Chapitre 1
Historique

Ce chapitre ne prétend nullement donner une vision complète de l'histoire de la cryptologie. Il existe de très bons livres consacrés à ce sujet, notamment le livre de Simon Singh [1] ou la « bible » du domaine, l'ouvrage de David Kahn [2] (en anglais). Nous allons simplement décrire quelques exemples marquants, choisis soit parce que les idées qu'ils utilisent sont reprises dans des algorithmes récents, soit parce qu'ils illustrent l'évolution entre les méthodes naïves et les méthodes actuelles.

Dans ce qui suit, nous emploierons les termes *crypter* ou *chiffrer* pour désigner l'action consistant à masquer un message, et les termes *décrypter* ou *déchiffrer* pour l'opération inverse. La méthode permettant ce masquage sera appelée algorithme de cryptage ou de chiffrement, ou plus simplement cryptosystème. Le message initial sera appelé texte clair et le message masqué sera appelé texte crypté ou texte chiffré.

I. De l'Antiquité au Moyen Âge

Il est vraisemblable que la cryptologie soit apparue peu ou prou en même temps que l'écriture. Dès l'utilisation des messages écrits, le besoin de masquer certaines informations à transmettre s'est fait sentir. Nous avons trouvé, par exemple, des traces de cryptage dans des hiéroglyphes tracés par les Égyptiens antiques.

Il est également bien connu que Jules César cryptait ses messages importants. Sa méthode consistait à passer du texte clair au texte crypté en décalant les lettres de l'alphabet de trois places. Plus précisément, la correspondance entre les lettres du texte clair et celles du texte crypté par Jules César est donnée par le tableau ci-dessous :

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

Clair	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Crypté	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

et ainsi le texte :

JE VAIS ENVAHIR LA GAULE

devient

MH YDLV HQYDKLU OD JDXOH

Dans le vocabulaire moderne, ce cryptosystème s'appelle ROT3 (rotation de trois crans de l'alphabet). De façon analogue, on appelle ROT4, ROT5, etc., les méthodes qui consistent à décaler les lettres de 4 crans, 5 crans, etc.
L'inconvénient évident de cette méthode de cryptage par décalage est qu'elle est très simple à décrypter. Il suffit de tester les 26 possibilités de décalages pour y parvenir. Par conséquent, la sécurité de cette méthode s'effondre si nous savons qu'elle a été utilisée.
De nos jours, pour masquer des textes dont on veut éviter la lecture accidentelle, tels des solutions à des devinettes, il est courant d'utiliser ROT13. Pourquoi 13 ? Tout simplement parce que l'alphabet étant composé de 26 lettres, l'algorithme ROT13 permet de chiffrer **et** de déchiffrer.

Définition 1.1.

On appelle **substitution monoalphabétique** tout cryptosystème consistant à remplacer chaque lettre du message clair par une autre lettre ou par un symbole.

Bien évidemment, les chiffrements par décalages sont des exemples simples de substitutions monoalphabétiques.

L'ordre des Templiers chiffrait ses messages à l'aide d'une substitutions monoalphabétiques. Le schéma ci-dessous donne pour chaque lettre, le symbole qui lui correspond.

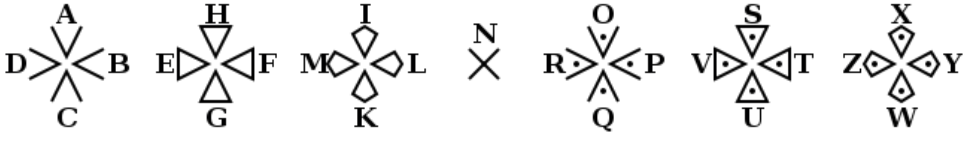

Fig. 1.1. Chiffre des Templiers

Cette disposition des symboles permet une mémorisation facile de cet alphabet, et facilite ainsi son utilisation. Si l'on se souvient des douze premiers symboles alors il est facile d'en déduire les douze derniers en ajoutant un point. Par ailleurs, la disposition des symboles évoque la croix de l'ordre du temple et ainsi, il suffit de mémoriser que le premier symbole est un chevron, le second un triangle et le troisième un « cerf-volant » pour retrouver aisément le schéma. Il ne reste plus qu'à se souvenir de l'ordre dans lequel sont écrites les lettres entourant les symboles...

