Introduction

Historiquement développé pour garantir le secret dans la messagerie, la communication militaire a eu la plus grande influence sur le chiffrement et le cryptage, au 19eme siècle La nécessité d'une communication commerciale privée et sécurisée a été dirigée par l'ère de l'information, qui a commencé dans les années 1980. Bien que l'Internet eût été inventé à la fin des années 1960, il n'a pas gagné un visage public jusqu'à ce que le World Wide Web qui est un protocole électronique qui permet aux gens de communiquer des informations, et de commercer à travers un moyen numérique. été inventé en 1989 Cette nouvelle méthode d'échange d'informations a provoqué un énorme besoin de sécurité de l'information. Une compréhension approfondie de la cryptographie et du chiffrement aidera les gens à développer de meilleures façons de protéger les informations précieuses puisque la technologie devient plus rapide et plus efficace..

le cryptage des informations est maintenant utilisé plus largement pour interdire l'accès ou la modification des informations sensibles et garantir la confidentialité dans les applications informatiques. Cependant, le cryptage n'est qu'un élément dans l'ensemble des dispositifs d'un système complexe. La protection qu'il assure n'est valable que si elle s'insère dans un ensemble cohérent. Le cryptage est le moyen le plus utilisé parmi les transactions sur des canaux non sécurisés de communication, tels que l'Internet. également utilisé pour protéger les données transférées entre les périphériques tels que les distributeurs automatiques de billets (DAB), téléphones mobiles, et beaucoup plus. Il peut être utilisé pour créer les signatures numériques, qui permettent à un message d'être authentifié.

Principe général du cryptage

Le cryptage ou chiffrement des données est généralement décrit à partir de la communication secrète d'informations entre deux interlocuteurs. Dans un système informatique, cette confidentialité intervient en fait sous plusieurs formes, notamment dans la protection du stockage de l'information (copie de fichiers), des accès (interrogation d'une base de données) et de sa transmission ("écoute").

Quel que soit son support physique, l'information est toujours représentée par un codage binaire. Ce codage est conforme à un standard, qui permet de communiquer avec les dispositifs de la machine (claviers, écrans, imprimantes, traceurs ...). Lorsque les ensembles de données à stocker ou à transmettre sont très volumineux (en particulier les images), on fait appel à des techniques de compression pour optimiser le volume et la vitesse de transmission.

Le cryptage repose sur une transformation du code vers une forme non standard, de façon en limiter l'utilisation.

La technique est très ancienne, mais elle a été largement développée et transformée avec l'utilisation intensive de l'informatique et des codages numériques. Ce domaine reste, pour des raisons évidentes, très secret, et les progrès les plus récents ne sont pas divulgués. Cette présentation se limite évidemment aux techniques et outils du domaine public, qui sont ceux effectivement utilisés dans les grandes applications informatiques.

Dans le schéma ci-dessous figurent les différentes branches de la cryptographie classique.

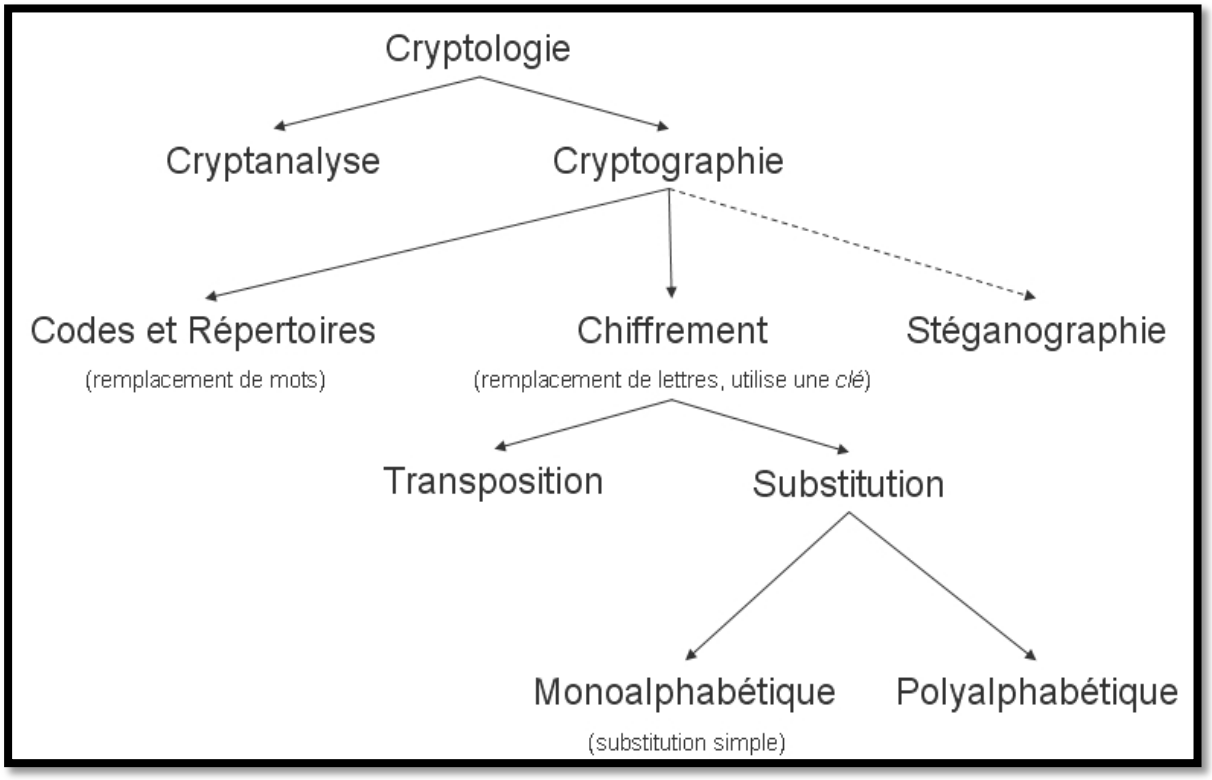


Fig. 3.1 – Domaines inclus dans la cryptologie.

