Chiffrement à clé secrète

Vincent Vajnovszki

- 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 ...

L'histoire

- 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

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.

Le chiffrement de César

et chiffrement à substitution simple

A	B	C	D	E	F	G	H	I	J	K	L	...
D	E	F	G	H	I	J	K	L	M	N	O	...

Attaque statistique : consistent à analyser statistiquement les textes cryptés pour déterminer les fréquences d'apparition des symboles

Chiffrement de César : algorithm secret

- 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)

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 :	C	O	D	E	P	O	L	Y	A	L	P	H	A	B
clé :	B	A	C	H	E	L	I	E	R	B	A	C	H	E
texte crypté	D	O	F	L	T	Z	T	C	R	M	P	J	H	F

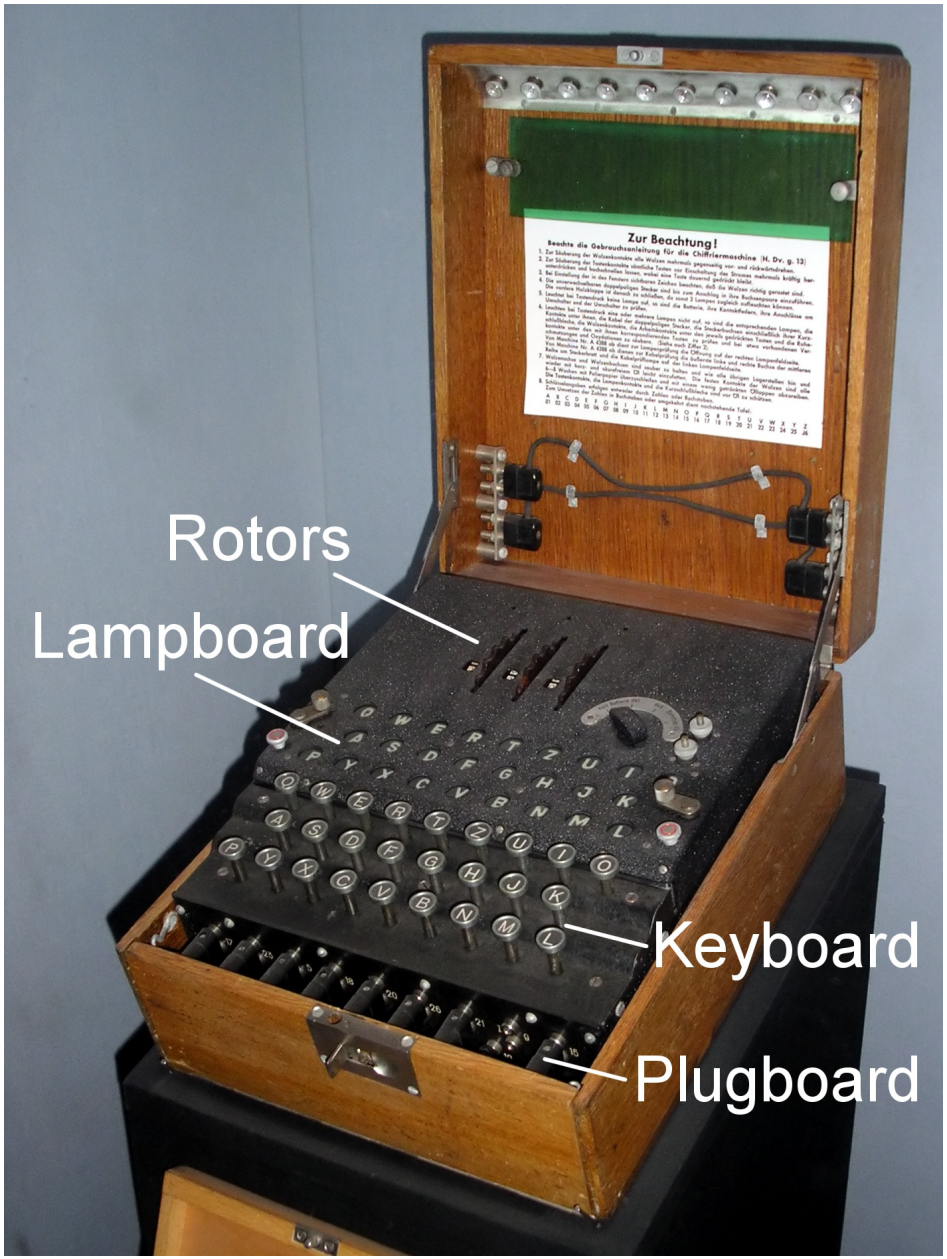
Ce chiffre a été longtemps réputé indécryptable. Un décrypteur qui connaît la longueur de la clé peut effectuer des analyses de fréquence.
Avec les moyens informatiques, si on sait dans quelle langue est le message et si on a un texte crypté assez long, on peut essayer diverses longueurs de clé et garder 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

