Gilles Bailly-Maitre

ARITHMÉTIQUE ET CRYPTOLOGIE

2e édition

ellipses

Références sciences

Arithmétique et cryptologie

2e édition

Gilles Bailly-Maitre

ellipses

Collection Références sciences

dirigée par Paul de Laboulaye
paul.delaboulaye@editions-ellipses.fr

Retrouvez tous les livres de la collection et des extraits sur www.editions-ellipses.fr

ISBN 9782340-046191

32, rue Bargue 75740 Paris cedex 15

www.editions-ellipses.fr

à mes zoupzounets : Antoine, Colin et Hélène...

Introduction

La cryptologie est étymologiquement la science du secret. Cette science regroupe plusieurs disciplines. La cryptographie, qui est l'art de masquer des messages afin que seuls les destinataires légitimes puissent les déchiffrer, en est une. Mais ce n'est pas la seule. Citons par exemple la cryptanalyse, qui est l'étude des méthodes de chiffrements dans le but de les rendre inutiles.
Si certaines méthodes de cryptologie sont utilisées depuis l'Antiquité, la recherche scientifique dans ce domaine est assez récente. L'essentiel des découvertes a eu lieu au cours du XX^{e} siècle. Par exemple, la cryptographie à clé publique est un immense domaine de recherche qui a été initié au début des années 70.
Depuis quelques décennies, la cryptologie s'est étendue à d'autres sujets tels la signature électronique, l'identification ou l'authentification. Ces nouvelles possibilités sont probablement plus utiles que de chiffrer des messages. Tout un chacun en use quotidiennement sans même s'en rendre compte. Par exemple, lorsque l'on utilise une carte bancaire pour effectuer une transaction, plusieurs algorithmes d'authentification sont effectués entre le terminal et la puce de la carte pour que celle-ci « prouve » son authenticité.
Le but de ce livre est d'expliquer les méthodes les plus couramment utilisées actuellement pour effectuer les principales tâches cryptologiques. Le point important est qu'il s'agit bien d'expliquer et non pas seulement de décrire ces méthodes.
Par exemple, un groupe international de chercheurs ont décidé d'une méthode standard de chiffrement (de cryptage) en novembre 2001, celle-ci s'appelle **AES** (initiales de « Advanced Encryption Standard »). Il existe de multiples sources où l'on peut trouver une description détaillée de cet algorithme, notre but ici est de tenter de comprendre la logique ayant présidé à sa construction. Grâce à des considérations historiques, informatiques et mathématiques, nous allons percer à jour les raisons des différentes étapes. En fin de compte, cette succession de calculs *a priori* très mystérieuse va devenir très logique.
La cryptologie moderne s'appuie sur des notions mathématiques avancées (telle les courbes elliptiques que l'on décrira succintement) portant sur plusieurs domaines. On va se limiter ici à l'étude des notions d'arithmétique nécessaires en cryptologie. L'arithmétique ayant le double avantage d'être accessible aux étudiants dès la première année universitaire et d'être à la base des principales méthodes de cryptologie actuelles.

La première partie de cet ouvrage présente une brève histoire de la cryptologie. Les étapes les plus marquantes sont présentées afin de comprendre l'évolution

des idées dans ce domaine. Nous verrons qu'il existe des principes intemporels et que certains concepts anciens sont encore utilisés dans les méthodes les plus modernes.

La deuxième partie est une longue parenthèse mathématiques où sont développées les notions et propriétés utiles pour comprendre la cryptologie moderne. Cette partie peut être lue comme un cours d'algèbre de base et d'arithmétique.

La dernière partie poursuit la description des principales méthodes de cryptologie, en se concentrant sur celles développées à l'ère informatique. Les grandes découvertes de la fin du XX^{e} siècle y sont décrites en détail. Les différentes tâches de la cryptologie actuelles et les moyens de les accomplir sont également présentés.
Cette seconde édition permet de corriger quelques erreurs présentes dans la première édition, de mettre à jour les données et compléter diverses informations mais également de parler des nouveautés de la dernière décennie notamment des ordinateurs quantiques et de la blockchain.