1 Substitution monoalphabétique

Chaque lettre est remplacée par une autre lettre ou symbole. Parmi les plus connus, on citera le chiffre de César, le chiffre affine, ou encore les chiffres désordonnés. Tous ces chiffres sont sensibles à l'analyse de fréquence d'apparition des lettres (nombre de fois qu'apparait une même lettre dans un texte). De nos jours, ces chiffres sont utilisés pour le grand public, pour les énigmes de revues ou de journaux.
Historiquement, on recense des procédés de chiffrement remontant au Xème siècle avant JC. On trouve par exemple, l'Atbash des Hébreux (-500), la scytale à Sparte (-400), le carré de Polybe (-125), . . . Des langues anciennes sont également parfois classifiées dans les codes secrets : le Rongo Rongo, le linéaire
A, les écritures du disque de Phaistos en sont des exemples. Intraduisibles à l'heure actuelle, on les place (à tort ?) dans ce domaine.

1.1 Chiffre de César (50 av. J-C)

Il s'agit d'un des plus simples et des chiffres classiques les plus populaires. Son principe est un décalage des lettres de l'alphabet. Dans les formules ci-dessous, p est l'indice de la lettre de l'alphabet, k est le décalage.

Pour le chiffrement, on aura la formule

$$C = E(p) = (p + k) \bmod 26$$

Pour le déchiffrement, il viendra

$$p = D(C) = (C - k) \bmod 26$$

Si on connait l'algorithme utilisé (ici César), la cryptanalyse par force brute est très facile. En effet, dans le cas du chiffre de César, seules 25 (!) clés sont possibles.

1.2 Analyse de fréquences

Lorsque la langue de départ et la technique de chiffrement sont connus, on peut exploiter les régularités du langage par le principe d'analyse de la fréquence d'une lettre. Cette technique ne fonctionne bien que si le message chiffré est suffisamment long pour avoir des moyennes significatives.

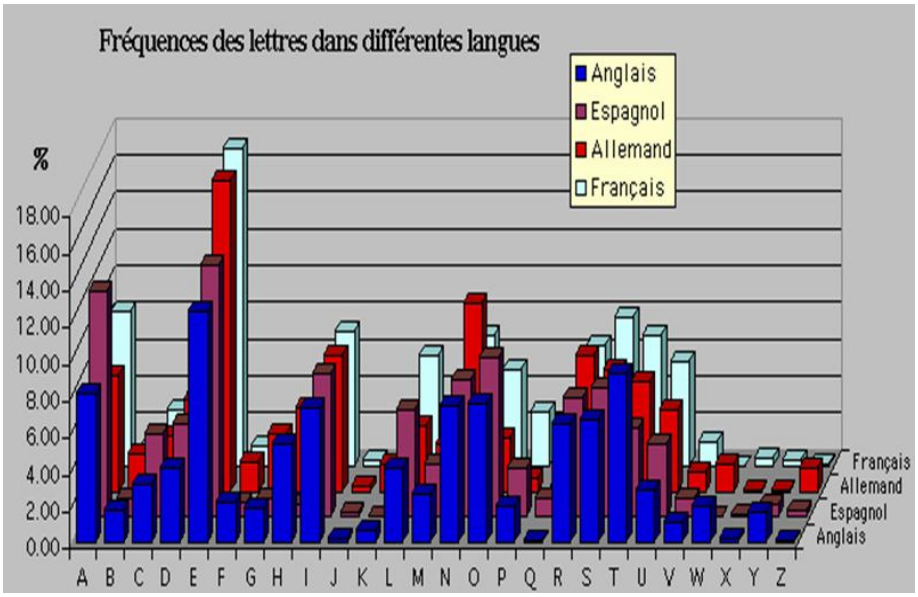


Fig. 3.2 – Exemple d'analyse de fréquence

Cependant, il existe également des cas où cette analyse ne fonctionne pas, comme le montre l'exemple ci-dessous.

Exemple : De Zanzibar à la Zambie et au Zaïre, des zones d'ozone font courir les zèbres en zigzags zinzins.

_pour éviter ce type d'attaque sur un texte chiffré, il existe différents moyens :

– On peut par exemple chiffrer le message par digrammes, trigrammes, etc.

Les 20 bigrammes les plus fréquents

Bigrammes	ES	DE	LE	EN	RE	NT	ON	ER	TE	EL	AN	SE	ET	LA	AI	IT	ME	OU	EM	IE
Nombres	3318	2409	2366	2121	1885	1694	1646	1514	1484	1382	1378	1377	1307	1270	1255	1243	1099	1086	1056	1030

Les 20 trigrammes les plus fréquents

Trigrammes	ENT	LES	EDE	DES	QUE	AIT	LLE	SDE	ION	EME	ELA	RES	MEN	ESE	DEL	ANT	TIO	PAR	ESD	TDE
Nombres	900	801	630	609	607	542	509	508	477	472	437	432	425	416	404	397	383	360	351	350

Fig. 3.3 – Bigrammes et Trigrammes

– On peut également utiliser des homophones :
Il s'agit de remplacer une lettre non pas par un symbole unique, mais par un symbole choisi au hasard parmi plusieurs. Dans sa version la plus sophistiquée, on choisira un nombre des symboles proportionnel à la fréquence d'apparition de la lettre. Ainsi, on obtient un renversement des fréquences ce qui permet de faire disparaître complètement les indications fournies par la fréquence.
On parle de Substitution Homophonique ou de Substitution à représentation multiple.