Aux Temps Modernes

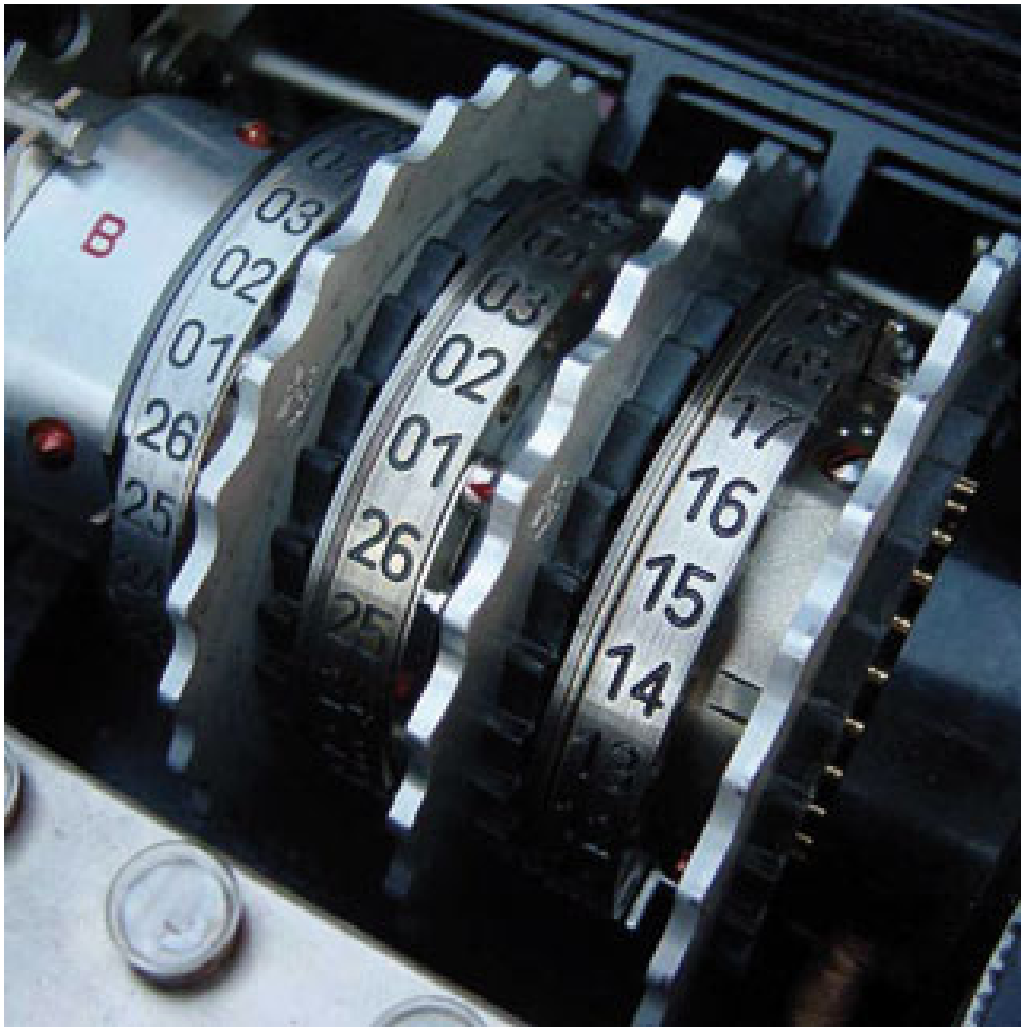
- Jefferson invente un dispositif mécanique de chiffrement, réinventé par Bazeries.
- De l'après-Première Guerre mondiale date l'invention d'un autre dispositif mécanique célèbre, la machine Enigma, qui sera perfectionnée et adoptée pendant la 2e Guerre mondiale par les Allemands.
- Elles sera cassée 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