Je tiens à remercier particulièrement Jacques Patarin, de l'université de Versailles, dont l'enthousiasme communicatif a largement contribué à mon goût pour la cryptologie.
Un grand merci à tous ceux qui ont contribué à l'amélioration de ce livre, par sa relecture et des suggestions pertinentes : Geneviève Moebs, Jérome Trosset dit « Adrameleck », Louis Rivoallan, Frédéric Manconi, Noël Fraissex, Jean-Philippe Furter, Étienne Matheron, Xavier Roart dit « Goto », sans oublier papy Jean-Paul et ma moman chérie.
Je tiens tout particulièrement remercier Axel Cypel pour sa lecture minutieuse de la première édition qui m'a grandement aidé pour rédiger le présent volume ! Merci aux relecteurs de la seconde édition : Texas Granger, ProfRogue, Key, FreshPineapple, Valoukanga, Matuidi Chariot et bien sûr un énorme merci à Blanche Heisler pour ses commentaires judicieux et son enthousiasme.
Il va sans dire que ma reconnaissance est immense envers mon épouse qui m'a supporté et soutenu pendant la préparation de cet ouvrage.

Table des matières

Première partie

Cryptologie à l'ancienne

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.

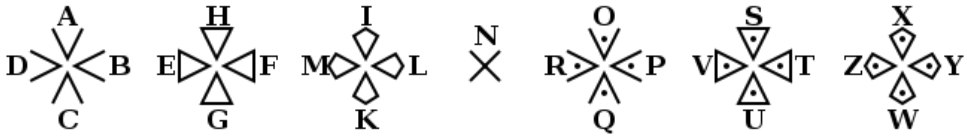

Fig. 1.1. Chiffre des Templiers

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.

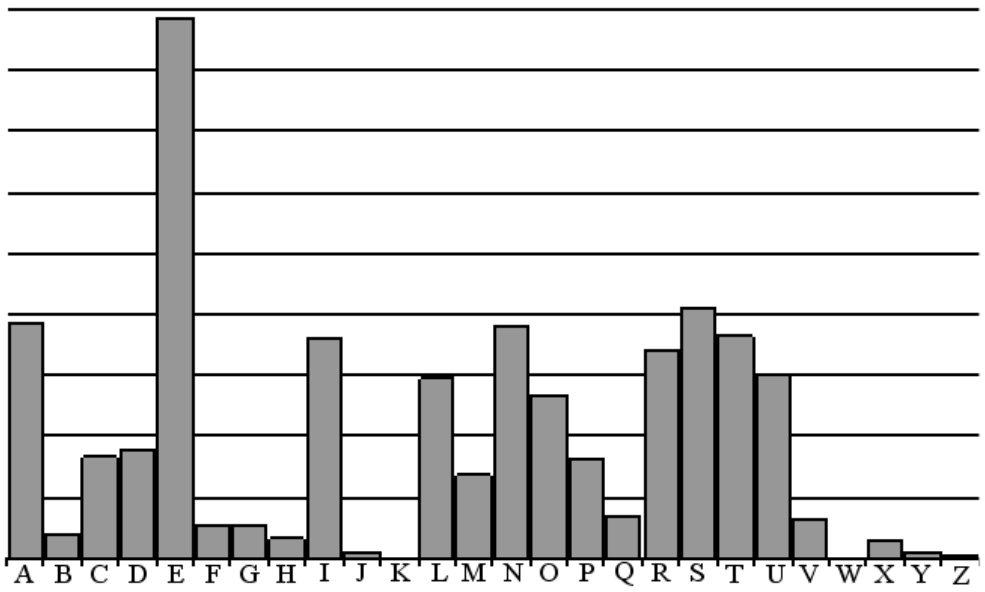

Fig. 1.2. Tableau des fréquences d'apparition des lettres en français

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.

Charles Babbage (1791-1871)

Il faut, tout d'abord, trouver la longueur du mot-clé. Pour cela, nous cherchons des séquences de 3 lettres ou plus se répétant dans le texte chiffré. Imaginons, par exemple, que nous trouvions dans le texte crypté la séquence ABC deux fois, les deux occurrences étant séparées par 35 lettres. Il est possible que deux triplets de lettres différentes cryptés différemment donnent la même chose, en l'occurrence ABC. Mais il est plus probable que ce soit la même suite de lettres (par exemple EST) qui ait été cryptée de la même façon. Si c'est le cas, le nombre de lettres séparant la répétition est un multiple de la longueur du mot-clé. Dans notre exemple, nous pouvons en déduire que la longueur du mot-clé est 5 ou 7. En trouvant plusieurs répétitions et en calculant le plus grand diviseur commun des distances les séparant, nous pouvons obtenir la longueur du mot-clé.