Lettre	%	chiffre	Letter	%	chiffre
A	9	091233474853677892	N	7	18585966719199
B	1	81	O	5	0005075472
C	3	134162	P	3	389095
D	3	010345	Q	1	94
E	16	061014162324444654 55577479828798	R	6	293540427789
F	1	31	S	8	1119813676869697
G	1	26	T	7	17203043496975
H	1	39	U	6	020861638590
I	8	3250567073838893	V	2	3452
J	1	15	W	0	60
K	0	04	X	0	28
L	5	2637516584	Y	0	24
M	3	222768	Z	0	01

Fig. 3.4 – Exemple de table pour un chiffrement par homophones

1.3 Chiffre affine

On dit qu'une fonction est affine lorsqu'elle est de la forme $x \to a * x + b$, c'est-à-dire un polynôme de degré 1. Une fonction linéaire est une fonction affine particulière.
L'idée est d'utiliser comme fonction de chiffrement une fonction affine du type

$$y = (ax + b)\ mod\ 26,$$

Où a et b sont des constantes, et où x et y sont des nombres correspondant aux lettres de l'alphabet
(A=0,B=1,...). On peut remarquer que si a = 1, alors on retrouve le chiffre de César où b est le décalage (le k du chiffre de César).

Propriété de neutralité : si b = 0, alors "a" est toujours chiffré "A" car il ne subit aucun décalage.
En effet, si aucun décalage n'a lieu, l'alphabet de départ se retrouve chiffré par lui-même, et donc ne subit aucune modification.

Pour le chiffre affine, la clé est constituée de $(k1, k2)\ ou\ k1, k2\ \in [0,25]$ et telle que

$$pgcd(k1,26) = 1.$$

Le chiffrement en lui-même est donné par :

$$c_i = f(m_i) = k_1 * m_i + k_2\ \ mod\ 26\ .$$

Pour le déchiffrement, il vient

$$m_i = f^{-1}(c_i) = k_1^{-1} * (c_i - k_2)\ mod\ 26\ .$$

Par le chiffre affine, on obtient 312 clés possibles. En effet, pour respecter la propriété de k1, il n'y a que 12 choix possibles. Et puisque k2 peut prendre n'importe quelle valeur dans [0, 25], il vient 12 * 26 = 312.

Exemple : Soit la clé = (k1, k2) = (3, 11)

Transformation de chiffrement :

$$c_i = f(m_i) = 3 * m_i + 11\ mod\ 26$$

Transformation de déchiffrement :

$$k_1^{-1} = 3^{-1}\ mod\ 26\ = 9[3 * 9\ mod\ 26 = 1]$$
$$m_i = f^{-1}(c_i) = 9 * (c_i - 11) mod\ 26$$

Ainsi, pour une suite de lettres telle que $'NSA' \rightarrow 13\ 18\ 0 \rightarrow 24\ 13\ 11 \rightarrow 'YNL'$.

1.3.1 Cryptanalyse du chiffre affine

1. Il faut tout d'abord établir la fréquence relative de chaque lettre du texte chiffré, par analyse de fréquence.

HGAHY RAEFT GAGRH DGAGM OEHIY RAAOT ZGAGJ GKFDG AZGSB INNTG KGRHE
NNIRG

On dénombre 12 fois la lettre G et 8 fois la lettre A. Supposons que le langage original du texte est le français.

2. Sur base de l'analyse de fréquences, il faut ensuite dériver les équations correspondantes. E, A, S, I,N étant les lettres les plus fréquentes en français, on en déduit les équations suivantes (en émettant l'hypothèse que le bigramme ES est plus fréquent que EA) :

$$E \rightarrow G \Rightarrow f(E) = G\ ,\qquad S \rightarrow A \Rightarrow f(S) = A$$

Il en découle que :

$$4 \rightarrow 6 \Rightarrow f(4) = 6\ ,\qquad 18 \rightarrow 0 \Rightarrow f(18) = 0$$

3. On peut maintenant résoudre les équations pour que retrouver k1 et k2.
f(4) = 6, f(18) = 0
4 * k1 + k2 ≡ 6 (mod26)
18 * k1 + k2 ≡0 (mod26)
14k1 ≡ −6 mod 26
k1 = 7 ⇨ k2 = 4.

4. La fonction de déchiffrement est donc la suivante : mi = 15 _ (ci − 4) mod 26.
Et donc,

HGAHYRAEFTGAGRHDGAGMOEHIYRAAOTZGAGJGKFDGAZGSBINNTGKGRHE
NNIRG

devient

TESTONSAPRESENTLESEQUATIONSSURDESEXEMPLESDECHIFFREMENTAFFINE

2 Chiffrement polygraphique

Il s'agit ici de chiffrer un groupe de n lettres par un autre groupe de n symboles. On citera notamment le chiffre de Playfair et le chiffre de Hill. Ce type de chiffrement porte également le nom de *substitutions polygrammiques*.

2.1 Chiffre de Playfair (1854)

On chiffre 2 lettres par 2 autres. On procède donc par digramme. On dispose les 25 lettres de l'alphabet (W exclu car inutile à l'époque, on utilise V à la place) dans une grille de 5x5, ce qui donne la clef.
La variante anglaise consiste à garder le W et à fusionner I et J.

Il y a 4 règles à appliquer selon les deux lettres à chiffrer lors de l'étape de substitution. Pour le déchiffrement, on procède dans l'ordre inverse.

1. Si les lettres sont sur des "coins", les lettres chiffrées sont les 2 autres coins.
Exemple : OK devient VA, RE devient XI ...
2. Si les lettres sont sur la même ligne, il faut prendre les deux lettres qui les suivent immédiatement à leur droite.
3. Si les lettres sont sur la même colonne, il faut prendre les deux lettres qui les suivent immédiatement en dessous.
4. Si elles sont identiques, il faut insérer une nulle (habituellement le X) entre les deux pour éliminer ce doublon.
Exemple : "balloon" devient "ba" "lx" "lo" "on".

