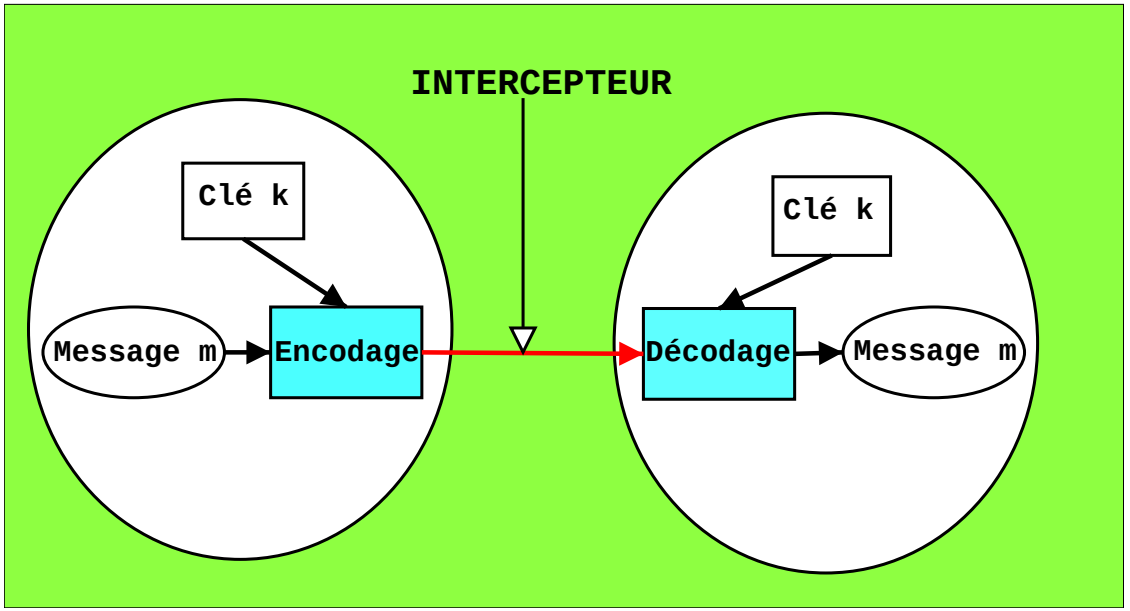
Introduction aux codes secrets

Michel Van Caneghem

Février 2003

Les codes secrets

Le chiffrage - déchiffrage ou codage, décodage est une science qui s'appelle la cryptographie



On ne peut envisager que l'ensemble du système, c'est à dire à la fois comment décoder et coder un message, mais aussi s'il est possible d'intercepter un message codé.

Pourquoi le secret

Traditionellement les codes secrets étaient l'affaire des militaires et des diplomates. Maintenant c'est également :

- **Le problème des banques (transfert d'argent, carte de crédits,...)**
- **Le problème des informaticiens (mot de passe, protection contre le piratage)**
- **Le problème des sociétés (coffre fort électronique)**
- **Le problème de tout le monde : Internet et la protection de la vie privée.**

Quel secret

Pour élaborer un système de protection il faut évaluer :

