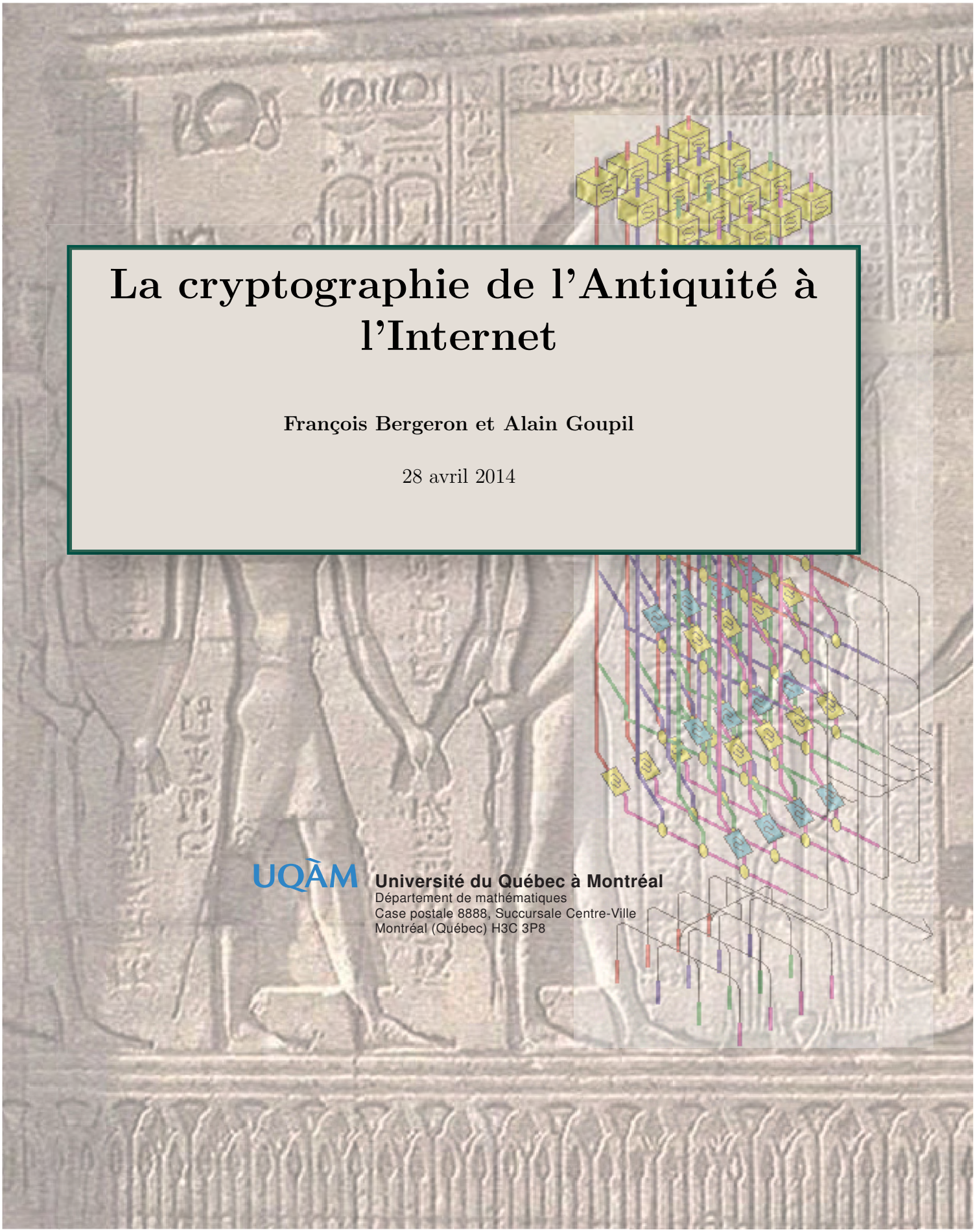


La cryptographie de l'Antiquité à l'Internet

François Bergeron et Alain Goupil

28 avril 2014

UQÀM **Université du Québec à Montréal**
Département de mathématiques
Case postale 8888, Succursale Centre-Ville
Montréal (Québec) H3C 3P8

Table des matières

Préface **iii**

1 Introduction **1**
1.1 Un peu d'histoire . . . 1
1.2 Le jargon de la cryptographie . . . 6
1.3 La cryptographie, les mathématiques et l'informatique . . . 7
1.4 Utilisation courantes de la cryptographie . . . 8

2 Quelques cryptosystèmes simples **11**
2.1 Introduction . . . 11
2.2 Chiffrement par décalage . . . 11
2.3 Chiffrement par substitution . . . 12
2.4 Le code de Vigenère . . . 13
2.5 Chiffrement par permutation de blocs de m lettres . . . 15
2.6 Chiffrement de Hill . . . 15
2.7 Chiffrement de Playfair . . . 18
2.8 Le système $ADFGVX$ . . . 20
2.9 Le chiffre de Vernam . . . 22
2.10 Quelques notions mathématiques . . . 25
2.11 Chiffrement affine . . . 32
2.12 Exercices . . . 32

3 Cryptanalyse des systèmes classiques **37**
3.1 Introduction . . . 37
3.2 Cryptanalyse des systèmes mono alphabétiques . . . 38
3.3 L'écriture automatique . . . 43
3.4 Cassage du chiffre de Vigenère . . . 45
3.5 L'indice de coïncidence . . . 47
3.6 Briser un codage de Hill . . . 49
3.7 Exercices . . . 50

3.8 Appendice : Fréquences de n-grammes 51

4 Probabilités **55**
4.1 La roulette des probabilités 55
4.2 Exemples autour du lancer de deux dés 56
4.3 Le jargon des probabilités 59
4.4 Le jeu de craps 66
4.5 Probabilité totale 69
4.6 Explication de l'indice de coïncidence 70
4.7 Exercices 71

5 La théorie de l'information **73**
5.1 Entropie et incertitude 73
5.2 Propriétés de l'entropie 78
5.3 Quantité d'information et entropie conditionnelle 79
5.4 Systèmes cryptographiques et théorie de l'information 83
5.5 Systèmes par substitution mono alphabétique 85
5.6 Exercices 86

6 Cryptographie moderne **89**
6.1 Introduction 89
6.2 Éléments de théorie des nombres 90
6.3 L'algorithme d'Euclide 93
6.4 Algorithme d'Euclide étendu 95
6.5 Exponentiation modulo n 97
6.6 Le système RSA 102
6.7 Sécurité du système RSA 103
6.8 Recherche de grands nombres premiers 104
6.9 Logarithme discret 105

7 Pour les mordus **109**
7.1 Courbes elliptiques 109
7.2 Cryptosystèmes elliptiques 113
7.3 Chaînes d'additions 116

Bibliographie **119**

Préface

Ces notes accompagnent le cours du même nom[1], offert par la Faculté des Sciences de l'Université du Québec à Montréal. Il est en grande partie basé sur un cours mis au point par Adriano Garsia, du Département de Mathématiques de l'*University of California San Diego*. Nous commençons donc, d'entrée de jeu, par le remercier de nous avoir encouragés dans ce projet.