Pour former ces grilles de chiffrement, on utilise un mot-clef secret pour créer un alphabet désordonné avec lequel on remplit la grille ligne par ligne. Ensuite, on comble la grille avec les lettres restantes de l'alphabet.

Fig. 3.5 – Exemple du chiffre de Playfair

2.2 Chiffre de Hill (1929)

Les lettres sont d'abord remplacées par leur rang dans l'alphabet. Les lettres $P_k \; et \; P_{k+1}$ deviennent $C_k \; et \; C_{k+1}$

$$\begin{pmatrix} C_k \\ C_{k+1} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} P_k \\ P_{k+1} \end{pmatrix} (mod\ 26)$$

Les composantes de cette matrice doivent être des entiers positifs. De plus la matrice doit être inversible dansZ_{26}. Cependant, sa taille n'est pas fixée à 2. Elle grandira selon le nombre de lettres à chiffrer simultanément.

Chaque digramme clair ($P_1 \; et \; P_2$) sera chiffré ($C_1 \; et \; C_2$) selon :

$$C_1 \equiv aP_1 + bP_2 \; (mod\ 26)$$
$$C_2 \equiv cP_1 + dP_2 \; (mod\ 26)$$

Exemple de chiffrement : Alice prend comme clef de cryptage la matrice

$$\begin{pmatrix} 9 & 4 \\ 5 & 7 \end{pmatrix}$$

Pour chiffrer le message "je vous aime" qu'elle enverra à SA MAMAN. Après avoir remplacé les lettres par leur rang dans l'alphabet (a=1, b=2, etc.), elle obtiendra

$$C_1 \equiv 9 * 10 + 4 * 5 \; (mod\ 26) = 110 (mod\ 26) = 6$$
$$C_2 \equiv 5 * 10 + 7 * 5 \; (mod\ 26) = 85 \; (mod\ 26) \; = 7$$

Elle fera de même avec les 3e et 4e lettres, 5e et 6e, etc. Elle obtiendra finalement le résultat de la figure 3.6.

Lettres	**j**	**e**	**v**	**o**	**u**	**s**	**a**	**i**	**m**	**e**
Rangs ($\boldsymbol{P_k}$)	10	5	22	15	21	19	1	9	13	5
Rangs chiffrés ($\boldsymbol{C_k}$)	6	7	24	7	5	4	19	16	7	22
Lettres chiffrées	F	G	X	G	E	D	S	P	G	V

Fig. 3.6 – Exemple de chiffrement de Hill

Pour déchiffrer, le principe est le même que pour le chiffrement : on prend les lettres deux par deux, puis on les multiplie par une matrice

$$\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} (mod\ 26)$$

Cette matrice doit être l'inverse de matrice de chiffrement (modulo 26). Ordinairement, cet inverse est $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

Exemple de déchiffrement : Pour déchiffrer le message d'Alice, SA MAMAN doit calculer :

$$\begin{pmatrix} 9 & 4 \\ 5 & 7 \end{pmatrix}^{-1} = \frac{1}{43} \begin{pmatrix} 7 & -4 \\ -5 & 9 \end{pmatrix} (mod\ 26) = (43)^{-1} \begin{pmatrix} 7 & -4 \\ -5 & 9 \end{pmatrix} (mod 26)$$

Comme pgcd(43, 26) = 1, 43^{-1} existe dans Z_{26} et $(43)^{-1}\begin{pmatrix}7 & -4\\ -5 & 9\end{pmatrix}(mod\ 26)$. A la matrice de déchiffrement :

$$\begin{pmatrix}9 & 4\\ 5 & 7\end{pmatrix}^{-1} = 23\begin{pmatrix}7 & -4\\ -5 & 9\end{pmatrix}(mod26) = \begin{pmatrix}161 & -92\\ -115 & 207\end{pmatrix}(mod26) = \begin{pmatrix}5 & 12\\ 15 & 25\end{pmatrix}(mod26)$$

La maman d'Alice prend donc cette matrice pour déchiffrer le message "FGXGE DSPGV". Après avoir remplacé les lettres par leur rang dans l'alphabet (A=1, B=2, etc.), il obtiendra :

$$P_1 \equiv 5 * 6 + 12 * 7(mod26) = 114(mod26) = 10$$
$$P_2 \equiv 5 * 6 + 25 * 7(mod26) = 265(mod26) = 5$$

Il fera de même avec les 3e et 4e lettres, 5e et 6e, etc. Il obtiendra finalement le résultat de la figure 3.7

Lettres chiffrées	F	G	X	G	E	D	S	P	G	V
Rangs chiffrés (C_k)	6	7	24	7	5	4	19	16	7	22
Rangs (P_k)	10	5	22	15	21	19	1	9	13	5
Lettres	J	e	v	o	u	s	a	i	m	e

Fig. 3.7 – Exemple de déchiffrement de Hill

3 Substitutions polyalphabétiques

3.1 Chiffre de Vigenère (1568)

C'est une amélioration décisive du chiffre de César. Sa force réside dans l'utilisation non pas d'un, mais de 26 alphabets décalés pour chiffrer un message. On parle du carré de Vigenère. Ce chiffre utilise une clef qui définit le décalage pour chaque lettre du message (A : décalage de 0 cran, B : 1 cran, C : 2 crans, ..., Z : 25 crans).