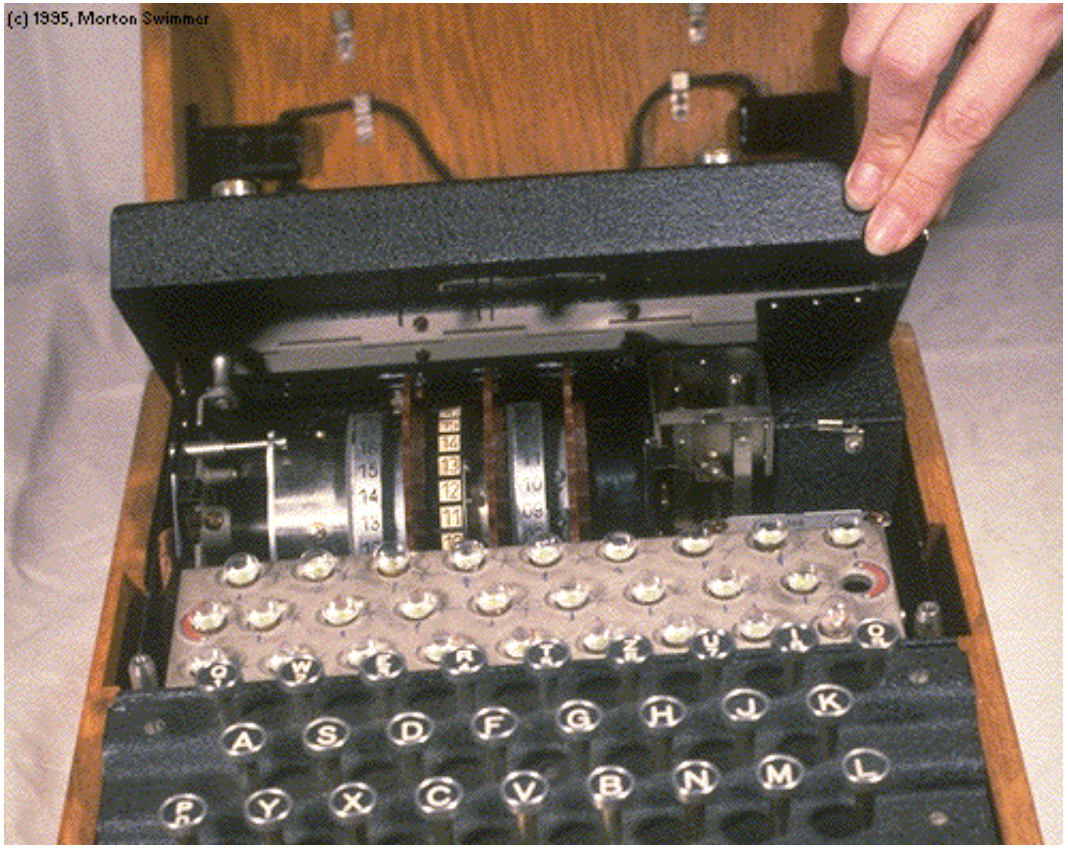
- ✗ **La sécurité recherchée (durée de vie de l'information)**
- ✗ **La taille de la clé**
- ✗ **La simplicité et l'efficacité des opérations de codage et décodage**

Pour évaluer la protection, il faut imaginer le pire des cas : c'est à dire le meilleur pour l'intercepteur :

- ✘ **Le décrypteur à une complète connaissance du système**
- ✘ **Le décrypteur a accès à une quantité importante de texte codé.**
- ✘ **Le décrypteur connait une partie en clair du message codé.**

La machine Enigma

utilisée par les allemands pendant la dernière guerre



Lettre de George Sand à Alfred de Musset

Je suis très émue de vous dire que j'ai
bien compris l'autre soir que vous aviez
toujours une envie folle de me faire
danser. Je garde le souvenir de votre
baiser et je voudrais bien que ce soit
là une preuve que je puise être aimée
par vous. Je suis prête à vous montrer mon
affection toute désintéressée et sans cal-
cul, et si vous voulez me voir aussi
vous dévoiler sans artifice mon âme
toute nue, venez me faire une visite.
.....*George Sand*

Réponse d’Alfred De Musset :

Quand je mets a vos pieds un éternel hommage,
Voulez-vous qu’un instant je change de visage ?
Vous avez capturé les sentiments d’un coeur
Que pour vous adorer forma le créateur.
Je vous chéris, amour, et ma plume en délire
Couche sur le papier ce que je n’ose dire.
Avec soin de mes vers lisez les premiers mots,
Vous saurez quel remède apporter à mes maux.

Alfred de Musset

Ce que l'on va examiner

Les premiers ordinateurs ont été construits pendant la guerre pour pouvoir casser les codes secrets allemand et autres...

