Histoire de la cryptographie

Frédéric Havet

MASCOTTE, commun I3S(CNRS/UNSA)-INRIA Sophia Antipolis

Fête de la science – 21-24 octobre 2010

Messages secrets

Depuis l'Antiquité, on cherche à envoyer des messages sans que des personnes extérieures ne puissent les intercepter.

Le plus vieux document chiffré date du XVI^{e} siècle avant J. C.

Deux manières complémentaires de faire:

- STEGANOGRAPHIE: cacher le mesage pour que l'ennemi ne le trouve pas.
- CRYPTOGRAPHIE: rendre le message incompréhensible par l'ennemi.

Chiffre des Hébreux

V^{e} siècle av. J.-C.: premières techniques de chiffrement par les Hébreux.

Plus connu Atbash pour **a**leph, **t**au, **b**eth, **sh**inest.
Chiffre par substitution alphabétique inversée.
A devient Z, B devient Y, C devient X, ...

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

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

J'UTILISE ATBASH devient Q'FGRORHV ZGYZHS

Que veut dire QV XLNKIVMWH O'SVYIVF ?

Chiffre des Hébreux

V^{e} siècle av. J.-C.: premières techniques de chiffrement par les Hébreux.

Plus connu Atbash pour **a**leph, **t**au, **b**eth, **sh**inest.
Chiffre par substitution alphabétique inversée.
A devient Z, B devient Y, C devient X, ...

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

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

J'UTILISE ATBASH devient Q'FGRORHV ZGYZHS

Que veut dire QV XLNKIVMWH O'SVYIVF ?
JE COMPRENDS L'HEBREU

Chiffre de César

César utilisait un chiffrement par décalage. Chaque lettre est remplacée par la lettre décalée de k dans l'alphabet.

César décalait toutes les lettres de 3. Ainsi A devient D, B devient E, ..., X devient A, Y devient B et Z devient C.

JE CHIFFRE PAR DECALAGE se code
MH FKLIIUH SDU GHFDODJH

Que veut dire IDFLOH D OLUH OH FRGH GH FHVDU?

Chiffre de César

César utilisait un chiffrement par décalage. Chaque lettre est remplacée par la lettre décalée de k dans l'alphabet.

César décalait toutes les lettres de 3. Ainsi A devient D, B devient E, ..., X devient A, Y devient B et Z devient C.

JE CHIFFRE PAR DECALAGE se code
MH FKLIIUH SDU GHFDODJH

Que veut dire IDFLOH D OLUH OH FRGH GH FHVDU?
FACILE A LIRE LE CODE DE CESAR

Inconvénients et avantages du chiffrement par décalage

Peu sûr: Si on sait que le chiffrement est par décalage alors on peut retrouver le message: il n'y a que 25 possibilités de décalage.

Simple: Très simple à utiliser et à se rappeler.

- officiers sudistes pendant la Guerre de Sécession.
- l'armée russe en 1915.
- de nos jours sur les forums internet: ROT13 (décalage de 13 lettres). But est d'empêcher la lecture involontaire: (d'une réponse à une devinette, de la fin d'un film, ...).

Chiffrement par substitution mono-alphabétique

Remplacement d'une lettre par une autre suivant une table.

Exemple: Atbash

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

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

Exemple: ROT-13

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

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

Chiffrement par substitution mono-alphabétique

On peut utiliser d'autres tables "plus compliquées".

A	B	C	D	E	F	G	H	I	J	K	L	M
R	H	N	Y	C	Q	F	U	W	A	J	O	Z

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
X	M	K	S	I	T	G	P	E	D	V	B	L

KRT QRNWOC YC TC IRKKCOCI PXC GCOOC GRHOC.

Chiffrement par substitution mono-alphabétique

On peut utiliser d'autres tables "plus compliquées".

A	B	C	D	E	F	G	H	I	J	K	L	M
R	H	N	Y	C	Q	F	U	W	A	J	O	Z

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
X	M	K	S	I	T	G	P	E	D	V	B	L

KRT QRNWOC YC TC IRKKCOCI PXC GCOOC GRHOC.
PAS FACILE DE SE RAPPELER UNE TELLE TABLE.

Chiffrement par substitution mono-alphabétique

On peut utiliser d'autres tables "plus compliquées".

A	B	C	D	E	F	G	H	I	J	K	L	M
R	H	N	Y	C	Q	F	U	W	A	J	O	Z

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
X	M	K	S	I	T	G	P	E	D	V	B	L

KRT QRNWOC YC TC IRKKCOCI PXC GCOOC GRHOC.
PAS FACILE DE SE RAPPELER UNE TELLE TABLE.

Avantage: il y a $26! \sim 10^{27}$ tables possibles.
Inconvénient: table est difficile à se rappeler.

Analyse fréquentielle

Découverte au IX^{e} siècle par Al-Kindi.