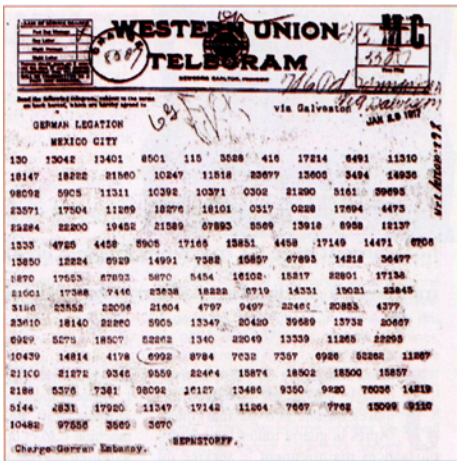
Notre but est de présenter, de façon toute simple, les idées de base de la cryptographie pour un large auditoire. Dans un contexte moderne, où de grandes quantités d'informations sont transmises ou enregistrées de façon codée, notre objectif est d'expliquer comment il se fait que ces codages puissent être facilement vulnérables aux attaques par ordinateur. C'est là un objet d'étude plus spécifique à la cryptanalyse, dont l'apprentissage dépasse largement le niveau de présentation de ce texte. Cependant, nous avons l'intention de bien illustrer le genre de techniques qui rendent possible cette cryptanalyse. Dans un deuxième temps, nous allons présenter des techniques modernes de cryptographie, via lesquelles la sécurité des données cryptées est beaucoup mieux assurée.

Ayant donné ce cours en Californie, le premier auteur a été à même de constater l'engouement que la cryptographie soulève chez un public provenant de disciplines très variées, surtout si l'approche choisie privilégie l'accessibilité. En transposant ce cours au contexte de l'UQAM, nous avons donc tenté de conserver un ton qui permet à tous de bien comprendre les sujets abordés. En particulier, toutes les notions mathématiques nécessaires sont introduites de façon informelle, au fur et à mesure qu'elles deviennent nécessaires. L'accent est donc toujours plus sur l'accessibilité et la clarté, que sur l'exactitude et la rigueur. Cela n'a pas toujours été facile, étant donné notre biais naturel (de mathématiciens) qui nous porte à écrire en saupoudrant les textes des mots «théorèmes», «propositions», «lemmes» et (bien entendu) «preuves». Qu'on se rassure, nous avons presque toujours réussi à résister à cette tentation particulière.

1. Mis au point par les auteurs.

Chapitre 1

Introduction



1.1 Un peu d'histoire

La cryptographie est l'étude des messages secrets. Le terme «cryptographie» vient en effet des mots grecs anciens : **kruptos** ($\kappa\rho\upsilon\pi\tau o\sigma$) qu'on peut traduire comme «secret» ou «caché» ; et **graphein** ($\gamma\rho\alpha\varphi\eta\nu$) pour écriture. Plus précisément, la cryptographie est l'étude des *codes secrets*, et non celle des messages simplement voilés (comme avec de l'encre invisible, par exemple). Les origines de la cryptographie semblent remonter à plus de 4000 ans. On a trouvé, sur une tombe égyptienne de cette époque, des inscriptions contenant des hiéroglyphes modifiés, et il semblerait qu'on ait cherché par ces modifications à obscurcir le sens des inscriptions. Quoi qu'il en soit, plusieurs indications archéologiques tendent à montrer que les «écritures secrètes» sont en fait aussi anciennes que l'invention de l'écriture elle-même.

Scytale

Le premier exemple indéniable de cryptographie remonte au moins au V^{e} siècle avant notre ère. En effet, les Spartiates (Grèce) du temps avaient développé une méthode originale pour l'échange de messages secrets. Celle-ci est basée sur le fait que deux copies identiques d'un bâtonnet, appelé *scytale*, soient en possession de l'envoyeur et du récepteur du message. Pour préparer un message, on enroule en spirale autour de la scytale une bandelette de parchemin (ou de cuir), pour ensuite écrire le message le long de la scytale (voir ci-contre). Une fois déroulée la bandelette ne contient plus qu'une suite apparemment incompréhensible de lettres. Cependant, pour décoder le message il suffit simplement d'enrouler la bandelette sur la scytale jumelle. Comme la méthode est assez simple, il leur fallait bien entendu la conserver secrète.

Le premier texte connu, traitant explicitement de cryptographie, semble être le traité de

Carré de Polybius

Aeneas Tacticus (circa 400 AD), sur la «Défense des fortifications». On sait aussi qu'un autre grec, Polybius (circa 200 AD), développa un système de codage des lettres de l'alphabet par des paires de symboles, utilisant ce qu'on appelle un «carré de Polybius» (voir ci-contre). Son idée a souvent été reprise par la suite. L'utilisation du carré de Polybius consiste à remplacer chaque lettre de l'alphabet par deux nombres, donnant la ligne et la colonne où se trouve cette lettre. Ainsi A deviens 11, B deviens 12, et ainsi de suite. Pour pouvoir utiliser un carré 5 par 5, on identifie les lettres I et J, obtenant ainsi un alphabet à $25 = 5 \times 5$ lettres. Cela ne rend pas les messages trop obscurs, puisqu'on comprend toujours aisément que la signification d'un mot comme $IOURNAL$ est bien $JOURNAL$.

En 44 avant notre ère, Jules César utilisait une simple méthode de substitution de lettres pour communiquer secrètement avec ses généraux. Dans son système cryptographique, connu comme le *code de César*, on place les 26 lettres de l'alphabet dans l'ordre habituel et le message codé est obtenu en décalant circulairement chaque lettre du message clair de trois positions. Autrement dit, on a

$$\begin{array}{cccccccccccccccccccccccc} A & B & C & D & E & F & G & H & I & J & K & L & M & N & O & P & Q & R & S & T & U & V & W & X & Y & Z \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ D & E & F & G & H & I & J & K & L & M & N & O & P & Q & R & S & T & U & V & W & X & Y & Z & A & B & C \end{array}$$

et, pour illustrer, le mot $POURQUOI$ [1] devient $SRXUTXRL$. Bien que le résultat semble tout à fait incompréhensible, nous allons voir qu'il est facile de «casser» le code de César. Ici le terme «casser» signifie qu'on a découvert comment décoder les messages secrets.

WRXWH →

← *DOHD MDFWD HVW*

Ce sont probablement les Arabes qui, les premiers, ont compris clairement les principes de la cryptographie, et commencé à développer la *cryptanalyse*. En particulier, ils ont découvert l'utilisation de l'analyse de la *fréquence des lettres* pour attaquer un système de codage. Dès 1412, al-Kalka-shandi inclus dans son encyclopédie une étude de plusieurs systèmes

1. Pour faciliter la lecture de ce texte, on utilisera les lettres majuscules pour écrire les messages à coder et décoder.

cryptographiques. Il y décrit clairement comment procéder au calcul de fréquence des lettres pour s'attaquer au décodage des messages secrets.

Du côté européen, en 1379, Gabriel de Lavinde fait de la cryptographie une «science» mieux comprise, en publiant le premier manuel sur le sujet. Il y présente sa compilation des systèmes de codage connus. Plusieurs ouvrages d'autres auteurs suivront. Plusieurs se mettent à décrire de nouveaux systèmes de codage, ainsi que des mécanismes pour faciliter ces codages. Ainsi, dans un traité publié en 1466, l'italien Leon Battista Alberti décrit la construction d'outils de codage comme son cadran, illustré ci-contre, qui facilitent les codages «polyalphabétiques». On attribue souvent au français Blaise de Vigenère le développement, en 1586, de ce qui fut longtemps considéré comme un «Chiffre Indeschiffrable». Cependant, la paternité de ce système reviendrait plutôt à Giovan Batista Belaso, en 1553, et Vigenère en aurait simplement clarifié certains aspects.