Nous étudierons :

- **Les codes secrets anciens**
- **Comment casser les codes secrets anciens ?**
- **Le DES**
- **Les codes a clés publiques : RSA**

Chiffre monoalphabétique

On remplace un alphabet par un autre. Il y a en tout $26!$ codes possibles. Un sous ensemble auquel on peut s'intéresser : les codes basés sur l'arithmétique modulaire.

La clé est constituée de deux nombres a et b. On chiffre avec la formule suivante :

$$C \equiv aP + b \pmod{26}$$

C représente le texte codé (Coded) et P représente le texte en clair. Chaque lettre est transformée en un chiffre ($A = 0$, $Z = 25$). On décode avec :

$$P \equiv a^{-1}(C - b) \pmod{26}$$

Chiffre monoalphabétique (2)

à condition que $a \wedge 26 = 1$. Il y a donc $12 \times 26 = 312$ codes possibles. Si $a = 1$ et $b = 3$, on obtient le code de César.

On remarque que les fréquences des lettres ne change pas avec un tel code : d'ou la méthode pour casser le code. On définit la probabilité de coincidence comme la probabilité que deux lettres prises au hasard soient identiques :

$$Ic = \sum_{\lambda=A}^{Z} p_\lambda^2 = \frac{\sum_{\lambda=A}^{Z} f_\lambda(f_\lambda - 1)}{n(n-1)}$$

$$\chi_p = \sum_{\lambda=A}^{Z} p_\lambda^2 \quad \chi_{francais} = 0,079 \quad \chi_{anglais} = 0,065 \quad \chi_{alea} = 0,038$$

Fréquence des lettres du français

A	0.083944
B	0.007669
C	0.033297
D	0.040699
E	0.145037
F	0.012109
G	0.009495
H	0.007973
I	0.081828

J	0.006377
K	0.000638
L	0.058405
M	0.029355
N	0.075570
O	0.053669
P	0.032087
Q	0.012613
R	0.070209

S	0.080091
T	0.074775
U	0.059808
V	0.015791
W	0.000067
X	0.004098
Y	0.003155
Z	0.001240

Les plus fréquentes : E (145), A (84), I (81), S (80), N (76), T (75), R (70).

Fréquence des lettres (2)

Le Scrabble

E	15	L	5	D	3	Q	1	H	2
A	9	O	6	C	2	V	2	X	1
N	6	U	6	P	2	G	2	J	1
S	6			M	3	F	2	Y	1
I	8					B	2	K	1
T	6							Z	1
R	6							W	1

$$N = 100, \qquad \chi_p = 0,069$$

Remarque

Attention : ces chiffres ne représentent qu'une moyenne. Voici un exemple ou la table de fréquence est fausse : PEREC, Georges, La disparition, Gallimard, L'imaginaire, Paris, 1969. qui est un livre écrit sans la lettre E. En voici un extrait :

«Il y avait au mur un rayon d'acajou qui supportait vingt-six in-folios. Ou plutôt, il aurait dû y avoir vingt-six in-folios, mais il manquait, toujours, l'in-folio qui offrait (qui aurait dû offrir) sur son dos l'inscription "CINQ". Pourtant, tout avait l'air normal : il n'y avait pas d'indication qui signalât la disparition d'un in-folio (un carton, "a ghost" ainsi qu'on dit à la National Library) ; il paraissait n'y avoir aucun blanc, aucun trou vacant. Il y avait plus troublant : la disposition du total ignorait (ou pis : masquait, dissimulait) l'omission : il fallait la parcourir jusqu'au bout pour savoir, la soustraction aidant (vingt-cinq dos portant subscription du "UN" au "VINGT-SIX", soit vingt-six moins vingt-cinq font un), qu'il manquait un in-folio ; il fallait un long calcul pour voir qu'il s'agissait du "CINQ".»

Un exemple de déchiffrage

texte en clair

Voici un premier texte qui est tres facile a decoder. Il faut quand meme que le texte soit assez long.