Exemple : chiffrer le texte "CHIFFRE DE VIGENERE" avec la clef "BACHELIER" (cette clef est éventuellement répétée plusieurs fois pour être aussi longue que le texte clair)

Clair	**C**	**H**	**I**	**F**	**F**	**R**	**E**	**D**	**E**	**V**	**I**	**G**	**E**	**N**	**E**	**R**	**E**
Clef	B	A	C	H	E	L	I	E	R	B	A	C	H	E	L	I	E
Décalage	1	0	2	7	4	11	8	4	17	1	0	2	7	4	11	8	4
Chiffré	D	H	K	M	J	C	M	H	V	W	I	I	L	R	P	Z	I

Fig. 3.8 – Application du carré de Vigenère

La grande force du chiffre de Vigenère est que la même lettre sera chiffrée de différentes manières d'où perte de la fréquence des lettres, ce qui rend inutilisable l'analyse de fréquence classique. La figure 3.9 illustre cette perte des fréquences dans une fable de Lafontaine, codée par substitution simple et par Vigenère.

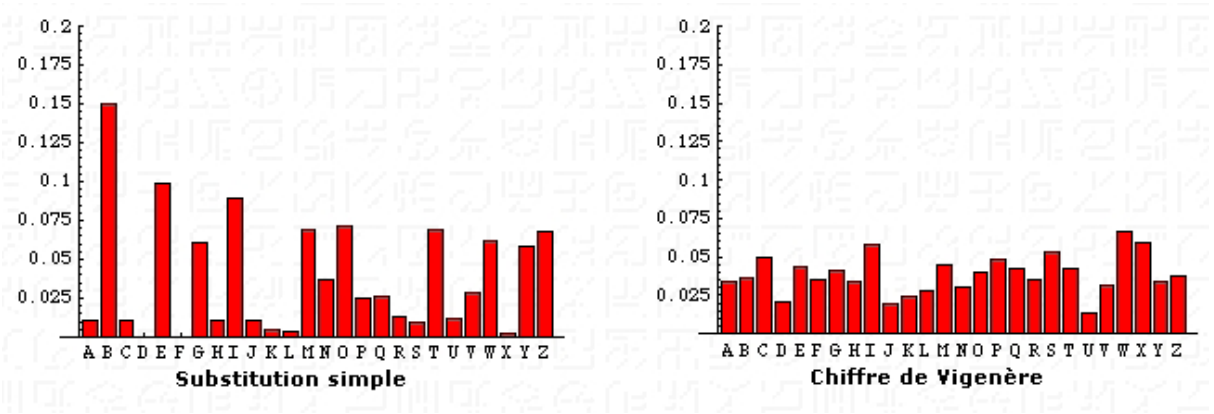


Fig. 3.9 – Perte de la fréquence des lettres

Pour utiliser le chiffrement de Vigenère, on a recours au Carré de Vigenère, illustré à la figure 3.10.

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

Fig. 3.10 – Le carré de Vigenère

La lettre de la clef est dans la colonne la plus à gauche, la lettre du message clair est dans la ligne tout en haut. La lettre chiffrée est à l'intersection des deux.
L'emploi du carré de Vigenère est souvent sujet à erreurs : la lecture en est pénible et, à la longue, fatigante. Beaucoup de cryptologues préfèrent se servir d'une "réglette", facile à construire, et d'un maniement plus rapide.

3.1.1 Cryptanalyse de Kasiski

Cette technique consiste à chercher des séquences de lettres qui apparaissent plus d'une fois dans le texte. En effet, dans ce cas, il n'y aura que deux possibilités :
– soit la même séquence de lettres du texte clair a été chiffrée avec la même partie de la clef,
– soit deux suites de lettres différentes dans le texte clair auraient par pure coïncidence engendré la même suite dans le texte chiffré (probabilité faible).

Exemple : Soit une séquence de 3 lettres répétée dans un message codé avec une distance d. On fait donc l'hypothèse sensée qu'il s'agit d'une même séquence de 3 lettres du texte initial, codée avec la même séquence de lettres de la clé. Par conséquent, si m est la longueur de la clé, pour que les 2 séquences soient codées avec les mêmes lettres de la clé, il est nécessaire que m divise d. m sera donc le pgcd des distances de séquences répétées. m représentera la longueur de la clé de chiffrement!

Il faut toutefois baser ce raisonnement sur les valeurs significatives du pgcd. Ainsi, dans l'exemple 3.11, les suites FCS et WTV ne seront pas prises en compte.

CS AZZMEQM,COXRWF,CSDZRMGFMJECV,X'IMOQJ JCLB NLFMK CCLBM WCCZBM KFIMSZJSZ CS URQ IUOU, CS ZLPIE ECZ RMWWTV, SB KCCJ QMJ FCSOVJ GCI ZICCKS, MK QMLL YL'CV ECCJ OKTFWTVM JIZ COXFWBIWVV, IV ACCICC C'OCKFM,JINWWB U'OBKSVUFM

Séquence	Position	Distance	Décomposition
COX	11-140	129	3.43
FCS	16-99	83	83
ZRM	20-83	63	$3^2 7$
FMJ	24-162	138	2.3.23
CLB	37-46	9	3^2
KCC	44-92	48	$2^3 3$
WTV	87-133	46	2.23
CCJ	93-126	33	3.11
ICC	110-155	45	$3^2.5$
MJI	136-163	27	3^3

Fig. 3.11 – Méthode de Kasiski

Ce renseignement est capital. Cela signifie que les caractères de rang 1, 4, 7, 10, ..., 3k+1, sont simplement décalés à la manière du chiffre de César. On peut donc appliquer maintenant l'analyse de fréquence à ces caractères et trouver la première lettre de la clef. Pour la deuxième lettre de la clef, on analysera les fréquences des caractères de rang 3k + 2 et pour le dernière lettre les fréquences des caractères de rang 3k.