Cadran d'Alberti

Dans un autre ordre d'idée, il est intéressant de constater que les techniques développées pour la cryptographie ont aidé d'autres domaines. Un des exemples les plus frappants est lié à la découverte de la Pierre de Rosette, trouvée en Égypte en 1799, qui contient trois copies d'un décret de Ptolémée V Épiphane, inscrit en hiéroglyphes (haut), en démotique (centre), et en grec (bas). On a vite été convaincu d'avoir trouvé la clé pour traduire les hiéroglyphes, jusque-là incompréhensibles, ayant constaté que les parties en grec et en démotique correspondaient au même texte. Utilisant les techniques alors connues de la cryptanalyse, Jean-Francois Champollion, parvint en 1822 à décoder le langage des hiéroglyphes.

Pierre de Rosette

Passant ici sous silence une longue période d'utilisation de codes dans les milieux diplomatiques et militaires, on arrive d'emblée au XX^{e} siècle, en particulier au moment de la guerre 1914–1918. En janvier 1917, les Britanniques réussirent à décoder un télégramme chiffré (voir en tête de chapitre) envoyé par le Ministre des affaires étrangères allemand, Zimmermann, au Président mexicain de l'époque. On y proposait au Mexique d'attaquer les États-Unis, avec l'aide des Allemands. Les Britanniques avisèrent aussitôt le Président des États-Units qui, le 2 avril, déclara la guerre à l'Allemagne. C'est cependant au cours de la Seconde Guerre mondiale que la cryptographie s'inscrit véritablement comme élément central des stratégies militaires. L'un des avantages marqués des Alliés, ayant très certainement contribué à leur victoire finale, fut leur capacité de décoder autant les messages secrets des Japonais que des Allemands. Le cas le plus connu (avec livres, documentaires et films à l'appui) est certainement l'histoire entourant le décodage du code Enigma par les Polonais et les Britanniques. La machine Enigma (voir ci-contre) fut brevetée par l'allemand Arthur Scherbius en 1919. Une conjonction d'espionnage classique [2] et d'efforts de mathématiciens polonais permirent de déduire la clé alors utilisée et ainsi de décoder les messages encodés

Machine Enigma

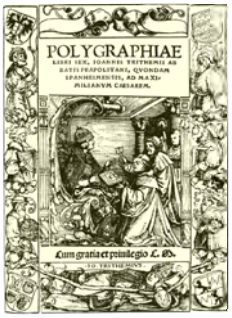
2. Les Français avaient obtenu des photos de manuels d'instructions pour Enigma

avec Enigma. Cependant, les Allemands modifièrent leurs procédures, et les Britanniques s'allièrent aux Polonais et aux Français pour développer des ordinateurs mécaniques (les Bombes de Turing) qui, avec l'aide de mathématiciens, de linguistes, et même de joueurs d'échecs, purent calculer «rapidement» les clés qu'on changeait maintenant plus souvent.

On trouve apparemment bien moins de détails sur les efforts cryptographiques durant la guerre froide, probablement parce que ces informations sont encore «Top Secret». Il est possible que certains des systèmes plus modernes aient été considérés secrètement à cette époque.

On en est maintenant à l'époque moderne des codes à clés publiques, basées sur diverses notions mathématiques, avec toutes les applications nouvelles suscitées par l'utilisation de l'Internet, sans mentionner les utilisations potentiellement plus problématiques comme celles liées au terrorisme. On a aussi des indications claires de tendances à venir comme la cryptographie quantique, basé sur les principes de la théorie des quanta. Enfin, les applications potentielles des outils développés pour la cryptanalyse sont aujourd'hui encore plus variées, incluant par exemple celles dans le domaine de l'étude de génomes.

Quelques textes de cryptographie.



1518

Pour illustrer l'ancienneté de la fascination qu'exerce la cryptographie, voici une liste de publications sur le sujet, datant toutes des XV^{e} et XVI^{e} siècles.

1470 Leone Battista Alberti, *Trattati in cifra*, est publié à Rome. Alberti y traite des théories et processus de chiffrement, méthodes de déchiffrement, et données statistiques.

1518 Abbott Johannes Trithemius (Trithème) écrit (sans le publier) son *Steganographia*, qui a circulé sous forme manuscrite pendant plus de cent ans.

1518 Trithème publie son *Polygraphiae libri sex*, incluant sa «tabula recta» pour faciliter l'utilisation de codes à la César. Celle-ci consiste en un carré de lettres dont les lignes sont constituées de l'alphabet successivement décalé d'une lettre (Voir tableau 1.1). Pour décrire un décalage, on n'a qu'à donner la lettre qui correspond au A (la première sur la ligne correspondante).

1526 Jacopo Silvestri publie *Opus novum ... principibus maxime vtilissimum pro cipharis.* Il y discute de six méthodes de chiffrement, incluant le code de César.

1540 Giovanni Battista Palatino publie *Libro nvova d'imparare a scrivere ... Con vn breue et vtile trattato de le cifere.*

1550 Girolamo Cardano publie *De subtilitate libri XXI.* Son texte contient une grande quantité d'information sur le codage.

	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

TABLE 1.1 – Tabula recta

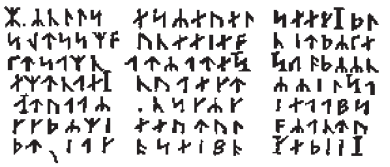
1553 Giovanni Battista Bellaso, *La cifra*. Il y présente la notion de «mot-clef» et un système polyalphabétique.

1586 Blaise de Vigenère publie un livre de 600 pages *Traité des chiffres*. II y discute plusieurs systèmes de codage.

1605 Francis Bacon publie *Proficience and Advancement of Learning Divine and Humane*. Il y décrit simplement la cryptographie et discute de qualités des systèmes de codage.

Quelques incursions de la cryptographie dans le monde littéraire

Dans plusieurs de ses romans, Jules Verne fait intervenir des messages codés. C'est le cas de *La Jangada* (1881), *Le voyage au centre de la Terre* et *Les enfants du capitaine Grant*. Dans le second chapitre du *voyage au centre de la Terre*, le message



joue un rôle prépondérant.

Dans *Le Scarabée d'or* (1843), de Edgar Allen Poe, on trouve le message codé suivant,

53++ !305))6* ;4826)4+.)4+) ;806* ;48 !8‘60))85 ;]8* :+*8 !83(88)5* ! ;
46(;88*96* ? ;8)*+(;485) ;5* !2 :*+(;4956*2(5*-4)8‘8* ; 4069285) ;)6
!8)4++ ;1(+9 ;48081 ;8 :8+1 ;48 !85 ;4)485 !528806*81(+9 ;48 ;(88 ;4(+ ?3
4 ;48)4+ ;161 ; :188 ;+ ? ;

obtenu par substitution de caractères pour les lettres. Un des personnages utilise l'analyse de fréquence de lettres pour décoder le message.

La clé.

En outre, dans l'*Aventure des hommes dansants* (1903), mettant en vedette Sherlock Holmes, Arthur Conan Doyle fait intervenir, comme élément principal de son histoire, un code dans lequel chaque lettre est remplacée par un petit homme dansant différent. L'un des messages est par exemple : . Sherlock Holmes réussit enfin à décoder les messages après en avoir accumulé suffisamment pour pouvoir appliquer une analyse de fréquence des dessins. Il obtient la grille de traduction ci-contre.