Imaginons par exemple, qu'avec cette méthode, nous ayons découvert que la longueur du mot-clé est 4.
Nous formons alors un premier texte, que nous appellerons texte-test, à partir du message chiffré en prenant sa première lettre puis sa 5^e puis sa 9^e lettre et ainsi de suite en prenant une lettre sur 4. Toutes les lettres du texte-test ainsi obtenues ont été codées par le même décalage.

Nous procédons alors à l'analyse statistique du texte-test. Notons que la fréquence d'apparition des lettres peut être notablement différente d'un texte standard, car celui-ci est issu par décalage d'un texte qui n'a aucun sens.
Néanmoins, dressons le tableau des fréquences d'apparition des lettres. Représentons l'histogramme et comparons-le à celui obtenu grâce au tableau des fréquences moyennes en français.

Imaginons, par exemple, que nous obtenions l'histogramme de la figure 1.3.

En comparant cet histogramme avec celui de la figure 1.2, nous nous rendons facilement compte qu'il y a eu un décalage de deux crans et que, donc, la première lettre du mot-clé est C. De manière générale, nous devinons la lettre du mot-clé en faisant coller le mieux possible les deux tableaux.

Ensuite, nous constituons un deuxième texte-test en prenant une lettre sur 4 du texte chiffré, mais en partant cette fois de la deuxième lettre. Par la même méthode, nous allons trouver la seconde lettre du mot-clé.

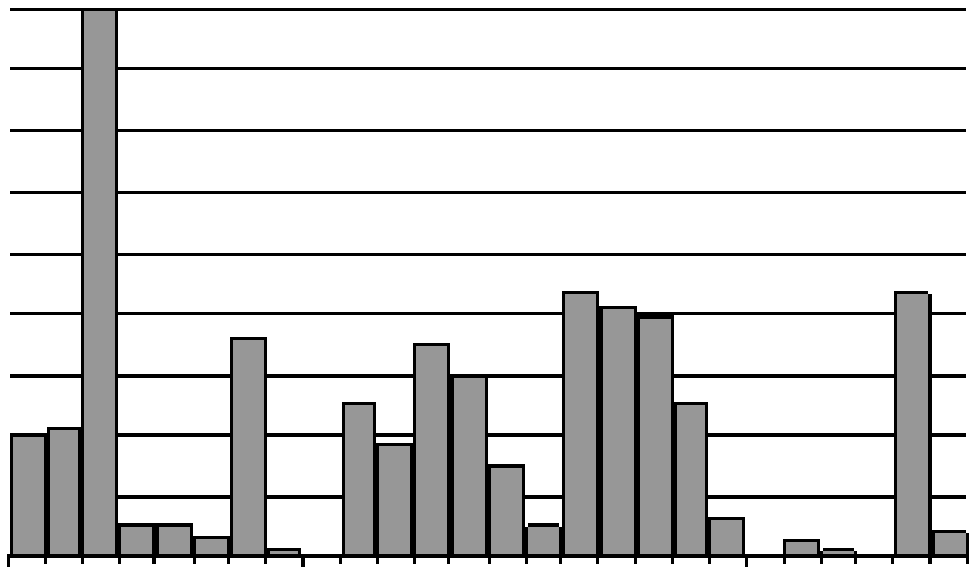

Fig. 1.3. Texte testé

De cette façon, nous allons révéler toutes les lettres du mot-clé et décrypter le message.
Il est tentant de croire que nous pouvons ainsi venir à bout de toutes les substitutions polyalphabétiques, mais il n'en est rien. Il est même possible de construire un cryptosystème de ce type qu'il est impossible d'attaquer. Décrivons-le sans plus attendre.

III. Le one-time pad ou masque jetable

Le cryptosystème appelé **one-time pad** ou **masque jetable** fut inventé par l'ingénieur en télécommunications américain Gilbert Vernam en 1917. Il consiste simplement à utiliser le chiffrement de Vigenère de telle sorte que l'attaque décrite précédemment ne puisse plus fonctionner. Pour cela, nous allons choisir un mot-clé (que nous appellerons aussi masque) dont le nombre de lettres sera égal au nombre de lettres du messsage à crypter. Par conséquent, tout le monde connaît la longueur du mot-clé. Mais la deuxième phase de l'attaque précédente nous amènerait à faire l'analyse statistique de messages composés d'une seule lettre !
De plus, pour éviter qu'un attaquant ne devine le mot-clé, il est construit en tirant au sort chacune de ses lettres. Il est ainsi totalement aléatoire.