Calculons le nombre de substitutions monoalphabétiques.
Il faut remplacer la lettre A par une des 26 lettres, donc il y a 26 choix possibles. Pour la lettre B, il nous reste donc 25 lettres ; pour C, plus que 24 ; etc. Au final, le nombre de façons de permuter les 26 lettres de l'alphabet est

$$26 \times 25 \times 24 \times \cdots \times 2 \times 1 = 26! \simeq 4 \times 10^{26}.$$

Il n'est alors maintenant plus question d'essayer toutes les possibilités, même avec un ordinateur ! Pour autant, choisir aléatoirement une substitution monoalphabétique nous permet-il vraiment d'obtenir une méthode de cryptage solide ?
La réponse à cette question est négative.
Si nous avons intercepté un message crypté de cette manière, et que nous ne connaissons pas la substitution employée, comment pouvons-nous faire pour déchiffrer ce message ? L'idée essentielle est d'utiliser les statistiques !
Tout joueur de scrabble sait que certaines lettres sont plus fréquentes que d'autres dans la langue française, mais dans quelle mesure exactement ? Pour le savoir, il a fallu compter les lettres dans de multiples textes publiés sur divers supports. Nous avons obtenu les résultats suivants (exprimés en pourcentage) :

E	:	17.76	O	:	5.34	B	:	0.80
S	:	8.23	D	:	3.60	H	:	0.64
A	:	7.68	C	:	3.32	X	:	0.54
N	:	7.61	P	:	3.24	Y	:	0.21
T	:	7.30	M	:	2.72	J	:	0.19
I	:	7.23	Q	:	1.34	Z	:	0.07
R	:	6.81	V	:	1.27	K	:	< 0.01
U	:	6.05	G	:	1.10	W	:	< 0.01
L	:	5.89	F	:	1.06			

La figure ci-dessous représente l'histogramme associé à ce tableau. Nous verrons par la suite qu'il est très utile.