Umberto Eco dans son roman de plus de 700 pages, *Le pendule de Foucault*, mentionne Thrithème et son ouvrage *Steganographia*, mais aussi bien d'autres aspects de la cryptographie, des mathématiques, de la biologie, etc.

Dan Brown, dans *Da Vinci Code*, fait intervenir plusieurs codes. Un livre à même été écrit pour discuter des divers codes de ce roman : *Breaking the Da Vinci Code : Answering the Questions Everybody's Asking* de Darrel L. Bock (mai 2004).

1.2 Le jargon de la cryptographie

La *cryptographie* est donc l'étude des méthodes d'envoi de messages *codés* de telle sorte que seul le destinataire puisse le *décoder*. Le message qu'on veut envoyer s'appelle le *texte clair* et le message codé, ou *encrypté*, s'appelle aussi *cryptogramme*.

Le processus de conversion d'un texte clair en message codé s'appelle *chiffrement*, ou *codage* ; et le processus inverse s'appelle *déchiffrement*, ou *décodage*. Pour effectuer un codage, on suit une méthode précise appelée *système de codage*, ou *système cryptographique*, ou même encore *cryptosystème*. Un codage se fait donc à l'aide d'un système cryptographique, et celui-ci nécessite très souvent l'utilisation d'une *clé* de codage. Cette clé (un mot, un nombre, une grille) est nécessaire [3] pour décoder le message chiffré. En d'autres termes, la clé modifie le comportement du mécanisme de codage et de décodage.

Les symboles utilisés dans un message sont appelés des *lettres* et l'ensemble des symboles

3. Nous verrons plus tard des situations avec une clé pour le codage, et une autre clé pour le décodage.

possibles s'appelle l'*alphabet*. On désignera souvent l'alphabet par $\mathcal{A}$. L'alphabet du texte clair peut être différent de l'alphabet du message codé. Le texte clair et le texte chiffré sont souvent découpés en *blocs*. L'intention derrière le découpage en blocs est habituellement d'envoyer le texte comme une succession de blocs qui sont encodés et décodés séparément.

La *cryptanalyse* est l'étude des méthodes qui permettent de découvrir le sens d'un message codé, sans connaître le message original. Il y a plusieurs situations possibles. On peut vouloir simplement trouver le sens du message codé, sans chercher à trouver la clé de codage. Mais, en général on voudra trouver d'abord quel est le système de codage, puis la clé de codage utilisée. Lorsqu'on a trouvé tous les éléments de la méthode utilisée pour coder des messages, on dit qu'on a *cassé*, ou *brisé*, le système cryptographique utilisé. Plus un système est «difficile» à briser, plus il est «sûr».

1.3 La cryptographie, les mathématiques et l'informatique

Avec le temps, les liens entre la cryptographie et les mathématiques sont devenus de plus en plus étroits. Les systèmes cryptographiques modernes sont maintenant tous formulés en termes mathématiques. Ils font intervenir des notions mathématiques importantes et, sans celles-ci, ils seraient impossibles à construire. D'autre part, en cryptanalyse, l'utilisation de techniques mathématiques et informatiques est devenue la norme. Ceci est principalement dû au fait qu'on a montré que tous les systèmes du passé sont soit extrêmement limités (une seule utilisation), soient relativement faciles à briser avec les bons outils mathématiques alliés aux ordinateurs modernes.

D'un point de vue tout à fait général, le processus de chiffrement peut être considéré comme une «fonction» f_k qui décrit comment encoder les messages. Comme on l'a déjà mentionné, les messages trop longs sont découpés en «blocs». La fonction f_k est la recette de codage de ces blocs, qu'elle transforme en blocs codés. Pour exploiter au mieux le potentiel de description de cette façon de voir les choses, on introduit les ensembles suivants. On a d'abord l'ensemble $\mathcal{M}$, de tous les blocs de messages clairs possibles, puis l'ensemble $\mathcal{C}$, de tous les blocs codés possibles. Ainsi, pour un bloc de message clair m, le bloc codé correspondant est $f_k(m)$. Tout ceci est représenté schématiquement à la Figure 1.1. La fonction d'encodage f_k dépend d'une *clé*, désignée ici par k. Cette clé est choisie dans un ensemble (généralement fini) $\mathcal{K}$ de clés possibles. La fonction de décryptage est la fonction *inverse* f_k^{-1} de la fonction d'encryptage f. Elle permet de récupérer le texte clair à partir du texte codé. En d'autres termes, on a

$$f_k(m) = c, \qquad \text{si et seulement si} \qquad f_k^{-1}(c) = m.$$

Décrire un cryptosystème correspond donc à décrire les ensembles $\mathcal{M}$, $\mathcal{C}$, et $\mathcal{K}$, ainsi que la

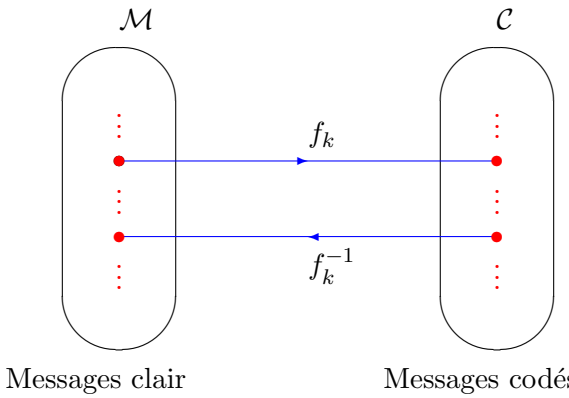


FIGURE 1.1 – Fonctions d'encodage et de décodage

façon de calculer les fonctions de codage

$$f_k : \mathcal{M} \longrightarrow \mathcal{C},$$

et de décodage

$$f_k^{-1} : \mathcal{C} \longrightarrow \mathcal{M};$$

qui dépendent toutes deux de la clé k, choisie dans l'ensemble $\mathcal{K}$.

De plus en plus, les calculs nécessaires au codage et au décodage sont exigeants. On utilise maintenant systématiquement les ordinateurs pour réaliser ces calculs. En fait, les systèmes modernes sont tout à fait impraticables sans un ordinateur. De plus, l'accès facile à des ordinateurs performants rend les anciens systèmes de codage (ceux d'avant les années 1970) presque complètement désuets, et force l'introduction des récentes techniques de cryptographie. Si l'on veut conserver ses secrets, il faut aussi tenir compte de l'évolution fulgurante de la puissance des ordinateurs.

1.4 Utilisation courantes de la cryptographie

Des systèmes cryptographiques de toute sorte sont utilisés de façon courante. Tous ne nécessitent pas le même niveau de sécurité, et les systèmes cryptographiques utilisés varient beaucoup en complexité. Dans certains cas les codages sont très simples, et dans d'autres on cherche à assurer la meilleure sécurité disponible. Les utilisations vont de la téléphonie cellulaire, aux transactions bancaires, en passant par le cryptage de certaines chaînes de

télévision ; sans mentionner les communications diplomatiques, militaires, ou encore terroristes ou criminelles. On comprend donc que, dans certains cas, les impératifs de facilité et de rapidité de codage l'emportent sur l'assurance d'une sécurité absolue.