Avec ces deux précautions, nous pouvons prouver qu'il est impossible de retrouver le message clair à partir du message crypté sans connaître le mot-clé, même avec une puissance de calcul infinie.
En effet, supposons que nous recevions le message codé WXYZ, tous les mots de 4 lettres peuvent donner ce résultat suivant la clé choisie. Par exemple, si la clé est UQYG, alors le texte clair est CHAT ; mais si la clé est LJEK, alors le texte clair est LOUP. Si nous ne connaissons rien de la clé, nous ne connaissons rien du texte clair.
Décrivons maintenant l'utilisation moderne du one-time pad. Nous allons en profiter pour introduire quelques notations utiles pour la suite.
Les textes que nous voulons crypter sont aisément transformés en une succession de bits (0 ou 1). C'est le cas dès que nous utilisons un ordinateur. Nous supposerons donc qu'un message est un élément de $I_n = \{0,1\}^n$, c'est-à-dire une chaîne formée de n bits. Nous générons de manière aléatoire une autre suite de n bits appelée K.
Le message crypté M' est obtenu par

$$M' = M \oplus K$$

où $\oplus$ est l'opération appelée XOR qui consiste à ajouter les bits deux à deux avec la convention $1 + 1 = 0$. Cette opération et ses propriétés sont décrites plus en détail dans le chapitre 8. Donnons juste un exemple de calcul

$$\begin{array}{rcccc} & 0 & 1 & 1 & 0 \\ \oplus & 1 & 1 & 0 & 0 \\ \hline & 1 & 0 & 1 & 0 \end{array}.$$

Donc, si $M = 0110$ et $K = 1100$, alors $M' = 1010$.
En pratique, cette méthode est contraignante car nous ne pouvons pas utiliser plusieurs fois le même masque pour chiffrer plusieurs messages, d'où le terme de masque « jetable ». En effet, imaginons que nous utilisions le même masque K pour crypter deux messages M_1 et M_2, notons M'_1 et M'_2 les messages obtenus. Nous avons

$$\begin{aligned} M'_1 \oplus M'_2 &= M_1 \oplus K \oplus M_2 \oplus K \\ &= M_1 \oplus M_2 \end{aligned}$$

car $K \oplus K$ n'est composé que de zéros. Donc, la connaissance des messages cryptés donne des informations sur les messages clairs. Par exemple, si quelqu'un finit par apprendre M_1, alors il peut en déduire immédiatement M_2.
De ce fait, pour pouvoir utiliser ce cryptosystème régulièrement, il faut en permanence générer des chaînes aléatoires de bits et les transmettre de façon sûre

au destinataire des messages, ce qui pose beaucoup de problèmes pratiques. Par exemple, la sécurité des communications transitant sur la ligne directe reliant les présidents américain et russe (le fameux « téléphone rouge », qui est en réalité un fax) est assurée par le one-time pad. Cela implique des voyages réguliers d'agents spéciaux entre Washington et Moscou avec gardes du corps et mallette attachée au poignet, malette contenant des pages couvertes de 0 et de 1.
Une autre difficulté est de générer des suites de bits réellement aléatoires en grandes quantités.
Pour toutes ces raisons, nous comprenons bien pourquoi le commun des mortels ne peut pas utiliser cette méthode et pourquoi la cryptologie reste un domaine de recherche très actif.

IV. La machine Enigma

Cette invention est conçue pour permettre de chiffrer de grosses quantités de messages de façon pratique tout en essayant d'obtenir une sécurité très importante, se rapprochant autant que possible de la sécurité du masque jetable.
Nous allons brièvement décrire son fonctionnement. Le lecteur intéressé pourra consulter à profit les multiples sites qui en parlent dont Ars Cryptographica (http ://www.apprendre-en-ligne.net/crypto/).
Enigma a été inventée par Arthur Scherbius (1878-1929), ingénieur en électricité allemand, qui en a commencé la commercialisation au début des années 20. Elle doit sa notoriété au fait qu'elle fut massivement utilisée par les armées allemandes lors de la seconde guerre mondiale. Elle se présente comme une grosse machine à écrire munie d'un panneau lumineux représentant les différentes lettres de l'alphabet.