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

texte codé

HAUOU GZBDQ YUQDF QJFQC GUQEF FDQER MOUXQ MPQOA PQDUX RMGFC GMZPY QYQCG QXQFQ JFQEA UFMEE QLXAZ S

Chiffre de Vigenère

vers 1560, le diplomate français Blaise de Vigenère, s'inspirant des travaux de Leon Batista Alberti et de Jean Trithème, mit au point une grille de 26 alphabets (autant que de lettres), chacun étant décalé d'un cran par rapport au précédent. On plaçait au-dessus le texte à coder, et on en remplaçait successivement chaque lettre par l'une de celles qui se trouvait en-dessous, en changeant de ligne à chaque fois, selon un ordre donné par un mot clé.

Le système polyalphabétique de Vigenère résista pendant environ trois siècles, jusqu'à ce que le mathématicien britannique Charles Babbage élabore la théorie de son décodage, vers 1854.

Chiffre de Vigenère (2)

Méthode de chiffre par substitution polyalphabétique. Une suite de lettre est prise comme clé : $k_1, k_2, \ldots, k_n$. On découpe alors le message en blocs de longueur n. Soit $p_1, p_2, \ldots, p_n$ un tel bloc. Il est alors codé de la manière suivante :

$$c_i \equiv p_i + k_i \pmod{26}$$

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
CDEFGHIJKLMNOPQRSTUVWXYZAB
LMNOPQRSTUVWXYZABCDEFGHIJK
EFGHIJKLMNOPQRSTUVWXYZABCD
```

(1) : $k_1 = +2$

(2) : $k_2 = +11$

(3) : $k_3 = +4$

Chiffre de Vigenère (3)

Message original : UNCOURSPASSIONNANT**, clé =** CLE

C	L	E	C	L	E	C	L	E	C	L	E	C	L	E	C	L	E
2	11	4	2	11	4	2	11	4	2	11	4	2	11	4	2	11	4
U	N	C	O	U	R	S	P	A	S	S	I	O	N	N	A	N	T
W	Y	G	Q	F	V	U	A	E	U	D	M	Q	Y	R	C	Y	X

voici le message codé :

WYGQF VUAEU DMQYR CYX

Déchiffrage du code de Vigenère

L' avantage principal de ce code est que l'on ne peut plus se servir directement des fréquences des lettres pour le déchiffrer. Il y a cependant des méthodes pour casser ce code.

Il faut essayer de deviner la taille de la clé (m). Ensuite si le texte est assez long (n), chaque ligne correspond à un code mono-alphabétique simple.

Remarque 1 : On calcule tout d'abord l'indice de coincidence Ic :

$$Ic = \frac{\sum_{\lambda=A}^{Z} f_\lambda(f_\lambda - 1)}{n(n-1)}$$

Déchiffrage du code de Vigenère (2)

Cet indice, qui représente la probabilité pour que 2 lettres tirées au hasard soient égales se décompose en 2 :

1. **Soit les deux lettres sont a une distance de m l'une de l'autre, et dans ce cas cette probabilité correspond à l'indice de coincidence de la langue du message. Cet évenement arrive avec la probabilité $1/m$.**
2. **Soit les deux lettres sont a une distance différente de m l'une de l'autre, et dans ce cas on peut considérer que le texte est aléatoire. La probabilité de coincidence est donc de $1/26$. Cet évenement arrive avec une probabilité de $(m-1)/m$**

Déchiffrage du code de Vigenère (3)

On a donc :

$$Ic = \frac{1}{m}\chi + \frac{m-1}{m} \times \frac{1}{26}$$

$$m = \frac{26\chi - 1}{26Ic - 1}$$

On trouve :

- **pourle français :** $m = 1.054/(26Ic - 1)$
- **pour l'anglais :** $m = 0.69/(26Ic - 1)$

Remarque : Test de Kasiski : Si une même séquence se répète alors la distance entre les deux séquences doit être un multiple de la taille de la clé.

Déchiffrage du code de Vigenère (3)

Un exemple en anglais :