Chapitre 2

Quelques cryptosystèmes simples

2.1 Introduction

Au cours d'un échange visant à communiquer de façon secrète, deux protagonistes, appelés ici l'émetteur et le récepteur, s'entendent sur la nature du système cryptographique à utiliser. Puis, après avoir choisi une clé secrète déterminant la manière dont le système effectuera le codage, l'émetteur fait parvenir cette clé au récepteur de façon à ce qu'aucun tiers (opposant) ne puisse intercepter celle-ci. Ils sont alors prêts à communiquer, mais ils n'ont pas l'assurance que les messages codés ne seront pas interceptés par l'opposant. Le but de l'opposant est de décoder le message secret transmis par l'émetteur au récepteur. Plus cette tâche est difficile, plus le système est considéré comme sûr.

Pour illustrer toute cette démarche, nous allons présenter quelques systèmes cryptographiques.

2.2 Chiffrement par décalage

On peut facilement modifier le code de César, mentionné à la section 1.1. Pour ce faire, on place les 26 lettres de l'alphabet dans l'ordre habituel et le message codé est obtenu en décalant circulairement chaque lettre du message en clair d'un nombre fixé de positions. Un décalage de 3 était utilisé par Jules César, mais on peut en fait utiliser n'importe quel autre décalage. La valeur, d, de ce décalage est la *clé* du système de codage.

Ce système est très vulnérable aux *attaques*. Ainsi, supposons qu'on ait intercepté un mes-

sage codé :

$$HXAZAY \quad KYZ \quad VXKVGXK \quad G \quad RK \quad ZAKX \tag{1}$$

en sachant que l'envoyeur a utilisé un système par décalage. Pour décoder le message, il nous faut trouver la clé de décalage. Si le message est très court, la meilleure façon de procéder est probablement d'essayer de décoder avec les 25 possibilités de valeurs pour la clé. La plupart du temps, seulement une de ces possibilités donnera un message qui a un sens. Toutes les autres clés donneront des messages aussi peu expressifs que (1). Une façon beaucoup plus élégante consiste à analyser la *fréquence des lettres* dans le message. On compte donc, pour chaque lettre, le nombre de fois que la lettre apparaît dans le message. On cherche ici à exploiter le fait que la lettre E est (en général [1]) la plus fréquente dans un texte français (ou anglais). Comme le décalage est le même pour chaque lettre, on aura gagné si on trouve le décalage pour E. Dans le message codé (1), la lettre la plus utilisée est le K. S'il est vrai que le E du message en clair correspond bien à K dans le message codé, alors le décalage doit être de 6, puisque K est 6 positions plus loin que E. On essaie alors de décoder avec ce décalage, pour obtenir

$$BRUTUS \quad EST \quad PREPARE \quad A \quad LE \quad TUER \tag{2}$$

ce qui semble bien être la solution. Si un doute subsiste, on peut toujours tenter l'approche exhaustive décrite ci-haut. Pour s'habituer au jargon de la cryptranalyse, on peut dire qu'on a réussi à casser le système cryptographique utilisé. Le cryptosystème par décalage ne résiste donc pas à une attaque basée sur l'analyse de fréquence des lettres. On est ainsi poussé, en tant qu'obsédé du secret, à développer des codes plus sûrs.

2.3 Chiffrement par substitution

Une première approche consiste à généraliser les codes par décalages en considérant des chiffrements par *substitution* plus élaborés. On suppose pour l'instant que l'alphabet des textes en clair est le même que l'alphabet des messages codés. Les blocs de messages en clair, ou codés, sont simplement constitués des lettres de l'alphabet. On a donc

$$\mathcal{M} = \mathcal{C} = \{A, B, C, \ldots, Z\}.$$

Pour coder les messages, on commence par se choisir une clé, à savoir une *permutation* quelconque des 26 lettres de l'alphabet. Cette permutation prend la forme d'une correspondance comme la suivante :

1. Ceci est vrai si le texte est assez long,à moins qu'on ne soit tombé sur l'une des oeuvres de Georges Perec, comme *La disparition*.

$$\begin{array}{cccccccccccccccccccccccc} A & B & C & D & E & F & G & H & I & J & K & L & M & N & O & P & Q & R & S & T & U & V & W & X & Y & Z \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ S & M & B & A & Z & Q & H & J & I & R & Y & C & W & G & N & T & U & D & O & F & E & L & V & K & P & X \end{array} \tag{3}$$

qu'on lit de haut en bas. Donnant le nom σ à la permutation décrite en (3), on écrit encore $\sigma(A) = S$, $\sigma(B) = M$, etc. Pour décoder un message, le receveur utilise la permutation inverse de σ, notée σ^{-1}. On l'obtient simplement en lisant la correspondance (3) de bas en haut, plutôt que de haut en bas.

Observons que pour choisir une permutation, on doit d'abord choisir la lettre correspondant à A et il y a de 26 possibilités, puis on choisit la lettre correspondant à B de l'une des 25 façons restantes (toutes les possibilités, sauf la lettre qui a été choisie pour A), et ainsi de suite pour chaque autre lettre. Il y a donc

$$26 \times 25 \times \ldots \times 2 \times 1 = 403291461126605635584000000$$

permutations possibles de l'alphabet français. On en conclut qu'on ne peut plus passer en revue toutes les clés possibles pour décoder le message.

Le chiffrement par décalage est un cas particulier du chiffrement par substitution, car un décalage est en fait une permutation particulière. Nous allons voir au chapitre 3 que ce système est vulnérable à une analyse de fréquence un peu plus poussée. Continuons donc notre recherche de cryptosystèmes plus sûrs.

Blaise de Vigenère (1523-1596)

2.4 Le code de Vigenère

En 1550, le diplomate français Blaise de Vigenère (1523-1596), secrétaire de Charles IX, voyage partout en Europe et se familiarise avec les méthodes cryptographiques connues pour des raisons professionnelles. Dix ans plus tard, il abandonne sa carrière de diplomate, et se consacre exclusivement à l'étude détaillée des écrits d'Alberti, de Trithème et de Porta. Il donne la forme finale à un nouveau chiffre puissant auquel on donne plus tard son nom. La force du chiffre de Vigenère vient du fait qu'une même lettre en clair peut être chiffrée de différentes manières. À partir de l'invention de l'imprimerie, vers 1450, l'emploi des messages codés se généralise dans les relations diplomatiques entre les états européens et chez les militaires. L'art du chiffrement et du déchiffrement évolue plus rapidement et des scientifiques renommés contribuent au développement des techniques utilisées. Le mathématicien Cardano, de Milan (célèbre pour sa résolution des équations du troisième degré), l'architecte Alberti de Florence, et l'abbé Trithème font évoluer la science du chiffre. Dans *Le traité*

Traicté des chiffres de Vigenère

des secrètes manières d'écrire, Vigenère décrit le chiffre qu'il qualifie «d'indeschiffrable». Le procédé de Vigenère, fondé sur la tabula recta de Trithème (Tab. 1.1), consiste à changer l'alphabet de substitution à chaque chiffrement d'une lettre, ce qui fait qu'on ne peut tenter de décrypter le message en utilisant simplement un calcul de fréquence des lettres.