Idée: examiner la fréquence des lettres d'un message chiffré.
En effet, la fréquence dépend de la lettre :

A	B	C	D	E	F	G	H	I	J	K	L	M
9,4	1,0	2,6	3,4	15,9	0,9	1,0	0,8	8,4	0,9	0,0	5,3	3,2

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
7,1	5,1	2,8	1,0	6,5	7,9	7,3	6,2	2,1	0,0	0,3	0,2	0,3

On a l'ordre suivant:
E,A,I,S,T,N,R,U,L,O,D,M,P,C,V,Q,G,B,F,J,H,Z,X,Y,K,W

Décoder par analyse fréquentielle

Pour décoder un texte chiffré par substitution, on fait une analyse de fréquences sur les lettres.

La lettre la plus fréquente est très certainement le E, la deuxième le A, la troisième le I, etc ....
Attention: inversion possible dans l'ordre. Surtout pour des lettres de fréquences proches et le texte court.

En général, cela permet d'identifier les lettres les plus fréquentes. On peut ensuite deviner les autres,...

Analyse de fréquences sur les bigrammes = bloc de deux lettres dans un mot. Les plus courants sont ES 3,15%, LE 2,46%, EN 2,42%, DE 2,15%, RE 2,09%, NT 1,97%, ...

Analyse fréquentielle et Scrabble

A	B	C	D	E	F	G	H	I	J	K	L	M
9,4	1,0	2,6	3,4	15,9	0,9	1,0	0,8	8,4	0,9	0,0	5,3	3,2

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
7,1	5,1	2,8	1,0	6,5	7,9	7,3	6,2	2,1	0,0	0,3	0,2	0,3

Voici les lettres du scrabble avec leur points ainsi que leur nombre dans le jeu.

A_1	B_3	C_3	D_2	E_1	F_4	G_2	H_4	I_1	J_8	K_{10}	L_1	M_2
9	2	2	3	15	2	2	2	8	1	1	5	3

N_1	O_1	P_3	Q_8	R_1	S_1	T_1	U_1	V_4	W_{10}	X_{10}	Y_{10}	Z_{10}
6	6	2	1	6	6	6	6	2	1	1	1	1

Répartition des lettres du premier Scrabble vient d'une analyse fréquentielle du New York Times.

Sherlock Holmes et l'analyse fréquentielle

Les Hommes dansants (The Adventure of the Dancing Men)

De curieux gribouillages apparaissent dans la propriété des Cubitt.

Holmes déchiffre le code de ces gribouillages grâce à l'analyse fréquentielle.

Indice de coïncidence

Introduit par W. Friedman en 1920.

IC mesure la probabilité de trouver une paire de lettres identiques.

$$IC = \sum_{q=A}^{q=Z} \frac{n_q}{n} \cdot \frac{n_q - 1}{n - 1}$$

avec n nombre total de lettres,
n_A nombre de A, n_B nombre de B, ...

Ce nombre ne varie pas après substitution mono-alphabétique.
Français: 0,0778 Anglais: 0,0667 Russe: 0,0529
Aléatoire: $1/26 = 0{,}0385$

Chiffrement par substitution poly-alphabétique

XVI^{e} siècle

- 1518: JeanTrithème *Polygraphiae*
- 1553: Giovan Battista Bellaso *La Cifra*
- 1563: Giambattista della Porta *De Furtivis Literarum Notis, vulgo de ziferis*
- 1586: Blaise de Vigenère *Traicté des chiffres ou secrètes manières d'escrire*

Principe: Clé littérale qui indique le décalage à appliquer.

Chiffre de Vigenère

Clé que l'on répète à l'infini indique le décalage.

Exemple: ABCD la clé.

Texte	P	R	E	N	O	N	S	U	N	E	X	E	M	P	L	E
Clé	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Chiffre	Q	T	H	R	P	P	V	Y	O	G	A	I	N	R	O	I

Avantages:

- Simple.
- Résiste à l'analyse fréquentielle.

Cryptanalyse du chiffre de Vigenère

Charles Babbage (?) 1854

Friedrich Wilhelm Kasiski 1863: Test permettant d'estimer la taille de la clé basé sur les écarts entre les séquences redondantes.

Une fois la taille de la clé trouvée: analyse fréquentielle pour chaque lettre de la clé.

Victoires de la cryptanalyse

1ere Guerre mondiale: Les Français décryptaient les messages pour les sous-marins allemands.

2nde Guerre Mondiale: Les Britanniques ont pu décrypter les communications de l'armée allemande.
Machine Enigma: substitutions poly-alphabétiques avec changement de clé permanent créé par des rotors.
On estime à plusieurs mois (voir années) l'écourtement de la Guerre.

Chiffre de Vernam ou masque jetable

Créé par par Gilbert Vernam en 1917 et perfectionné par Joseph O. Mauborgne.