La machine Enigma

L'innovation essentielle apportée par cette machine est la mécanisation du cryptage. Cela permet de crypter des messages bien plus nombreux et de façon

bien plus rapide, qu'en le faisant « à la main ». Grâce à un mécanisme ingénieux que nous allons décrire de façon schématique ci-dessous, Enigma permet d'utiliser facilement un cryptosystème de Vigenère avec un très long mot-clé (plus long que la plupart des messages à envoyer). Par conséquent, il semble *a priori* impossible de déchiffrer les textes ainsi cryptés. L'autre avantage de cette invention, c'est qu'elle est conçue pour être très commode à utiliser.
Le principe de fonctionnement est simple : lorsque nous pressons une touche du clavier d'Enigma, nous fermons un circuit électrique, ce qui allume une lettre sur le panneau d'affichage qui donne la lettre correspondante du texte crypté. Lorsque nous relâchons cette touche, un mécanisme modifie la structure du circuit électrique interne.
Cette prouesse est réalisée de la façon suivante : sur le trajet électrique entre les touches du clavier et le panneau d'affichage se trouvent des composants appelés « rotors ». Un rotor comporte 26 connecteurs en entrée (un pour chaque lettre) et 26 connecteurs en sortie qui sont reliés deux à deux de façon à mélanger les lettres. De plus, ces connecteurs sont disposés de façon circulaire, de part et d'autre du rotor. La figure ci-dessous représente un rotor démonté :

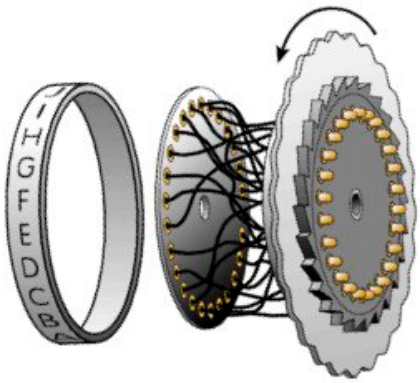

Fig. 1.4. Un rotor de la machine Enigma

Relâcher une touche du clavier provoque une rotation d'un rotor d'un cran ce qui modife le schéma électrique. S'il n'y avait qu'un rotor, alors, au bout de 26 lettres, nous retrouverions le schéma électrique initial et donc nous obtiendrions un chiffrement de Vigenère avec un mot-clé de 26 lettres de long, ce qui est bien insuffisant. Sur le modèle le plus courant de la machine Enigma, il y a trois rotors, le premier se décalant d'un cran à chaque frappe de touche, le second se décalant d'un cran lorsque le premier finit un tour et le troisième se décalant d'un cran lorsque le deuxième finit un tour. De ce fait, avant de

retrouver le schéma électrique initial, il faut frapper $26 \times 26 \times 26 = 17\,576$ fois sur le clavier. Enigma utilise donc un chiffrement de Vigenère avec un mot-clé formé de $17\,576$ lettres.
La faiblesse de ce dispositif est que le mot-clé obtenu n'est pas du tout aléatoire, il suffit même de démonter un exemplaire d'Enigma et d'examiner les rotors pour le connaître. Comme il était évident que, tôt ou tard, l'ennemi serait en possession d'une machine Enigma, celle-ci a été conçue pour éviter ce problème. Voici les moyens qui ont été mis en oeuvre dans ce but :

– La position initiale des rotors est réglable manuellement. Les lettres de l'alphabet sont gravées sur le pourtour des rotors pour faciliter le réglage. Par exemple, si deux personnes se mettent d'accord sur le réglage OUF, cela signifie qu'avant de crypter il faut régler le premier rotor sur la position initiale : « O », le second sur la position « U » et le troisième sur la position « F ». Cela permet en fait d'avoir 17576 réglages, et donc 17576 mots-clés.

– Il est possible d'échanger la place des rotors au sein d'Enigma. Par exemple, deux personnes se mettront d'accord sur le placement $1 - 3 - 2$ pour dire que le rotor 1 est en première position, puis que le rotor 3 est en seconde place et le rotor 2 en troisième place.
Comme il y a 6 façons de ranger les 3 rotors, cela permet de multiplier par 6 le nombre de mots-clés.