Description du chiffre de Vigenère

Le chiffrement dépend d'un mot clé, dans l'exemple ci-dessous c'est le mot $PERMUTE$. Pour coder un mot en clair, comme $SECURITE$, on consulte la tabula recta à la ligne commençant par la première lettre, P, du mot-clé. On remplace alors la première lettre, S, du mot en clair par son correspondant H sur cette ligne. On procède de même pour la seconde lettre E du mot en clair, mais on utilise maintenant la ligne commençant par la seconde lettre E du mot clé. On continue ainsi pour les autres lettres, en recommençant au début du mot clé si nécessaire. Le chiffrement du mot en clair $SECURITE$ donne donc le mot codé $HITGLBXT$. Comme autre exemple un peu plus long, chiffrons le texte suivant

$$LEDOUTEESTLECOMMENCEMENTDELASAGESSE \tag{4}$$

avec le mot-clé $NAVIRE$, en ignorant les espaces. Le résultat est

$$YEYWLXRENBCIPOHUVRPEHMEXQEGIJETENAV$$

À partir de la 7$^{\text{e}}$ lettre on recommence à coder avec le décalage N, etc. Dans le chiffrement par décalage et le chiffrement par substitution, un caractère en clair est toujours transformé dans le même caractère codé. De tels procédés sont appelés *chiffrements mono alphabétique*, par opposition au chiffre de Vigenère qui n'est pas mono alphabétique. En effet, lorsque le mot clé est de longueur m, chaque lettre en clair est codée de m façons différentes. Dans un tel cas, on dit avoir un *chiffrement poly alphabétique*. Le nombre de mots-clés possibles (sans faire attention au sens), pour la longueur m se calcule de la façon suivante. On a 26 choix pour la première lettre du mot, puis encore 26 choix pour la seconde lettre, et ainsi de suite. On a donc, au total, $26 \times 26 \times \ldots \times 26 = 26^m$ mots-clés possibles de longueur m. Par exemple, pour $m = 8$, on a

$$26^8 = 208827064576$$

clés possibles. Le système de Vigenère ne sera déchiffré qu'au milieu du XIX$^{\text{e}}$ siècle, et demeurera à la base de la plupart des machines à chiffres, jusqu'au début du XX$^{\text{e}}$ siècle.

2.5 Chiffrement par permutation de blocs de m lettres

Il est possible d'utiliser des permutations de m lettres ($m < 26$), plutôt que des permutations de 26 lettres, sur des blocs de m lettres en clair. Si par exemple on veut chiffrer

$$\textit{LES CHEMISES DE L'ARCHIDUCHESSE} \tag{5}$$

avec la permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 2 & 4 \end{pmatrix}.$$

On commence par séparer le texte en clair en blocs de 5 lettres. Si le dernier bloc contient moins de 5 lettres, on lui ajoute des lettres qui ne sont pas susceptibles de brouiller le message comme $X, Q \ldots$ etc. Puis dans chaque bloc de 5 lettres, on mélange les lettres entre elles à l'aide de la permutation. Dans la phrase qui suit on choisit de considérer les espaces entre les mots comme faisant partie de l'ensemble des lettres. On a les 6 blocs

LES_C, HEMIS, ES_DE,_LARC , HIDUC, HESSE

Chaque bloc $x_1x_2x_3x_4x_5$ est transformé en un bloc $x_{\sigma(1)}x_{\sigma(2)}\ldots x_{\sigma(5)}$ et on obtient les blocs transformés suivants

SLCE_, MHSEI, _EESD, A_CLR, DHCIU, SHEES

et le texte chiffré est finalement

SLCE MHSEI EESDA CLRDHCIUSHEES

Il est amusant de constater que les mots sont maintenant découpés de façon différente.

2.6 Chiffrement de Hill

De prime abord, le système poly alphabétique de Hill [2] semble éviter les défauts du chiffre de Vigenère. On commence par regrouper les lettres du texte en clair en blocs de m caractères. On numérise ces blocs, puis on les code au moyen d'une certaine *matrice*. Pour numériser les lettres, on procède très souvent de la façon suivante

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*	,	!	?
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28

(6)

2. Mis au point par Lester S. Hill en 1929.

Tous les calculs s'effectuent ici *modulo* le nombre de lettres de l'alphabet de référence. On trouvera à la Section 2.10 une description simple des notions mathématiques pertinentes, ainsi qu'une explication des raisons qui nous poussent ici à choisir un alphabet avec un nombre *premier* de lettres. On travaillera donc avec 29 lettres, tout simplement en ajoutant les caractères de ponctuation «,», « !» et « ?».

La méthode est présentée ici par l'exemple, plutôt que de façon théorique, avec des blocs de longueur 2. Comme on l'a annoncé, les blocs sont d'abord numérisés pour produire des blocs de nombres (écrits verticalement pour les besoins de la cause). Ceci se fait tout simplement en numérisant chaque lettre du bloc. Ainsi, on a

$$PD \rightsquigarrow \begin{pmatrix} 15 \\ 3 \end{pmatrix}$$

La matrice $\mathbf{T}$ de codage utilisé ci-dessous est

$$\mathbf{T} := \begin{pmatrix} 11 & 13 \\ 5 & 6 \end{pmatrix},$$

et on code un couple

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

en le *multipliant* par $\mathbf{T}$:

$$\begin{pmatrix} 11 & 13 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 11\,x_1 + 13\,x_2 \pmod{29} \\ 5\,x_1 + 6\,x_2 \pmod{29} \end{pmatrix}$$

C'est là la *multiplication matricielle* modulo 29. On obtient ainsi

$$\mathbf{T} \begin{pmatrix} 15 \\ 3 \end{pmatrix} \equiv \begin{pmatrix} 1 \\ 6 \end{pmatrix} \pmod{29}$$

puisque

$$\begin{aligned} 11 \cdot 15 + 13 \cdot 5 &= 204, \\ 5 \cdot 15 + 6 \cdot 11 &= 93. \end{aligned}$$

et que $204 \equiv 1 \pmod{29}$ et $93 \equiv 6 \pmod{29} = 6$. Le calcul du *modulo* consiste simplement à trouver le reste de la division par 29. Pour des raisons surtout esthétiques, la réponse est reconvertie en blocs de lettres en utilisant (6) à l'envers. Globalement on a donc

$$PC \longmapsto BG.$$

Pour coder un message plus long, comme le message *PASDEREPONSE*, on numérise d'abord le message pour obtenir la matrice

$$\begin{matrix} PA & SD & ER & EP & ON & SE \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{matrix}$$
$$\begin{pmatrix} 15 & 18 & 4 & 4 & 14 & 18 \\ 0 & 3 & 17 & 15 & 13 & 4 \end{pmatrix}$$

où chaque colonne correspond à deux lettres consécutives du message. La multiplication, modulo 29, de cette matrice par la matrice $\mathbf{T}$ donne

$$\begin{pmatrix} 11 & 13 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 15 & 18 & 4 & 4 & 14 & 18 \\ 0 & 3 & 17 & 15 & 13 & 4 \end{pmatrix} \equiv \begin{pmatrix} 20 & 5 & 4 & 7 & 4 & 18 \\ 17 & 21 & 6 & 23 & 3 & 27 \end{pmatrix} \pmod{29}$$

Qu'on transforme en bloc de lettres pour obtenir

$$\begin{pmatrix} 20 & 5 & 4 & 7 & 4 & 18 \\ 17 & 21 & 6 & 23 & 3 & 27 \end{pmatrix}$$
$$\begin{matrix} \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ UR & FV & EG & HX & ED & S! \end{matrix}$$

Le résultat final est donc la transformation

$$PASDEREPONSE \longmapsto URFVEGHXEDS!$$

Plus généralement on numérisera un message en k blocs[3] de m lettres, pour obtenir une matrice $\mathbf{M}$, avec m lignes et k colonnes. La matrice $\mathbf{M}$ est alors encodée en la multipliant, modulo p, par $\mathbf{T}$ (la matrice de codage), de taille $m \times m$. Le résultat, $\mathbf{TM}$, est finalement retransformé en lettres.

