Müller (Lycée cantonal de Porrentruy) :

http://www.apprendre-en-ligne.net/crypto/index.html

Ce site contient une introduction historique et des développements sur l'état des techniques cryptographiques, illustrés de nombreuses applets java et de sources mathematica.

Cryptologie : science des messages secrets. Contient la cryptographie, art de rendre inintelligible un message, et la cryptanalyse, art de trouver le message clair caché.

La cryptographie utilise un chiffre pour coder un message. Le déchiffrement est l'opération inverse, par une personne autorisée à retrouver le message clair.

La cryptanalyse est l'ensemble des techniques permettant à une personne non autorisée de trouver le contenu d'un message.

L'histoire de la cryptologie est très ancienne. En fait, l'écriture même est une façon de coder des données qui est inintelligible aux illettrés...

On trouve des exemples de chiffres chez les Hébreux et les Grecs. Un code célèbre était utilisé par Jules César.

Al-Kindi, au 9e siècle, rédige le premier manuel de décryptage contenant la technique d'analyse des fréquences.

Au 15e siècle, Alberti propose un dispositif de codage polyalphabétique permettant d'éviter l'analyse des fréquences.

Au 16e siècle, on trouve les noms de Trithème, Cardan, Della Porta et surtout Vigenère, qui perfectionne la technique de Trithème.

On peut aussi mentionner le Grand Chiffre de Louis XIV, dû à Rossignol, qui ne fut décrypté qu'à la fin du 19e siècle par Bazeries.

Aux Temps Modernes, Jefferson invente un dispositif mécanique de chiffrement, réinventé par Bazeries. De l'après-Première Guerre mondiale date l'invention d'un autre dispositif mécanique célèbre, la machine Enigma, qui sera perfectionnée et adoptée pendant la 2e Guerre mondiale par les Allemands. Elles sera cassée par le Polonais Marian Rejewski, mais le décryptage nécessitait néanmoins des efforts de calculs

considérables, où s'illustrent Alan Turing et l'équipe du centre de Blechley Park, notamment par la conception d'ordinateurs dédiés à ce travail.

A l'époque contemporaine, la cryptologie devient de plus en plus dépendante des progrès de l'algorithmique et des performances des ordinateurs. Diffie et Hellman introduisent le concept de clé publique en 1976. Parallèlement, le système de cryptage DES est développé et est largement utilisé. Rivest, Shamir et Adleman en 1977 inventent le système RSA, le plus utilisé actuellement des systèmes à clé publique. Plus récemment, des systèmes ont été imaginés utilisant les ressources de la physique quantique...

Les systèmes cryptographiques vont évidemment de pair avec les "ennemis" qui essaient de décrypter les messages chiffrés.

On suppose toujours que les ennemis connaissent les algorithmes de chiffrement ; il est illusoire de penser conserver cet algorithme secret.

Par contre, on doit penser à trois niveaux possibles de sécurité :

- l'ennemi dispose de messages chiffrés en quantité appréciable ;
- l'ennemi dispose en plus d'échantillons de texte en clair avec leur version chiffrée ;
- l'ennemi dispose de la faculté d'obtenir la version chiffrée de textes qu'il a fournis lui-même.

Le troisième niveau correspond à la sécurité maximum possible.

3.2 Les cryptosystèmes historiques

Jusqu'aux systèmes contemporains, tous les systèmes cryptographiques étaient basés sur des techniques de substitutions et de décalages.

Le chiffre le plus simple consiste à remplacer chaque lettre du texte clair par un autre symbole (substitution simple). Ce symbole peut être une lettre du même alphabet, d'un autre alphabet ou un dessin... La fonction de substitution (table) représente la clé de chiffrement et sert aussi au déchiffrement.

Ce chiffre est facile à décrypter par analyse des fréquences des lettres. Par ailleurs, la connaissance d'un fragment de texte clair avec son chiffrement suffit à le casser. Enfin, la connaissance de tout ou partie de la clé suffit aussi.