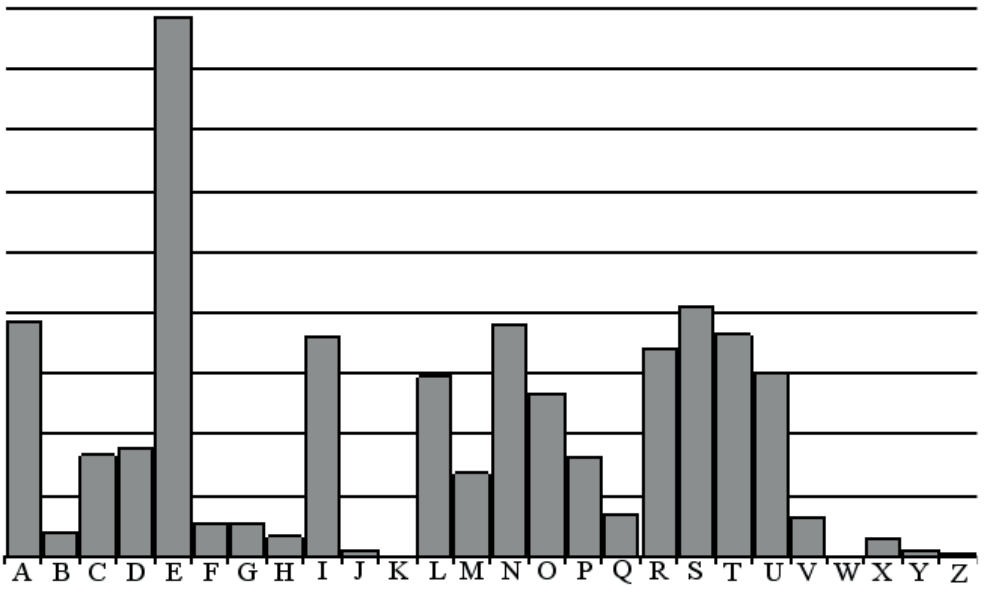

Fig. 1.2. Tableau des fréquences d'apparition des lettres en français

En comparant les données de ce tableau avec les fréquences d'apparition des lettres du message chiffré, nous pouvons en général identifier quel symbole correspond au « E » et quels sont les symboles probables pour représenter les lettres « S,A,N,T et I ». Puis, en identifiant les mots de liaison et les articles, nous pouvons généralement trouver le symbole pour les lettres, T, U, L et N. Nous pouvons également utilser le fait que les digrammes (groupes de deux lettres) les plus fréquents en français sont ES (4 %), EN (2,6 %), LE (2,2 %) et DE (2,2 %). Il est important de préciser que cela est spécifique au français. En anglais par exemple, les 4 digrammes les plus fréquents sont : HE, TH, IN et ER. Revenons au français, les digrammes fréquents formés de deux lettres identiques sont dans l'ordre : EE, LL, TT, NN, MM, RR, PP, FF, CC. Les trigrammes les plus fréquents, quant à eux, sont des terminaisons : ENT, AIT, ANT, ou des petits mots tels que : LES, QUE, DES et EST. Finalement, en devinant certains mots, nous trouvons le sens d'autres symboles. De cette manière, si le message codé est suffisament long, nous parvenons à le déchiffrer.

Cette méthode de cryptanalyse s'appelle l'analyse statistique de la fréquence d'apparition des lettres de l'alphabet. Au Moyen Âge, un érudit possédant ces connaissances (même de manière approximative) était capable de déchiffrer la plupart des messages cryptés.

Jusqu'au XVI^{e} siècle, la principale méthode de chiffrement était la substitution monoalphabétique. Cependant, une autre méhode était également utilisée : la transposition.

Définition 1.2.

On appelle **transposition** tout cryptosystème consistant à changer l'ordre des lettres du message clair pour obtenir le message chiffré.

Exemple 1.3. Une méthode de transposition classique consiste à écrire le message en lignes dans un tableau comme ci-dessous.

V	O	I	C	I
	U	N		E
X	E	M	P	L
E		D	E	
C	R	Y	P	T
A	G	E		

Nous obtenons alors le texte crypé en « lisant » ce tableau en colonnes. Le texte VOICI UN EXEMPLE DE CRYPTAGE donne ainsi V XECAOUE RGINMDYEC PEP IEL T.
Avec cette méthode, il suffit de se mettre d'accord sur le nombre de colonnes du tableau (ici 5), pour pouvoir s'échanger des messages. L'inconvénient majeur de cette méthode c'est que sa sécurité est réduite à néant si nous savons qu'elle a été employée. Il n'est, en effet, pas difficile d'essayer toutes les possibilités de taille pour le tableau et de trouver rapidement la bonne.

Quelle que soit la méthode utilisée pour mélanger les lettres, il n'est en général pas très difficile de les remettre dans le bon ordre. Donc les systèmes de chiffrement par transposition n'offrent que peu de résistance, pour peu que l'on essaie de les attaquer. Durant le Moyen Âge, la sécurité de ce système ne reposait que sur la croyance, irrationnelle mais répandue, qu'il était impossible de déchiffrer les textes cryptés. Par conséquent, la plupart des gens n'essayaient pas... Il est également possible de combiner une transposition et une susbstitution monoalphabétique mais cela ne renforce que peu la sécurité. Finalement, jusqu'au XIX^{e} siècle, l'avantage était clairement du côté de la cryptanalyse. Il n'existait pas de moyen pour réellement protéger les informations contenues dans un message, jusqu'à l'invention du chiffrement de Vigenère.

II. Le chiffrement de Vigenère

II.1. Description

Le diplomate français Blaise de Vigenère a introduit en 1586 une nouvelle manière de crypter les messages, rendant l'analyse statistique de la fréquence d'apparition des lettres inefficace.

Blaise de Vigenère (1523-1596)

Son idée essentielle, pour atteindre ce but, est qu'il faut que le chiffrement d'une lettre soit variable suivant la position de cette lettre dans le message clair. Par exemple, la lettre A pourra être remplacée par P si elle est en début de message et par F si elle est en deuxième position. On appelle ce type de chiffrement une **substitution polyalphabétique**.