Le décodage se fait de façon tout à fait analogue. Il suffit simplement de remplacer la matrice $\mathbf{T}$ par sa matrice *inverse*, $\mathbf{T}^{-1}$, dans le processus décrit ci-haut. Bien entendu, cela nécessite qu'on puisse calculer cette matrice inverse. Nous allons voir comment faire ce calcul à la Section 2.10. Dans le cas de notre exemple, cette matrice inverse est

$$\begin{pmatrix} 11 & 13 \\ 5 & 6 \end{pmatrix}^{-1} = \begin{pmatrix} 6 & 16 \\ 24 & 11 \end{pmatrix} \pmod{29}$$

Pour se rassurer que ceci décode bien les messages, on peut vérifier que

$$\mathbf{T}^{-1} \begin{pmatrix} 1 \\ 6 \end{pmatrix} = \begin{pmatrix} 15 \\ 3 \end{pmatrix} \pmod{29}$$

3. Ici k est égal à la longueur du message divisée par m. On ajoute parfois des lettres bidons au message, pour faire en sorte que sa longueur puisse se diviser par m.

Pour les mordus

Il est intéressant de constater que le chiffrement par permutation est un cas particulier du chiffrement de Hill. En effet, chaque permutation σ de m caractères peut être représentée par une matrice M_σ de format $m \times m$. La matrice M_σ est caractérisée par

$$M_\sigma(i,j) = \begin{cases} 1 & \text{si } \sigma(i) = j \\ 0 & \text{autrement} \end{cases}$$

Ainsi, à la permutation $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 2 & 4 \end{pmatrix}$ correspond la matrice 5×5

$$M_\sigma = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

La matrice M_σ est appelée une matrice de permutation et on observe que

$$(x_3, x_1, x_5, x_2, x_4) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$$

C. Wheatstone (1802-1875)

2.7 Chiffrement de Playfair

En 1854, Sir Charles Wheatstone invente un algorithme de cryptage baptisé *Playfair* en l'honneur de son ami Lyon Playfair, baron de St-Andrews, qui a fait militer pour l'adoption de ce chiffre par le gouvernement britannique. Sa simplicité et sa solidité, comparées aux techniques de substitution ont provoqué son succès immédiat dans le monde de la cryptographie, en particulier chez les anglais durant la guerre de Boers et la Première Guerre mondiale. Il a aussi été utilisé par plusieurs forces armées, durant la Seconde Guerre mondiale. Ainsi, lorsque la frégate PT-109 du Lieutenant John F. Kennedy fut coulée par un navire japonais au large des îles Salomon, J.F. Kennedy a pu atteindre, avec les survivants de son équipage, le rivage de l'île Plum Pudding en territoire ennemi. Il y a émis un message crypté avec le chiffre de Playfair. Une opération de sauvetage a pu être organisée et tous les survivants furent récupérés.

Construction du carré de Playfair

Afin d'encrypter un message avec l'algorithme de Playfair, on choisit un mot-clé qu'on écrit dans un tableau 5×5 de gauche à droite à partir du haut sans répétition de lettre. Les lettres restantes de l'alphabet sont insérées à la suite du mot-clé selon l'ordre alphabétique, avec les lettres I et J dans la même case. Dans l'exemple suivant, on utilise le mot-clé ADELAIDE.

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

Table 2.1 – Tableau de Playfair du mot ADELAIDE

Préparation du message en clair pour l'encodage

Le texte a être encodé est d'abord expurgé de toute ponctuation, espace. etc. ne laissant que des lettres de l'alphabet sous forme majuscule. Les chiffres sont épelés en mots et les J sont convertis en I. On regroupe ensuite les lettres du texte en blocs de deux lettres, appelés des *bigrammes*, séparés par des espaces. Il est important dans le processus d'encodage qu'il n'y ait aucun bigramme constitué de 2 lettres identiques. Ainsi les doublets sont éliminés en insérant un X entre 2 lettres identiques et en décalant le reste à droite. Si la chaîne de lettres est de longueur impaire, on complète la dernière paire en ajoutant un X à droite pour ne produire que des bigrammes. Le message est maintenant prêt à être encodé.

Encodage

L'encodage est obtenu en transformant chaque bigramme en un nouveau bigramme à l'aide du carré de Playfair selon les règles suivantes :

— **Règle de la même ligne.** Lorsque les lettres d'un bigramme sont sur la même rangée, on les remplace par les lettres situées immédiatement à leur droite avec la convention que le voisin de droite de la lettre à la fin d'une rangée est la première lettre de cette rangée. Ainsi la transformation du bigramme DE par le carré du

tableau 2.1 donne

$$D\,E \to E\,L$$

— **Règle de la même colonne.** Lorsque les deux lettres sont sur la même colonne, on les remplace par les lettres immédiatement en bas de celles-ci avec la convention que la lettre voisine de la lettre du bas d'une colonne est la première lettre au haut de cette colonne.

— **Règle des coins de rectangle.** Si les deux lettres X_1X_2 ne sont ni sur la même ligne ni sur la même colonne alors on identifie le rectangle dont les deux lettres forment les extrémités d'une de ses diagonales. On remplace les deux lettres par les lettres qui forment l'autre diagonale du rectangle en commençant par la lettre sur la rangée de X_1

$$\begin{array}{ccc} y_2 & \cdots & X_2 \\ \vdots & & \vdots \\ X_1 & \cdots & y_1 \end{array}$$

En guise d'illustration, UX est remplacé par SZ selon l'application de ces règles avec le carré du tableau 2.1. Toujours avec ce même carré, la phrase

28 av. Mississipi

donne le codage

$$ZAOFUGZHQLWKEUUEXEUEUH$$

Jolly good, n'est-il pas !

Georges Painvin (1886-1980)

2.8 Le système *ADFGVX*

Le système *ADFGVX* a été introduit en 1918 par le Colonel allemand Fritz Nebel. Le fait qu'il ait été décrypté, par le Lieutenant Georges Jean Painvin, a permis au Grand État-Major français de bloquer la dernière offensive allemande.

La clé est constituée de deux parties. On a d'abord une grille de 6 lignes et 6 colonnes. On étiquette les lignes, de haut en bas, et les colonnes de gauche à droite, avec les lettres A, D, F, G, V et X respectivement. On remplit alors, au hasard, les 36 cases de la grille avec les 26 lettres de l'alphabet $\{A, B, C, \ldots, Z\}$ et les 10 nombres $\{0, 1, 2, \ldots, 9\}$. Le résultat est

appelé *grille ADFGVX*. Ainsi :

	A	*D*	*F*	*G*	*V*	*X*
A	*C*	*O*	8	*X*	*F*	4
D	*M*	*K*	3	*A*	*Z*	9
F	*N*	*W*	*L*	0	*J*	*D*
G	5	*S*	*I*	*Y*	*H*	*U*
V	*P*	1	*V*	*B*	6	*R*
X	*E*	*Q*	7	*T*	2	*G*

(7)

est une telle grille *ADFGVX*. La deuxième composante de la clé est une permutation des nombres $1, 2, \ldots, 2\,n$, pour n un entier fixé. Par exemple

