ANNE-CLAIRE BROUARD DÉCEMBRE 2006

PROJET :

LA CRYPTOGRAPHIE

LM 206 : INITIATION A SCILAB

UNIVERSITÉ PIERRE ET MARIE CURIE

PROFESSEUR : NICOLAS LIMARE

La cryptographie est le fait de transformer un message pour tenter de le rendre inintelligible par toute autre personne que son destinataire, ce qui permet de protéger les données qu'il contient.

Elle est très utile depuis longtemps dans le domaine militaire, et a connu plus récemment un nouvel essor avec Internet.

Après un bref rappel historique, nous étudierons plus particulièrement les 2 codes Scilab suivants :

- le chiffre de Vigenère,
- la signature chiffrée RSA.

♦♦♦♦♦♦♦♦♦

I - Historique

Un des plus anciens systèmes connus est le Chiffre de César. Ce système est un algorithme de cryptage par décalage.

Le Chiffre de César consistait à remplacer chaque lettre de l'alphabet par la lettre venant trois places après et donc, pour déchiffrer le message, le destinataire devait remplacer chaque lettre par la lettre venant trois places avant dans l'alphabet. Le destinataire devait donc connaître la clé du chiffrage, c'est-à-dire l'information lui permettant de déchiffrer le message.

Ce système n'était pas très efficace car le code est facile à casser puisqu'il n'y a que 26 possibilités à essayer.

Un système plus efficace consiste à substituer chaque lettre par une autre. La clé de cryptage consiste alors simplement dans la possession d'un tableau de correspondance qui donne pour chaque lettre de l'alphabet la lettre qu'on lui associe dans le code.

Pour éviter d'être obligé de donner le tableau de correspondance, on peut utiliser une phrase qui sert à établir la relation entre l'alphabet normal et l'alphabet modifié. Il faut supprimer les lettres en double de cette phrase et compléter la phrase à partir de la lettre où on en était en rajoutant les lettres non encore utilisées que l'on place dans l'ordre alphabétique. Ainsi la clé de cryptage peut se transmettre oralement.

Celui qui intercepte le message devra essayer toutes les permutations possibles de l'alphabet, ce qui est impossible car il existe 26! possibilités. Ce système de chiffrage resta inviolé pendant plus de quinze siècles.

Ce sont les Arabes qui les premiers ont cassé ce code en utilisant des méthodes statistiques basées sur la fréquence moyenne d'utilisation de telle ou telle lettre dans telle ou telle langue. Par exemple, en français, la lettre « e » est la plus fréquente.

Vigenère a donc inventé un nouveau système de chiffrement au 16ème siècle pour tenter de supprimer cette faiblesse.

II- Le chiffre de Vigenère

Il s'agit également d'un système de chiffrement par substitution mais qui ne substitue pas toujours une lettre donnée par une même lettre, contrairement aux systèmes de chiffrage précédents qui se contentaient d'utiliser la même lettre de substitution. C'est donc un système « plus solide » que le précédent.

Le message à coder choisi se présente sous forme d'une matrice ligne de 13 nombres, chaque nombre correspondant à la position de la lettre dans l'alphabet en commençant la numérotation par 0, ce qui permet de travailler modulo 26 : les lettres de l'alphabet sont donc numérotées de 0 à 25. Cette matrice est le message à envoyer.

On choisit comme clé une matrice de 3 nombres.

On construit une nouvelle matrice de même longueur que la matrice du message et l'on remplit cette nouvelle matrice en répétant les 3 nombres de la clé.

Le codage consiste alors simplement à faire l'addition, modulo 26, des deux matrices : la matrice somme est le message codé.

Pour décoder le message codé, il suffit de soustraire, modulo 26, la matrice constituée par la répétition de la clé de la matrice correspondant au message codé.

Ce système peut sembler résistant au décodage statistique. En fait on peut quand même casser le code par le système de décodage statistique suivant, découvert au milieu du XIXe siècle à la fois par Charles Babbage et Friedrich Wilhelm Kasiski.

On repère dans le message des mots qui se répètent. On calcule la distance entre ces répétitions. La probabilité pour que celles-ci codent le même mot (avec la même partie de la clé) est très grande. La distance entre ces deux mots semblables est alors un multiple de la longueur de la clé de cryptage. On refait cela chaque fois que l'on voit des mots qui se répètent. Le pgcd de ces différentes distances est la longueur de la clé. Comme de temps en temps il y aura des répétitions parasites provenant de mots différents codés avec des parties différentes de la clé, la longueur de la clé ne sera pas toujours le pgcd mais plutôt un nombre divisant la grande majorité des distances trouvées.