– Un tableau de connexions (représenté ci-contre), situé entre les touches du clavier et les rotors est ajouté. Celui-ci permet aux opérateurs de relier des couples de lettres entre elles.
Par exemple, si l'opérateur a décidé de connecter la lettre A avec la lettre J, une pression sur la touche A allumera la lettre du panneau lumineux correspondant initialement à J et réciproquement. Tout va se passer comme si les lettres A et J avaient été échangées dans le message initial.
Le nombre de façons de choisir un couple de lettres parmi 26 est $\frac{26\times25}{2}$.
En pratique, les opérateurs sont munis de 6 cables pour relier 6 couples de lettres (dans la dernière version d'Enigma, ce nombre passera à 10).
Cette modification va encore ajouter des informations sur lesquelles les utilisateurs doivent se mettre d'accord pour pouvoir communiquer. Il faut, en effet, y ajouter 6 couples de lettres pour indiquer les branchements du tableau de connexion par exemple : $AJ - SO - KR - IZ - EL - DT$.
Il y a un grand nombre de façons de choisir les connexions puisqu'il y en a

$$\frac{26 \times 25}{2} \times \frac{24 \times 23}{2} \times \frac{22 \times 21}{2} \times \frac{20 \times 19}{2} \times \frac{18 \times 17}{2} \times \frac{16 \times 15}{2}$$
$$= 72\,282\,089\,880\,000$$

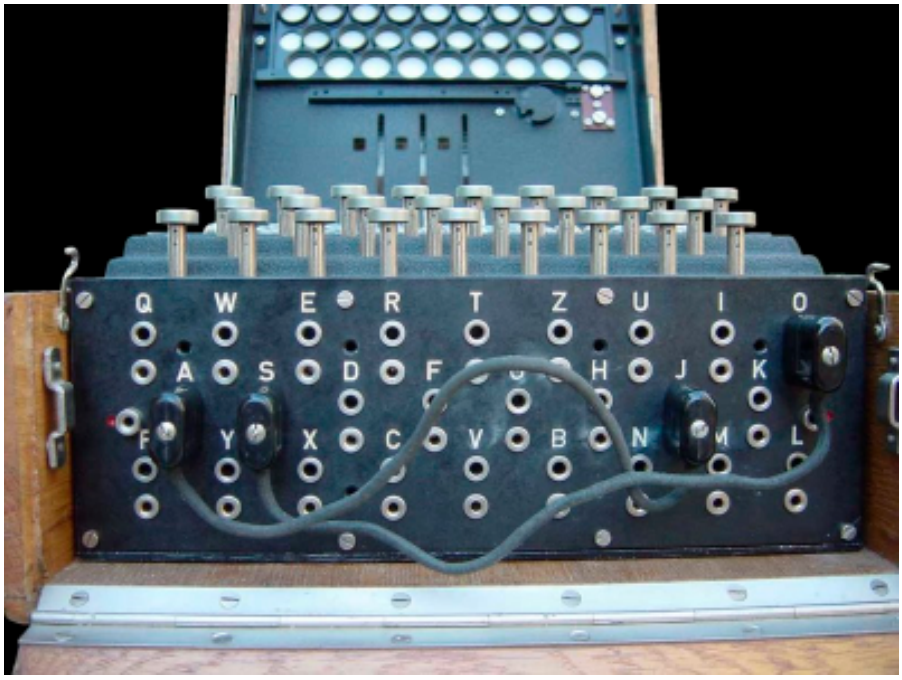

Fig. 1.5. Panneau de connexions de la machine Enigma. Sur cette photo, deux couples de lettres sont reliés : A – J **et** S – O.

Au final, une clé secrète de la machine Enigma est de la forme

$$\begin{cases} \mathsf{AJ} - \mathsf{SO} - \mathsf{KR} - \mathsf{IZ} - \mathsf{EL} - \mathsf{DT} \\ 1 - 3 - 2 \\ \mathsf{OUF} \end{cases}$$

Le nombre de clés possibles est

$$72\,282\,089\,880\,000 \times 6 \times 17576 \simeq 7,6 \times 10^{18}.$$

Pour résumer, Enigma permet très facilement d'utiliser un chiffrement de Vigenère avec un mot-clé long de 17576 lettres, choisi parmi $7,6 \times 10^{18}$ possibilités. Comme ce nombre est gigantesque, il semble de manière intuitive que la sécurité obtenue est proche de celle du one-time pad. L'histoire a pourtant prouvé le contraire.
La raison essentielle de la faiblesse d'Enigma est, paradoxalement, le nombre de mots-clés, qui est finalement bien insuffisant ! En effet, par rapport au nombre total de mots de 17576 lettres de long qui est $26^{1756} \simeq 4,9340 \times 10^{2484}$, le nombre $7,6 \times 10^{18}$ est infime.