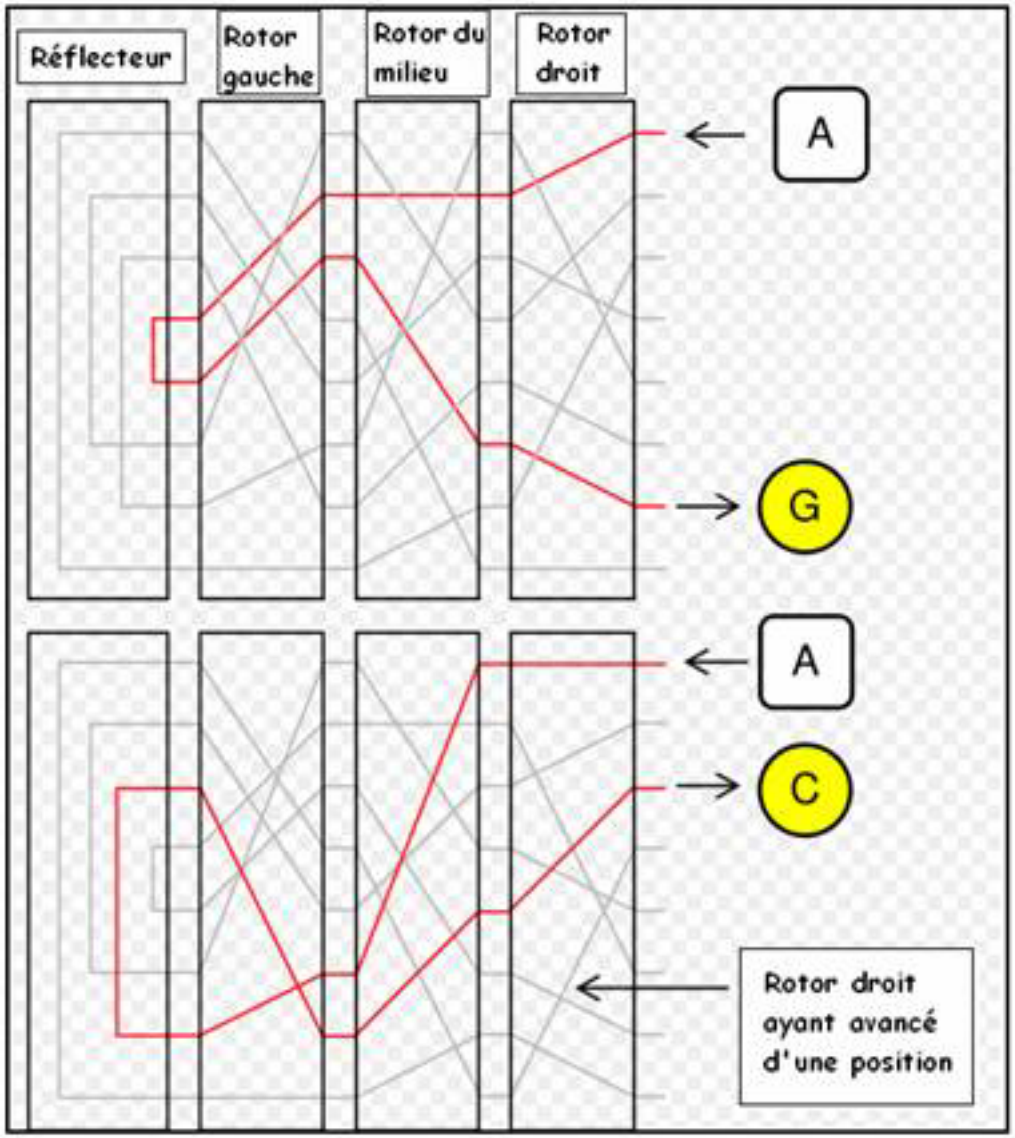
Machine Enigma

Machine Enigma



Detail




Réflecteur
Rotor gauche
Rotor du milieu
Rotor droit
A
G
A
C
Rotor droit ayant avancé d'une position

au centre de Blechley Park

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 (clé secrète)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 ;
  - en plus d'échantillons de texte en clair avec leur version chiffrée ;
  - de la faculté d'obtenir la version chiffrée de textes qu'il a fournis lui-même (correspond à la sécurité maximum possible)