On divise alors le texte codé en autant de sous-ensembles qu'il existe de nombres dans la clé de chiffrage, chaque sous-ensemble contenant les lettres codées grâce au même nombre de la clé, c'est-à-dire les lettres ayant la même position dans le texte modulo la longueur de la clé. Ainsi dans chaque sous-ensemble, chaque lettre est codée par le même décalage. On applique ensuite la méthode statistique décrite précédemment à chaque sous-ensemble.

III - Le système de chiffrage RSA

C'est un système de chiffrage révolutionnaire, inventé en 1977 par Ron Rivest, Adi Shamir et Len Adleman, qui permet de ne plus avoir de clé de cryptage secrète : seule la clé de décryptage est secrète. Ainsi il est possible, sans avoir rencontré au préalable une personne pour échanger une clé secrète, de lui communiquer une information qu'elle sera la seule à pouvoir décoder.

Principe de l'encryptage et du décryptage d'un message :

On génère deux « grands » nombres premiers p et q. On calcule le produit n = pq et l'indicatrice d'Euler $\varphi(n) = (p-1)(q-1)$.

On choisit un entier d>1 premier avec $\varphi(n)$. On calcule l'inverse e de d modulo $\varphi(n)$ c'est-à-dire on calcule e tel que $e*d = 1$ modulo $\varphi(n)$.

La clé publique est le couple (e,n) et la clé secrète est le couple $(d, \varphi(n))$. m est le message secret à envoyer.

Le message crypté M est le nombre m^e modulo n.

Pour le décrypter, on élève le message crypté M à la puissance d modulo n. Dans ce cas, ceux qui envoient le message possèdent la clé publique. Seul celui qui reçoit le message possède la clé secrète.

La force du système de codage repose sur le fait que les algorithmes de factorisation ont une complexité beaucoup plus grande que les algorithmes de multiplication et qu'il est donc impossible en pratique de retrouver p et q. Ainsi on ne peut pas calculer $\varphi(n)$ et donc trouver e, l'inverse de d modulo $\varphi(n)$.

Principe de signature chiffrée RSA :

Le système de chiffrage RSA est un système de cryptage permettant de signer un message, c'est-à-dire de rendre possible l'authentification de l'expéditeur. Le but est d'authentifier un message comme provenant bien d'une certaine personne. Cette fois ci, c'est la personne qui envoie le message qui possède la clé secrète et ce sont les personnes qui reçoivent le message codé qui possèdent le clé publique. Tout le monde peut décoder le message mais celui-ci ne peut avoir été codé que par la personne possédant la clé de cryptage.

Le deuxième code réalisé permet d'effectuer une signature RSA.

On introduit d'abord deux nombres premiers 47 et 53 qui permettent d'obtenir n.

On génère ensuite un nombre au hasard en utilisant la fonction Rand. On incrémente progressivement celui-ci pour obtenir un nombre premier avec $\varphi(n)$ en utilisant la fonction Pgcd. On rend ensuite ce nombre inférieur à $\varphi(n)$ en utilisant la fonction Modulo et on obtient le nombre d. On obtient ainsi la clé secrète $(d, \varphi(n))$.

On génère ensuite e en utilisant la fonction Bezout ($de + k\varphi(n) = 1$) et donc la clé publique (e,n).

On génère de nouveau grâce à la fonction Rand un nombre qui sera le message m que l'on signera. On le choisit inférieur à n.

La signature consiste à élever ce nombre à la puissance d Modulo n : $s = m^d \pmod n$. La difficulté est alors qu'on ne peut pas élever un grand nombre à une grande puissance sans faire

apparaître des arrondis : la solution est alors d'effectuer des simplifications Modulo n au fur et à mesure. Pour cela, on remplace la fonction puissance par une succession de multiplications par m : dès que le résultat dépasse n, on le réduit grâce à la fonction Modulo. Le message signé est le couple (m, s).

```
m = floor (n*rand( ) ) ;
s = 1 ;
for i = 1 : k ;
s = pmodulo (s*m, n) ;
end ;
[m, s]
```

On vérifie ensuite la signature en élevant s à la puissance e modulo n de la même façon que précédemment, et l'on s'assure alors que l'on retrouve bien le message m.

♦♦♦♦♦♦♦♦♦

L'utilisation de Scilab pour l'implémentation du chiffre de Vigenère et l'implémentation de la signature chiffrée RSA m'a permis de mieux appréhender concrètement ces systèmes de chiffrage et de réaliser l'importance des diverses fonctions Scilab d'arithmétique pré-définies telles que par exemple les fonctions Pgcd, Bezout ou Modulo.

Un prolongement intéressant que j'aimerais avoir l'occasion d'étudier ultérieurement serait d'apprendre les techniques d'attaque contre le code RSA, c'est-à-dire comment décoder les messages sans connaître la clé, ce qui revient principalement à factoriser rapidement de grands entiers.