3.1.2 Cryptanalyse de Friedman

Cette méthode utilise la notion d'Indice de Coïncidence (IC). Celui-ci est défini comme la probabilité que deux lettres choisies aléatoirement dans un texte soient identiques. Soient n le nombre de lettres dans le texte, n1 = nombre de A, ...,n26 = nombre de Z.
La probabilité de tirer deux A parmi les n lettres du texte est :

$$P(2\ fois\ A) = \frac{C_2^{n_1}}{C_2^n} = \frac{\frac{n_1(n_1-1)}{2}}{\frac{n(n-1)}{2}} = \frac{n_1(n_1-1)}{n(n-1)}$$

Plus généralement, la probabilité de tirer 2 lettres identiques est donnée par :

$$IC = \sum_{i=1}^{26} \frac{n_i\,(n_i-1)}{n(n-1)}$$

La figure 3.12 donne quelques indices calculés sur des textes contemporains dans différentes langues.

langue	allemand	anglais	espagnol	français	Italien	norvégien	Suédois
IC	0.072	0.065	0.074	0.074	0.075	0.073	0.071

Fig. 3.12 – IC des principales langues européennes

Plus n est grand, plus la probabilité est correcte.

Le test de Friedman a pour premier objectif de déterminer si un texte a été chiffré avec un chiffre monoalphabétique ou polyalphabétique. Comme second bénéfice, il suggère la longueur du mot-clef si le chiffre est polyalphabétique. Pour réaliser ces buts, le test de Friedman s'appuie sur l'IC.

Remarques importantes :
– Pour un langage de 26 lettres où chaque lettre a la même fréquence (1/26), l'IC vaut 0.038.
– Pour tout chiffre monoalphabétique, la distribution des fréquences est invariante, donc l'IC sera le même que pour le texte clair (respect des fréquences).
– Donc, si on applique ce test à un texte chiffré avec un chiffre monoalphabétique, on devrait trouver IC égal environ à 0.074 (en français). Si IC est beaucoup plus petit (p. ex. 0.050), le chiffre est probablement polyalphabétique.

Exemple de cryptanalyse par le test de Friedman
Soit le message suivant, supposé écrit en français, chiffré avec Vigenère (369 lettres) :

PERTQ UDCDJ XESCW MPNLV MIQDI ZTQFV XAKLR PICCP QSHZY DNCPW EAJWS ZGCLM
QNRDE OHCGE ZTQZY HELEW AUQFR OICWH QMYRR UFGBY QSEPV NEQCS EEQWE EAGDS ZDCWE OHYDW QERLM FTCCQ UNCPP QSKPY FEQOI OHGPR EERWI EFSDM XSYGE UELEH USNLV GPMFV EIVXS USJPW HIEYS NLCDW MCRTZ MICYX MNMFZ QASLZ QCJPY DSTTK ZEPZR ECMYW OICYG UESIU

GIRCE UTYTI ZTJPW HIEYI ETYYH USOFI XESCW HOGDM ZSNLV QSQPY JSCAV QSQLM QNRLP QSRLM XLCCG AMKPG QLYLY DAGEH GERCI RAGEI ZNMGI YBPP

On va considérer les sous-chaînes obtenues en prenant les lettres à intervalle donné :
– Intervalle de 1 : PERTQ UDCDJ XESCW MPNLV ... (texte original)
– Intervalle de 2 : PRQDD XSWPL ... et ETUCJ ECMNV ...
– Intervalle de 3 : PTDJS MLIIQ ..., EQCXC PVQZF... et RUDEW NMDTV
– ...
On calcule ensuite les IC pour toutes ces sous-chaînes, procédé illustré à la figure 3.13.

Intervalle	Indice de coïncidence
1	0.0456107
2	0.0476954, 0.0443098
3	0.044249, 0.0494469, 0.0426771
4	0.0465839, 0.0453894, 0.0449116, 0.0425227
5	0.0799704, 0.0925583, 0.0836727, 0.0795282, 0.0684932
6	0.0512956, 0.0407192, 0.0371585, 0.0382514, 0.0661202, 0.0431694

Fig. 3.13 – Indice de coïncidence des sous-chaines

On remarque que quand l'intervalle est de 5, l'IC correspond plus ou moins avec l'IC caractéristique du français (en tout cas, c'est cette ligne qui s'approche le plus de 0.074, les autres lignes étant plutôt proches de 0.038). La longueur de la clef utilisée est donc probablement 5.

D'autre part, si un message en français de longueur n et d'indice de coïncidence IC est chiffré avec un carré de Vigenère, alors on peut définir r, la longueur du mot-clef composé de lettres distinctes, qui est donné par la formule suivante

$$r \approx \frac{(0.074 - 0.0308)n}{(n-1)IC - 0.038n + 0.074} \approx \frac{0.036n}{(n-1)IC - 0.038n + 0.074}$$

En appliquant cette formule au texte précédent, on trouve r = 4.69, ce qui confirme ce que l'on avait trouvé ci-dessus.

3.2 Chiffre de Vernam (One Time Pad - 1917)

Le masque jetable est défini comme un chiffre de Vigenère avec la caractéristique que la clef de chiffrement a la même longueur que le message clair.

clair	**M**	**A**	**S**	**Q**	**U**	**E**	**J**	**E**	**T**	**A**	**B**	**L**	**E**
clef	X	C	A	A	T	E	L	P	R	V	G	Z	C
décalage	23	2	0	0	19	4	11	15	17	21	6	25	2
chiffré	J	C	S	Q	N	I	U	T	K	V	H	K	G

Fig. 3.14 – Exemple de One Time Pad

Pour utiliser ce chiffrement, il faut respecter plusieurs propriétés :
– choisir une clef aussi longue que le texte à chiffrer,

– utiliser une clef formée d'une suite de caractères aléatoires,
– protéger votre clef,
– ne jamais réutiliser une clef.

Exemple illustrant l'inviolabilité :

Soit le texte chiffré : cuskqxwmfwituk
Soit le masque jetable possible : bgfbcdfbfdecdg
Résultat : BONJOURLATERRE
Soit un autre masque jetable : quauwtedbdisjg
Résultat : MASQUESJETABLE

Il est donc impossible de déterminer le bon masque !