Cette idée est, certes, brillante mais il reste à trouver une façon de la mettre en application. Voici la méthode proposée par Vigenère :
L'expéditeur et le destinataire du message se mettent d'accord sur un mot-clé, par exemple CHAT. À chaque lettre de ce mot nous associons son rang dans l'alphabet et nous noterons U_k le rang de la k-ième lettre du mot-clé diminué de 1. Nous prolongeons alors la suite (U_n) de façon périodique. Avec le mot CHAT, nous obtenons $U_1 = 2$ (car C est la 3^e lettre de l'alphabet), $U_2 = 7$ (car H est la 8^e lettre de l'alphabet), $U_3 = 0$ et $U_4 = 19$. Nous sommes arrivés à la fin de notre mot-clé, donc nous continuons la suite en répétant le motif initial, ce qui donne $U_5 = 2$, $U_6 = 7$ et ainsi de suite... La suite (U_n) est donc très facile à calculer pour peu que nous connaissions le mot-clé.

Pour crypter le message clair, nous décalons la première lettre de U_1 crans, la seconde de U_2 crans et ainsi de suite, la $n-$ième case est décalée de U_n crans.

Essayons de crypter le texte : « Voici un exemple » avec le mot-clé CHAT. Nous décalons la première lettre : V de 2 crans, nous obtenons X. Ensuite nous devons décaler O de 7 crans et nous obtenons V. La troisième lettre n'est pas décalé donc le I reste I. Ainsi de suite, nous obtenons le résultat suivant :

Clair	V	O	I	C	I		U	N		E	X	E	M	P	L	E
Décalage	2	7	0	19	2		7	0		19	2	7	0	19	2	7
Crypté	X	V	I	V	K		B	N		X	Z	L	M	I	N	L

Pour déchiffrer, il suffit d'appliquer les mêmes décalages mais dans le sens inverse. Par exemple, pour la première lettre X, on la décale de deux crans vers l'arrière et on retrouve V.
Examinons le texte crypté ci-dessus, nous remarquons que les deux I du premier mot sont cryptés par deux lettres différentes et que les deux X du texte crypté ne correspondent pas à la même lettre du texte clair.

Comme cela était souhaité, la méthode d'analyse statistique décrite dans le paragraphe précédent ne permet plus de déchiffrer les messages cryptés avec ce système.

Pour coder un texte en utilisant cette méthode, il est commode d'utiliser la table de Vigenère présentée ci-après. On procède de la façon suivante :

– Tout d'abord, il faut repérer la lettre à coder dans le première ligne, cela nous indique une colonne du tableau. Dans notre exemple, nous commençons à crypter le lettre V, cela nous donne la 22^e colonne.

– Ensuite, il faut repérer la lettre du mot-clé correspondante dans la première colonne. Pour nous, il s'agit de la lettre C qui nous indique donc la troisième ligne.

– Il suffit ensuite de lire le contenu de la case du tableau se trouvant à l'intersection de la colonne et de la ligne que nous avons repérés dans les étapes précédentes pour avoir la lettre du texte crypté. Ainsi, à l'intersection de la 22^e colonne et de la 3^e ligne, nous lisons la lettre X.

Pour décrypter, il suffit de trouver le lettre du message chiffré dans la ligne débutant par la lettre du mot-clé. Dans notre exemple, il faut repérer le X dans la troisième ligne. La lettre se trouvant alors en haut de la colonne correspondante est la lettre du message clair.

Table de Vigenère :

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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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

II.2. Cryptanalyse

Cette méthode de cryptage très ingénieuse est restée invaincue près de 300 ans. Cela explique pourquoi le chiffre de Vigenère fut surnommé : « le chiffre indéchiffrable ». C'est vraisemblablement le mathématicien anglais Charles Babbage qui, le premier, a trouvé une méthode d'attaque. Mais il ne l'a pas rendue publique. Le premier à avoir publié une méthode, en 1863, est l'officier prussien Friedrich Wilhelm Kasiski (1805-1881). Nous allons décrire brièvement cette méthode.