Le perfectionnement ultime 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
Gilbert Vernam et Joseph Mauborgne 1920 : la méthode du One Time Pad = masque jetable
Basé sur :

- une clé privée
- générée aléatoirement
- utilisée une seule fois, puis détruite

Même époque, le Kremlin et la Maison Blanche étaient reliés par le fameux téléphone rouge, c'est-à-dire un téléphone dont les communications étaient cryptées grâce à une clé privée selon la méthode du masque jetable. La clé privée était alors échangée grâce à la valise diplomatique (jouant le rôle de canal sécurisé)

Un système de chiffrement à clé secrète (ou système symétrique) repose sur le partage entre deux personnes en communication, d'une même clé secrète utilisée à la fois pour le chiffrement des données et pour son déchiffrement
Principe :

- Alice crypte le message grâce à un algorithme symétrique à l'aide de la clé
- Le message crypté est communiquer par un moyen quelconque (les données sont illisibles)
- Bob à l'aide de la clé retrouve le message clair

Principe

Le chiffrement par la méthode du masque jetable consiste à combiner le message en clair avec une clé présentant les caractéristiques très particulières suivantes :

1. La clé doit être une suite binaire aussi longue que le message à chiffrer.
2. Les bits composant la clé doivent être choisis de façon totalement aléatoire.
3. Chaque clé, ou *masque*, ne doit être utilisée qu'une seule fois (d'où le nom de masque jetable).

Chiffrement de Vernam : Clé secrète

Si les trois règles ci-dessus sont respectées strictement, le système offre une sécurité théorique absolue (Claude Shannon en 1949)
L'argument théorique est le suivant (dans son principe) :

- si on ne connaît que le texte chiffré, et
- que toutes les clés sont équiprobables
- alors tous les textes clairs sont possibles et avec la même probabilité

Même une attaque par force brute ne donnera aucune information

Désavantages

Si la clé est intercepté par un pirate

- celui-ci peut lire les messages ET,
- en écrire et se faire passer pour un des correspondants

Contraintes

- La clé doit être échangée préalablement à la communication par un canal sûr
- la taille de la clé doit être de taille comparable au message crypté
- la clé doit être aléatoire
- la clé doit être utilisée une seule fois
- dans le cas d'échange entre n personnes,
  - il faut distribuer $\frac{n \cdot (n-1)}{2}$ clés,
  - les temps de chiffrement pour chaque clé = un temps global important

Génération de nombres aléatoires

Les systèmes informatique sont intrinséquement déterministes, le hasard est éliminé.
Pourtant, dans certaines applications on a besoin des valeurs aléeatoires.

- quelles sont ces applicatins ?
- qu'est-ce que c'est le hasard en informatique ?
- comment on le génére ?

Nécessité du hasard

1. *Les simulations* : pour reproduire des phénomènes aléatoires;
2. en *cryptographie*, où, pour pour cacher l'information on la *mélange* avec des valeurs aléatoires;
3. en *théorie de l'information* (au sense de Kolmogorov), ou les suites au plus fort contenu en information sont celles dificilement prédictibles

Qu'est-ce que c'est un nombre aléatoire

Les suites suivantes sont-elles aléatoires ?

- 01234567890123
- 31415926535897
- 82845904523536

Hasard faible : *uniformité* (bon mélange). Peut être généré par des algorithmes rapides. Utile en simulation.
Hasard moyen : Imprévisible pour un observateur ayant des moyens de calcul raisonnables. Peut être produit par des algorithmes efficaces. Utile en cryptographie
Hasard fort : Imprévisibilité totale. Peut être produit par des moyens physiques. Utile en cryptographie et théorie de l'information. Voir : `http://lavarand.sgi.com/`

Générateur pseudo aléatoire

Alice et Bob veulent utiliser le masque jetable de Vernam : il leur faut une source qui génère des bits uniformément distribués.
Savoir si une telle source peut exister ou si tout ce qui arrive est prédéterminé est une question philosophique.
En pratique, on utilise des générateurs de bits aléatoires tant matériels que logiciels. Ce sont des dispositifs qui utilisent, par exemple, le caractère aléatoire de la décomposition radioactive ou le temps écoulé entre deux frappes de clavier