Un perfectionnement consiste à chiffrer des digrammes, trigrammes, voire des mots entiers. Cela rend la clé encore plus vulnérable.

Pour mémoire, indiquons que diverses permutations de groupes de lettres peuvent être utilisées comme surchiffrement.

Le chiffre de César est un système de substitution dont la clé est une simple lettre. Son rang moins 1 dans l'alphabet indique la quantité dont on décale circulairement les lettres du texte en clair. Si la clé est E, cinquième lettre, le A est codé E, le B est codé F, ... , le Z est codé D.

Ce chiffre ne résiste pas à une analyse des fréquences.

Le chiffre de Vigenère perfectionne le chiffre de César en ce qu'il est polyalphabétique. Le décalage des lettres dans l'alphabet n'est pas constant, mais déterminé par la clé qui est un mot.

Par exemple, si la clé est `BACHELIER`, le chiffrement se passe ainsi :

texte clair : `CODEPOLYALPHABETIQUEDEVIGENERE`
clé : `BACHELIERBACHELIERBACHELIERBAC`
texte crypté : `DOFLTZTCRMPJHFPBMHVEFLZTOIEFRG`

Ce chiffre est très sûr et a été longtemps réputé indécryptable.
Un décrypteur qui connaît la longueur de la clé peut effectuer des analyses de fréquence. Avec les moyens informatiques, si on sait dans quelle langue est le message et si on a un texte crypté assez long, on peut essayer diverses longueurs de clé et garder celle qui donne une analyse des fréquences qui se rapproche de la norme ("Indices de coïncidence")...
Ce code peut en fait être cassé facilement. Babbage utilise des répétitions de suites de caractères pour faire des hypothèses sur la longueur de la clé. Bazeries attaque le chiffre en faisant une hypothèse sur la présence d'un mot connu dans le message en clair.

Le perfectionnement ultime du code de Vigenère est d'utiliser une clé aussi longue que le texte à coder et de ne jamais utiliser la même clé. On utilise alors un *masque jetable*.
On peut par exemple utiliser un texte de référence comme clé (la Bible...)
Le décryptage de Babbage devient inopérant, mais pas nécessairement celui de Bazeries...

Le problème de la communication de la clé peut être résolu par une double communication :

A fait à B une demande ce clé. Il envoie un texte aléatoire t sans signification. B choisit une clé c. Il renvoie $u = t + c$ à A. Le texte u est indistingable d'un texte aléatoire. En faisant la différence $u – t$, le premier interlocuteur A calcule c, qu'il peut alors utiliser pour chiffrer son message.
Quel est le point faible de ce protocole ?

Les chiffres de type Vigenère sont encore utilisés à cause de la grande facilité de chiffrement et de déchiffrement à la volée. Le point faible est celui de la transmission des clés.
Pour la transmission de celles-ci, on peut alors utiliser une communication faisant appel à un chiffrement plus sûr mais plus gourmand en temps, de type RSA, pour des clés de taille plus petite que les messages à chiffrer. Ceci permet d'utiliser des clés aléatoires.

Une variante est le procédé autoclave, qui utilise le texte lui-même comme clé de chiffrement après un initialisation avec une clé.

texte clair :	`EXEMPLEDUPROCEDEAUTOCLAVE`
clé :	`CLEEXEMPLEDUPROCEDEAUTOCL`
texte crypté :	`GIIQMPQSFTUIRVRGEXXOWEOXP`

3.3 Le cryptosystème parfait

Un cryptosystème parfait est tel que, même si le décrypteur connaît l'algorithme de cryptage et une quantité arbitraire d'exemples de texte clair et du même texte crypté, il est incapable de décrypter un échantillon crypté.

On montre que c'est possible. Il existe donc un chiffre indécryptable. Bien plus, on sait le décrire...