Néanmoins, la cryptanalyse d'Enigma a nécessité des années de travail par des mathématiciens brillants, notamment le britannique Alan Turing (1912-1954). Cette cryptanalyse est passionnante et le lecteur est vivement invité à la découvrir, par les moyens déjà cités (livre ou web). Pour y parvenir, il a fallu développer des trésors d'ingéniosité afin de réduire le nombre de clés à tester mais aussi construire des machines toujours plus sophistiquées pour tester rapidement toutes ces clés. La réponse à la mécanisation du codage fut la mécanisation du décryptage qui a conduit aux débuts de l'informatique.

V. Et après...

L'Enigma était une superbe utilisation de substitution polyalphabétique, mais celle-ci échoua. Il fut inventé par la suite des machines améliorées, telle SIGABA capable de générer un nombre de mots-clés bien plus considérable grâce à un ruban perforé qui définissait une marche des rotors variable d'un message à l'autre. Mais bien que SIGABA ne pût jamais être attaquée, l'échange des rubans était fastidieux et le gain en sécurité se faisait au détriment de la facilité d'utilisation.
L'échec d'Enigma conduisit à l'idée (peut-être fausse) qu'il est impossible d'implémenter un système de substitution polyalphabétique à la fois sûr et pratique, ce qui a contribué à leur abandon progressif. En effet, si, pour fonctionner, un cryptosystème nécessite des échanges fréquents d'objets physiques, autant utiliser le one-time pad !
Une autre raison de l'abandon de cette méthode est l'arrivée de l'informatique qui permet de créer des moyens de chiffrement totalement différents dont nous allons étudier les exemples les plus connus en détail dans le chapitre 8.
Notons que dans la lutte incessante entre la cryptographie et la cryptanalyse, les améliorations techniques apportées par les siècles amènent à trouver des moyens de chiffrer de plus en plus puissants et il est de plus en plus difficile de les attaquer.
L'arrivée de l'informatique ne déroge pas à cette règle. Contrairement à une idée reçue, sans doute liée au fait que les ordinateurs peuvent aisément casser les anciennes techniques de cryptage, l'utilisation de l'informatique améliore la sécurité des chiffrements et rend la tâche des cryptanalystes de plus en plus ardue.
Actuellement, il existe plusieurs façons de crypter des messages – nous allons en étudier par la suite – pour lesquelles il n'y a pas d'attaque réalisable connue.

VI. Exercices

Exercice 1.1. Moyen

Le texte suivant est chiffré avec l'algorithme de chiffrement de Vigenère :

HEGMMAEHQUPHAIJDEELZPVOQKEINEZJADILLPKVYDPAMHJSQDWSEZWZT
ZAPSOWQKZLGQZAZYOLJPEIASTHWTANIWVVDLABGWHDMJDMOBWCEOEW
WPWAKSIYWMWHNMEPQXMJELAUSWPPPWDIOBJLRCMSOZAVLFVAAGQLAZ
KELWBQZYDINPNQALMIAVJPEZQFREXWMEEJLOSIJAZPLWLXTREAZPHDMJD
MOBWCOEAKPHDMJLRZASTHEBOLWWKMCMKCKOVAIEOIWZUPVPIAYPUJM
ERKEJFREVLZCKCJEIWQLDKABLTRCTSEI

Quelle est la longueur vraisemblable du mot-clé ?

Exercice 1.2. Moyen

Dans un texte composé aléatoirement (en tirant chaque lettre au hasard), on choisit deux lettres. Quelle est la probabilité p_0 que ces lettres soient identiques ?
Si on prend cette fois un texte écrit en français, et de nouveau on choisit deux lettres au hasard, on note p_1 la probabilité d'obtenir deux lettres identiques. Pouvez-vous comparer p_0 et p_1 ?

Exercice 1.3. Facile

Dans des fouilles gallo-romaines on trouve une tablette comportant le texte suivant :

YRV RUGUHV VRQW GH FRPEDWWUH OHV JDXORLV MXVTX D OD YLFWRLUH PDLV DYDQW FHOD GH WXHU DVWHULA HW REHOLA.

Jules Cesar

Qu'en pensez-vous ?