Principe: Substitution poly-alphabétique avec une clé particulière.

- Clé aussi longue que le message à chiffrer.
- Les caractères de la clé doivent être aléatoires.
- Chaque clé (masque) doit avoir une utilisation unique (jetable).

Chiffre de Vernam ou masque jetable

Avantage: Théoriquement sûr et simple à coder/décoder.

Inconvénient: Difficulté de la mise en place.

- Transmission de la clé difficile $\Rightarrow$ valise diplomatique.
- Génération de clé aléatoire impossible $\Rightarrow$ nombres pseudo-aléatoires. (risque de faille).
- Utilisation impérativement unique. Si on a deux messages codés avec la même clé on peut très souvent les décoder.

Reste cependant très utilisé: ambassades, téléphone rouge.

Systèmes à clés publiques

Grand progrès: systèmes à clés publiques.
Caractéristiques principales:

- simplicité;
- clé publique;
- difficulté à briser le code.

Idée proposée en 1976 par Diffie & Hellman.
Création du premier système en 1978 par Rivest, Shamir & Adleman: le crypto-système RSA

Crypto-système RSA

Une personne **Alice** veut recevoir un message d'une autre personne **Bob**.

Le crypto-système RSA comprend 3 étapes.

1. Choix de la clé et sa publication par Alice.
2. Chiffrement du message par Bob et envoi.
3. Déchiffrement du message par Alice par clé privée.

Choix de la clé RSA

Alice choisit deux grands entiers naturels premiers p et q (100 chiffres chacun ou plus) et fait leur produit $n = pq$.

Puis elle choisit un entier e premier avec $(p-1)(q-1)$.

Enfin, elle publie dans un annuaire, par exemple sur le web, sa clé publique RSA: (n, e).

Exemple: $p = 53$, $q = 97$ donc $n = 5141$ et $e = 7$.

Chiffrement RSA

- **Bob** va chercher la clé d'**Alice**: (n, e).
- Il transforme son message. Par exemple, il remplace chaque lettre par son rang dans l'alphabet.
  "JEVOUSAIME" devient : "10 05 22 15 21 19 01 09 13 05".
- Il coupe son message en blocs de même longueur, chacun représentant un nombre plus petit que n.
  Attention: Les blocs doivent être assez longs sinon l'analyse fréquentielle s'applique.

  Son message devient : "010 052 215 211 901 091 305"
- Chaque bloc B est chiffré par la formule $C = B^e \mod n$, où C est un bloc du message chiffré que **Bob** enverra à **Alice**
  Message chiffré: "0755 1324 2823 3550 3763 2237 2052". .

Déchiffrement RSA

Alice calcule à partir de p et q, qu'elle a gardés secrets, la clé d de déchiffrage (c'est sa clé privée).
d doit satisfaire $e \times d \mod ((p-1)(q-1)) = 1$. Ici, d=4279.

Chacun des blocs C du message chiffré est déchiffré par la formule $B = C^d \mod n$.

Alice retrouve : "010 052 215 211 901 091 305"

Principe de RSA

- Il est facile de multiplier deux grands nombres premiers.
- Il est très difficile de déterminer les facteurs premiers d'un grand nombre.
  Il est actuellement impossible pour certaines catégories de nombres de "factoriser" ceux de plus de 300 chiffres.

RSA est partout

Plusieurs centaines de millions de programmes l'utilisent.

- transactions sécurisées sur internet;
- systèmes d'exploitation (Apple, Microsoft, Sun);
- confidentialité du courrier électronique;
- très grand nombre d'institutions.

Sécurité de RSA

Théoriquement elle est basée sur 2 conjectures:

1. "casser" RSA nécessite la factorisation du nombre n en le produit initial des nombres p et q.

2. avec les algorithmes classiques, le temps que prend cette factorisation croit exponentiellement avec la longueur de n.

Pratiquement RSA a résisté à toutes les attaques depuis 30 ans.

Quelques précautions

Clés longues: 256 bits se fait en quelques heures sur 1 PC.
768 bits plus grands nombres factorisés.
Des experts estiment des clés de 4096 bits sont sures.

Autres précautions:

- Prendre p et q de taille proche.
- Eviter que $p+1$, $p-1$, $q-1$ et $q+1$ soient faciles à factoriser.
- Prendre e suffisamment grand $(> \sqrt[4]{n})$

Faiblesse de RSA

Les conjectures sur lesquelles reposent la sécurité de RSA sont peut-être fausses.

Ordinateurs quantiques:

Algorithme de Shor: algorithme quantique pour factoriser un nombre n en temps $O((\log n)^3)$ et en espace $O(\log n)$.

En 2001, par un groupe d'IBM, factorisa 15 en 3 et 5, en utilisant un calculateur quantique de 7 qubits.