Il s'agit du masque jetable aléatoire. On a déjà vu la technique du masque jetable. Mais le masque jetable classique est vulnérable à l'attaque si un décrypteur dispose d'échantillons de texte clair et crypté, en reconnaissant le texte utilisé comme clé.
La solution consiste donc à utiliser un masque jetable aléatoire, c'est-à-dire que la clé est une suite aléatoire de symboles équiprobables.

On peut montrer que ce code est effectivement indécryptable.

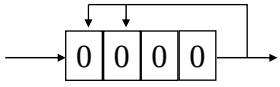
Les problèmes sont les suivants :
- on utilise en général des suites pseudo-aléatoire ;
- il faut transmettre la clé.

Mais cette technique est effectivement utilisée dans certains cas, lorsqu'on peut transmettre une clé par un moyen absolument sûr. Cette clé est alors gardée en réserve jusqu'au jour où on doit effectivement transmettre un message. Dans ce cas, on a en général le moyen de fabriquer des suites "vraiment" aléatoires.

A noter que les casinos publient les listes des numéros sortis sur leurs roulettes, de façon que chacun puisse vérifier leur régularité. Ces listes constituent d'excellentes suites "aléatoires"...

3.4 Les registres à décalage

On peut utiliser des registres à décalage pour produire des suites pseudo-aléatoires de bits. On va illustrer cet exemple par le registre suivant (correspondant au polynôme primitif $x^4 + x + 1$).



En partant du contenu 1000 pour ce registre, on trouve successivement
1000 0100 0010 0001 1100 0110 0011 1101
1010 0101 1110 0111 1111 1011 1001 1000 ... et la suite est périodique de période 15, période maximale avec un registre de taille 4.

On peut considérer la suite des bits de poids le plus fort :
0 0 0 1 0 0 1 1 0 1 0 1 1 1 1 0 0 0 1 0 0 1 1 0 1 0 1 1 1 1 0 ...

Elle possède des propriétés a priori intéressantes pour une suite pseudo-aléatoire.

Parmi toutes les suites engnedrées par un registre à décalage, elle a la période la plus longue. En plus, on peut montrer facilement que toute suite consécutive de 15 bits contient 8 fois le bit 1 et 7 fois le bit 0. (En fait, les registres utilisés ont comme longueur quelques centaines.)

Néanmoins, si on appelle a_n, b_n, c_n, et d_n les bits contenus dans le registre au temps n, on constate qu'ils sont donnés par les récurrences :

$$\begin{cases} d_n = c_{n-1} \\ c_n = b_{n-1} \\ b_n = a_{n-1} + d_{n-1} \\ a_n = d_{n-1} \end{cases}$$

En éliminant, on trouve une récurrence d'ordre 4 portant sur d :

$$d_n + d_{n-3} + d_{n-4} = 0 \pmod 2$$

On retrouve en fait les coefficients du polynôme de départ.

En conséquence, si on connaît n et un échantillon de texte clair et son chiffrement de longueur $2n$, on en déduit la valeur de $2n$ éléments de la suite pseudo-aléatoire d_n . On écrit les récurrences donnant les n derniers en fonction des n premiers, ce qui fournit n relations permettant de retrouver les coefficients et donc de casser le chiffre.

Ici, supposons qu'on connaisse ... 1 0 0 1 1 0 1 0 ... comme partie de la suite d_n . Si on cherche $d_n = xd_{n-1} + yd_{n-2} + zd_{n-3} + td_{n-4}$, où les coefficients cherchés sont x, y, z et t, on obtient :

$$\begin{cases} x + t = 1 \\ x + y = 0 \\ y + z = 1 \\ x + z + t = 0 \end{cases}$$

On trouve facilement comme solution unique $x = y = 0$, $z = t = 1$. La suite utilisée comme clé est alors complètement connue.

Les suites pseudo-aléatoires sont en général obtenues par des relations linéaires de ce type (car elles donnent une bonne approximation d'équirépartition). Mais la période de la suite n'est pas le bon paramètre pour trouver comment elle est engendrée.

Les générateurs pseudo-aléatoires sont en général très peu sûrs pour une utilisation cryptographique...