Le système du masque jetable, avec les précautions indiquées ci-dessus, est absolument inviolable si l'on ne connaît pas la clef. Il est couramment utilisé de nos jours par les États. En effet, ceux-ci peuvent communiquer les clefs à leurs ambassades de manière sûre via la valise diplomatique.
Le problème de ce système est de communiquer les clefs de chiffrement ou de trouver un algorithme de génération de clef commun aux deux partenaires.
De plus, la création de grandes quantités des clefs aléatoires devient vite problématique. N'importe quel système couramment utilisé pourrait exiger des millions de caractères aléatoires de façon régulière.
La distribution des clés est également complexe. La longueur de la clé étant égale à celle du message, une bonne organisation est nécessaire.

4 Transpositions

Elles consistent, par définition, à changer l'ordre des lettres. C'est un système simple, mais peu sûre pour de très brefs messages car il y a peu de variantes. Ainsi, un mot de trois lettres ne pourra être transposé que dans 6 (=3!) positions différentes. Par exemple, "col" ne peut se transformer qu'en "col", "clo", "ocl", "olc", "lco" et "loc".
Lorsque le nombre de lettres croît, il devient de plus en plus difficile de retrouver le texte original sans connaître le procédé de brouillage. Ainsi, une phrase de 35 lettres peut être disposée de 35! = 10^{40} manières différentes. Ce chiffrement nécessite un procédé rigoureux convenu auparavant entre les parties.
Une transposition rectangulaire consiste à écrire le message dans une grille rectangulaire, puis à arranger les colonnes de cette grille selon un mot de passe donné (le rang des lettres dans l'alphabet donne l'agencement des colonnes).

Exemple :

A la figure 3.15, on a choisi comme clef GRAIN pour chiffrer le message SALUT LES PETITS POTS. En remplissant la grille, on constate qu'il reste deux cases vides, que l'on peut remplir avec des nulles (ou pas, selon les désirs des correspondants).

G	R	A	I	N
2	5	1	3	4
S	A	L	U	T
L	E	S	P	E
T	I	T	S	P
O	T	S	(A)	(B)

A	G	I	N	R
1	2	3	4	5
L	S	U	T	A
S	L	P	E	E
T	T	S	P	I
S	O	(A)	(B)	T

Fig. 3.15 – Application d'une transposition

5 Machines à rotor

L'entre-deux-guerres voit le début de la mécanisation de la cryptographie. Des outils mécaniques, comme les cylindres chiffrants, sont mis à disposition des opérateurs, et des machines électromécaniques, sont mises au point. Ces machines fonctionnent sur le principe des rotors et des contacts électriques, afin de réaliser des formes de substitution polyalphabétique dont la clef a une longueur gigantesque de l'ordre de centaines de millions de lettres, au lieu de quelques dizaines dans les méthodes artisanales, comme le chiffre de vigenère. Le tableau 3.1 donne quelques systèmes électro-mécaniques. En supplément, on pourra citer les machines Purple (Japon), Nema (Suisse) et Sigaba (USA) pour les plus célèbres, mais il en existe encore beaucoup d'autres (machines de Hebern (USA), CCM (USA), variantes d'Enigma,...)

Origine	**Nom**	**Périodes des rotors**
USA	Hagelin M-209	101 405 850(26*25*23*21*19*17)
Allemagne	Enegma	17 576 $(26 * 26 * 26)$
Angleterre	Typex	26*(26-k)*26 avec $(k = 5; 7; 9)$
Pologne	Lacida	26 040 (24*31*35)

Tab. 3.1 – Quelques machines et leur période.

Fig. 3.16 – Différents systèmes électromécaniques

Fonctionement d’enigma comme exemple:

Le codage Enigma effectué par la machine Enigma est à la fois simple et astucieux. Chaque lettre est remplacée par une autre, et la substitution opérée change d'une lettre à l'autre. La machine Enigma est alimentée par une pile électrique qui alimente le circuit électrique constitué de plusieurs éléments en chaîne :

le tableau de connexions : il permet d'échanger des paires de l'alphabet, deux à deux, au moyen de fiches. Il y a 6 fiches qui permettent donc d'échanger 12 lettres. Un tableau de connections est donc une permutation très particulière où on a échangé au plus 6 paires. Par exemple, dans le tableau suivant (avec simplement 6 lettres), on a échangé A et C, D et F, tandis que B et E restent invariants.

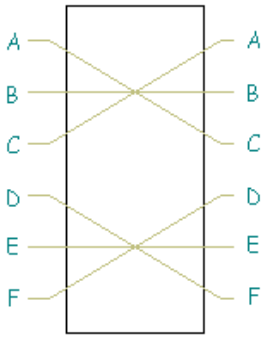
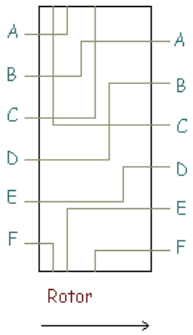


Tableau de connexions

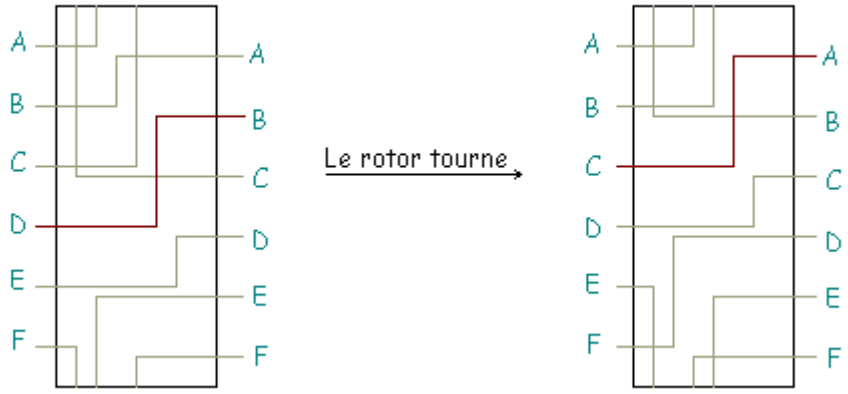
Entrée	Sortie
A	E
B	A
C	F
D	B
E	D
F	C

les rotors : un rotor est également une permutation, mais cette fois quelconque. A chaque lettre en entrée correspond une autre lettre.