Générateur congruentiel

On se donne les entiers M, a et b, ainsi qu'une *graine* x_0. On calcule itérativement x_n par

$$x_{n+1} = (a \times x_n + b) \bmod M$$

avec $x_n \in \{0, 1, \ldots, M-1\}$

Avec des valeurs convenablemént choisis pour x_0, a et b, $\{x_n\}_{n \geq 1}$ est une suite pseudo aléatoire

Générateurs par registres à décalage linéaire

Généralisation des générateurs congruentiels linéaires

X_n est calculé par des combinaisons linéaires de $X_{n-1}, X_{n-2}, \ldots, X_{n-k}$:

$$X_n = (a_1 \cdot X_{n-1} + a_2 \cdot X_{n-2} + \ldots + a_k \cdot X_{n-k}) \mod m.$$

La graine est : $(X_{k-1}, X_{k-2}, \ldots, X_0)$.

Ces générateurs sont particulièrement intéressants si m est un nombre premier (leur période maximale est alors $m^k - 1$).

Il existe même des puces spécialisées réalisant les opérations nécessaires. On parle alors de *registres à décalage linéaire*, abrégé par *LFSR* (= Linear Feedhack Shift Registers).

Exemple : Le LFSR

$$X_n = X_{n-1} \oplus X_{n-2} \oplus X_{n-4},$$

avec la graine

$$X_0 = 0, X_1 = 1, X_2 = 1, X_3 = 0$$

- correspond à $m = 2$, $a_1 = 1$, $a_2 = 1$, $a_3 = 0$, $a_4 = 1$ et au $k = 4$.
- Les valeurs ainsi obtenues sont : $X_4 = 1$, $X_5 = 0$, $X_6 = 0$, $X_7 = 0$, $X_8 = 1$, $X_9 = 1$, ...

Pour assurer la sécurité du système C. Shannon (de manière informelle) :

- La confusion doit cacher les structures algébriques et statistiques
- La diffusion (effet avalanche) doit permettre à chaque bit du texte claire d'avoir une influence sur une grande partie du texte chiffré

Schéma de Feistel

(avec la notation utilisée par Knudsen dans *Partial and Higher Order Differentials and Applications to DES*).

$$C_1 = (C_1^L, C_1^R) = (C_0^R, f(C_0^R, K) + C_0^L)$$

- C_0 est le texte claire;
- l'exposant représente la sous-partie du bloc considérée : L à gauche, R à droite.
- K est la clé;
- $+$ est une opération dans un groupe commutatif (souvent un *XOR* bit-à-bit);
- f est une fonction non-linéaire.

En général

$$C_i = (C_i^L, C_i^R) = (C_{i-1}^R, f(C_{i-1}^R, K_i) + C_{i-1}^L) \text{ pour } i \geq 1$$

- C_i correspond au bloc de texte chiffré à la sortie du tour i (après la i-ieme itération).
  - C_i^R correspond au bloc de droite à la sortie du tour i et
  - C_i^L correspond au bloc de gauche à la sortie du tour i;
- K_i est la clé du tour i, elle est calculée grâce à un *key schedule* de la clé principale;

Exemple

- le est le texte claire $C_0 = 01100101\ 00010011$ de longueur $n = 16$,
- la clé $K_1 = 11010001$
- $f(x, y)$ le *XOR* bit-à-bit de x et y

On a :

- $C_0^R = 00010011$ et donc $C_1^L = 00010011$
- $C_0^L = 01100101$ et
  - $f(C_0^R, K_1) = 00010011 + 11010001 = 11000010$
  - $f(C_0^R, K_1) + C_0^L = 11000010 + 01100101 = 10100111$, et finalement
  - $C_1^R = 10100111$

Donc, après le premier tour (après la première itération)
$C_1 = (C_1^L, C_1^R) = 0001001110100111$.
Cette opération est reprise $r - 1$ fois (r est le nombre de tours)
avec K_i une permutation circulaire de K_{i-1} (par exemple
$K_2 = 10100011$).

DES = Data Encryption Standard
Une étude fut demandée en 1996 pour estimer le cout et les performances des attaques pour un budget donné

Attaquant	**Budget en euros**	**Outil**	**Clé = 56 bits**
Hacker	300	Logiciel	38 ans
PME	7500	Circuit	18 mois
Grande Entreprise	225 K	Circuit ASIC	19j. 3h
Multinationale	7,5 M	ASIC	6 min
Gouvernement	225 M	ASIC	12 s