```
OOBQB PQAIU NEUSR TEKAS RUMNA RRMNR ROPYO DEEAD ERUNR
QLJUG CZCCU NRTEU ARJPT MPAWU TNDOB GCCEM SOHKA RCMNB
YUATM MDERD UQFWM DTFKI LROPY ARUOL FHYZS NUEQM NBFHG
EILFE JXIEQ NAQEV QRREG PQARU NDXUC ZCCGP MZTFQ PMXIA
UEQAF EAVCD NKQNR EYCFI RTAQZ ETQRF MDYOH PANGO LCD
```

Ic = 0,0477, m = 2,87 (3)

```
PQA   150    DER    57    CZCC   114
RTE    42    RUN   117    MNB     42
ROPY   81    UNR    12    ARU     42      UEQ 54
```

Tous multiples de 3. => longueur de la clé : 3

Déchiffrage du code de Vigenère (4)

On réécrit en 3 rangées :

```
1 : OQQUUTAUAMRYEDUQUZUTAPPUDGEOAMYTDDFDKRYU
2 : OBANSESMRNOOEENLGCNERTATOCMHRNUMEUWTIOAO
3 : BPIERKRNRRPDARRJCCRUJMWNBCSKCBAMRQMFLPRL
..1 : FZUMFEFXQQQEQUXZGZQXUAADQEFTZQMOAOD
..2 : HSENHIEINERGANUCPTPIEFVNNYIAERDHNL
..3 : YNQBGLJEAVRPRDCCMFMAQECKRCRQTFYPGC
```

Ic1 = 0.072, Ic2 = 0.064, Ic3 = 0.064

Il reste à déchiffrer chacune des lignes. On profite de l'idée consistant à décaler les codes des caractères d'une ligne par rapport à l'autre et calculer l'indice de coincidence pour les deux lignes mises bout à bout.

Déchiffrage du code de Vigenère (5)

On trouve les maxima suivant :

Ligne 1 et Ligne 2 : décalage 14
Ligne 1 et Ligne 3 : décalage 12
Ligne 2 et Ligne 3 : décalage 24 (12 - 14 + 26)

$$k_1 = 0, \quad k_2 = 14, \quad k_3 = 12$$

```
ABCDEFGHIJKL-M-NOPQRSTUVWXYZ
OPQRSTUVWXYZ-A-BCDEFGHIJKLMN
MNOPQRSTUVWX-Y-ZABCDEFGHIJKL
```

On peut appliquer le décalage sur tout le texte et chercher la lettre la plus fréquente. *Voir démonstration*

Le devoir 1

Le but du devoir est de faire un programme Maple qui casse automatiquement le code de Vigenère

- **Décrire (1 page max) le codage de Vigenère et faire les programmes qui codent et décodent**
- **Décrire (2 pages max) la méthode pour casser ce code et faire les programmes correspondant (calcul de la longueur de la clé et décryptage du texte)**
- **Decoder le code 4 du “Cypher Challenge”**
- **Decoder le message que je vais envoyer à chaque binôme.**

Histoire des codes secrets

Le livre de Simon Singh, "Histoire des codes secrets" (Editions JC Lattès, 1999), propose une série de 10 codes secrets à décoder. Un prix de 10.000 livres (100.000 francs environ) est offert à celui qui décodera l'ensemble. [trop tard tout est déja découvert]

Le langage des textes à décoder est variable. Il s'agit du français pour le code 1 ; après il y a un peu de tout.

- **Le code 1 est un code de substitution classique (chaque lettre est remplacée par une autre).**
- **Le code 2 est un chiffre de César.**
- **Le code 3 est un code utilisant la substitution homophonique. Chaque lettre est remplacée par une autre, mais certaines lettres du texte codé remplacent une lettre identique dans le message original. Cela s'applique spécialement aux langues qui comptent peu de lettres dans leur alphabet. Particulièrement ardu.**
- **Le code 4 est un codage de type Vigenère.**