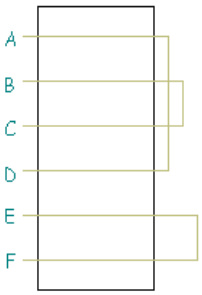
On peut composer les rotors, c'est-à-dire les mettre les uns à la suite des autres. La machine Enigma disposede 3 à 6 rotors. Parmi ces rotors, seuls 3 sont utilisés pour le codage, et on a le choix de les placer dans l'ordre que l'on souhaite (ce qui constituera une partie de la clé). les rotors sont cylindriques, et ils peuvent tourner autour de leur axe. Ainsi, à chaque fois qu'on a tapé une lettre, le premier rotor tourne d'un cran, et la permutation qu'il engendre est changée.

le rotor transforme initialement D en B. Lorsqu'il tourne d'un cran, cette liaison électrique D--->B se retrouve remontée en C--->A et, lorsque la prochaine lettre sera tapée, le rotor transformera cette fois D en C.



Chaque rotor possède donc 26 positions. A chaque fois qu'une lettre est tapée, le premier rotor tourne d'un cran. Après 26 lettres, il est revenu à sa position initiale, et le second rotor tourne alors d'un cran. On recommence à tourner le premier rotor, et ainsi de suite... Quand le second rotor a retrouvé sa position initiale, c'est le troisième rotor qui tourne d'un cran.

Le réflecteur : Au bout des 3 rotors se situe une dernière permutation qui permet de revenir en arrière. On permute une dernière fois les lettres 2 par 2, et on les fait retraverser les rotors, et le tableau de connexion.

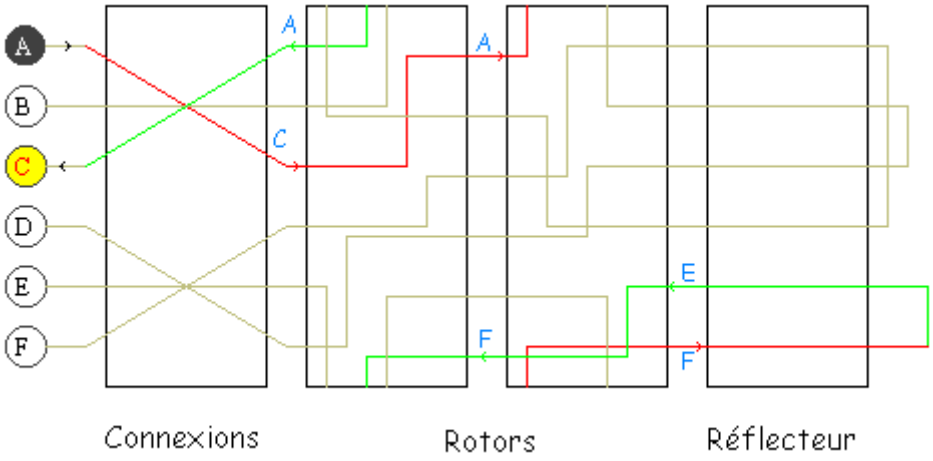


Le réflecteur

Résumons sur la machine simplifiée suivante (6 lettres, 2 rotors) comment est codée la lettre A :

- on traverse le tableau de connexions : on obtient C.
- on traverse les 2 rotors : on obtient successivement A et F.
- on traverse le réflecteur où on obtient E, puis on renvoie dans les rotors pour obtenir F, A et finalement C après le tableau de connexions.

Remarquons que si on avait tapé C, le courant aurait circulé dans l'autre sens et on aurait obtenu A.



Il y a trois éléments à connaitre pour pouvoir coder un message avec la machine Enigma.

- ✓ **<u>la position des 6 fiches du tableau de connexion</u> :** D'abord, il faut choisir 12 lettres parmi 26. C'est donc le nombre de combinaisons de 12 parmi 26, soit 26! /(12!14!). Maintenant, il faut choisir 6 paires de lettres parmi 12, soit 12!/6!, et comme la paire (A,D) donne la même connexion que la paire (B,A), il faut encore diviser par 2^6. On trouve finalement 100 391 791 500.
- ✓ **<u>l'ordre des rotors</u>** : il y a autant d'ordre que de façons d'ordonner 3 éléments : 3!=6.
- ✓ **<u>la position initiale des rotors :</u>** chaque rotor ayant 26 éléments, il y a 26*26*26=17576 choix.

Point forts : le nombre de clés énorme, et la réversibilité

L'une des failles de la machine Enigma est que jamais la lettre A ne sera codée par un A. Cela élimine un certain nombre de cas à inspecter.une des autres faiblesses dépend plutôt du protocole utilisé par les allemands : certains opérateurs (par exemple, ceux qui informaient de la météo) prenaient peu de précautions et commençaient toujours leurs messages par les mêmes mots (typiquement "Mon général..."). Les anglais connaissaient ainsi pour une partie du message à la fois le texte clair et le texte codé, ce qui aide à retrouver la clé

Conclusion

Bien que les progrès soient constants en cryptographie, les techniques ne cessant de se perfectionner et de se complexifier. La technologie, pour être précise l'informatique, permet désormais une fiabilité quasi-inviolable grâce au système de clefs publiques et clefs privées ou encore de clefs à usage unique. Cependant là encore avec l'évolution d'autres problèmes se posent, sachant que celle-ci est constante et ne cesse d'explorer des domaines de plus en plus complexes.