4 9 5 15 2 8 16 12 13 17 1 18 3 19 10 7 6 11 14 20

Nous allons montrer comment utiliser le système en codant la phrase

HQ REQUESTS FRONT LINE SITUATION BY TELEGRAM. HQ 7TH CORP

La première étape consiste à écrire le message dans un rectangle de n colonnes. Si nécessaire, on ajoute des lettres quelconques pour remplir les dernières cases. Pour notre exemple, on obtient le rectangle 5×10 suivant

H	*Q*	*R*	*E*	*Q*	*U*	*E*	*S*	*T*	*S*
F	*R*	*O*	*N*	*T*	*L*	*I*	*N*	*E*	*S*
I	*T*	*U*	*A*	*T*	*I*	*O*	*N*	*B*	*Y*
T	*E*	*L*	*E*	*G*	*R*	*A*	*M*	*H*	*Q*
7	*T*	*H*	*C*	*O*	*R*	*P*	*S*	*E*	*D*

Les trois dernières lettres *S*, *E* et *D* ont justement été ajoutées aux fin de remplissage. La prochaine étape consiste à placer, dans ce rectangle, au-dessus de chaque lettre la paire qui correspond aux «coordonnées» dans la grille *ADFGVX*. Ainsi, la première lettre du message est un *H*, qui se trouve dans la ligne *G* et la colonne *V* de la grille *ADFGVX*. On aura donc

G	*V*
H	

La prochaine lettre étant *Q*, on trouve

X	*D*
Q	

et ainsi de suite

V	*X*		*X*	*A*		
R		,	*E*		,	$\ldots$

Après avoir ainsi modifié le rectangle, on étiquette les colonnes correspondant à ces coordonnées selon les valeurs de la permutation choisie comme second élément de la clé, pour obtenir enfin :

4	9	5	15	2	8	16	12	13	17	1	18	3	19	10	7	6	11	14	20
G	V	X	D	V	X	X	A	X	D	G	X	X	A	G	D	X	G	G	D
	H		Q		R		E		Q		U		E		S		T		S
A	V	V	X	A	D	F	A	X	G	F	F	G	F	F	A	X	A	G	D
	F		R		O		N		T		L		I		N		E		S
G	F	X	G	G	X	D	G	X	G	G	F	A	D	F	A	V	G	G	G
	I		T		U		A		T		I		O		N		B		Y
X	G	X	A	F	F	X	A	X	X	V	X	D	G	D	A	G	V	X	D
	T		E		L		E		G		R		A		M		H		Q
X	F	X	G	G	V	A	A	A	D	V	X	V	A	G	D	X	A	F	X
	7		T		H		C		O		R		P		S		E		D

Le message codé est obtenu en lisant les colonnes dans l'ordre de ces étiquettes, de 1 à $2\,n$. On commence donc, pour notre exemple par le contenu de la colonne 1 : $GFGVV$, puis la colonne 2 : $VAGFG$, etc. On obtient donc le message

$$\begin{array}{cccccc} GFGVV & VAGFG & XGADV & GAGXX & XVXXX & XXVGX \\ DAAAD & XDXFV & VVFGF & GFFDG & GAGVA & AAGAA \\ XXXXA & GGGXF & DXGAG & XFDXA & DGGXD & XFFXX \\ AFDGA & DDGDX & & & & \end{array}$$

Gilbert Vernam (1890-1960)

2.9 Le chiffre de Vernam

En 1917 pendant la Première Guerre mondiale, l'américain Gilbert Vernam reçoit le mandat de la compagnie AT&T d'inventer une méthode d'encryption incassable. Celui-ci met au point un cryptosystème qui, lorsque correctement utilisé, est démontré incassable. C'est grâce à l'amélioration du major Joseph O. Mauborgne, en 1918, que l'objectif visé fut réellement atteint. On emploie encore ce chiffre dans des situations délicates, comme le téléphone rouge entre Moscou et Washington. La version améliorée du *chiffre de Vernam* est en fait un chiffre de Vigenère dont la caractéristique est que la clé de chiffrement a la même longueur que le message en clair.

Pour chiffrer un texte de manière sûre avec le chiffre de Vernam, on doit

1- *choisir une clé aussi longue que le texte à chiffrer*;

2- *utiliser une clé constituée de caractères choisis aléatoirement*;

3- *garder la clé secrète*;

4- *ne jamais utiliser 2 fois la même clé*;

5- *écrire des textes en clair ne contenant que les lettres de l'alphabet français.*

Utilisons le chiffre de Vernam pour coder la phrase suivante [4] :

DIEUNEJOUEPASAUXDES

Nous avons besoin d'une clé aléatoire aussi longue que le message, nous choisissons la suivante :

XCAATELPRVGZCRSTJEQ

Chaque lettre de la clé donne le décalage de la lettre correspondante du message en clair. On obtient le codage :

AKEUGIUDLZVZURMQMZI

Pour décoder, il suffit de procéder à l'opération inverse. Nous allons voir plus tard que, sans la clé secrète, il est impossible de récupérer quelque indice que ce soit sur la nature du message en clair. C'est ce qu'on appelle un *code parfait*.

Dans sa description originale, Vernan n'avait cependant pas une façon réellement aléatoire de construire la clé de codage, et on peut casser son système avec relativement peu d'information.

Description originale de Vernan

On commence avec deux listes aléatoires d'entiers, respectivement de longueur p et q, avec p et q des nombres premiers distincts :

$$\begin{aligned} u &= (u_1, u_2, \ldots, u_p) \\ v &= (v_1, v_2, \ldots, v_q). \end{aligned}$$

Par exemple,

$$\begin{aligned} u &= (3, 1, 2) \\ v &= (7, 3, 8, 4, 5) \end{aligned}$$

4. Affirmation d'Albert Einstein en réaction à la physique quantique.

qui sont de longueur 3 et 5. On construit ensuite deux listes de longueur pq, en répétant q fois la liste u ; et répétant la liste v, p fois. Ainsi, pour notre exemple, on obtient les listes :

$$(3,1,2,3,1,2,3,1,2,3,1,2,3,1,2),$$

constituée de 5 copies de la liste $(3,1,2)$, et

$$(7,3,8,4,5,7,3,8,4,5,7,3,8,4,5),$$

constituée de 3 copies de la liste $(7,3,8,4,5)$. Ces deux nouvelles listes sont alors additionnées terme à terme :

$$\begin{array}{rl} & (3,\ 1,\ 2,\ 3,\ 1,\ 2,\ 3,\ 1,\ 2,\ 3,\ 1,\ 2,\ 3,\ 1,\ 2) \\ + & (7,\ 3,\ 8,\ 4,\ 5,\ 7,\ 3,\ 8,\ 4,\ 5,\ 7,\ 3,\ 8,\ 4,\ 5) \\ \hline w = & (10,\ 4,\ 10,\ 7,\ 6,\ 9,\ 6,\ 9,\ 6,\ 8,\ 8,\ 5,\ 11,\ 5,\ 7) \end{array}$$

pour produire une liste $w = (w_1, w_2, \ldots, w_n)$, de longueur $n = pq$ (dans notre cas, $pq = 15$). C'est la clé qui sera utilisée. Quitte à ajouter des lettres, on fait en sorte que le message à envoyer soit de longueur égale à pq. S'il est plus long, on doit choisir des valeurs plus grandes de p et de q. Pour coder le message, on numérise chaque lettre du texte en clair comme cela est décrit en (6) à la section **??**, pour obtenir

$$m = (m_1, m_2, \ldots, m_n).$$

Le message codé s'obtient en calculant

$$\begin{array}{rl} & (m_1,\ m_2,\ \ldots,\ m_n) \\ + & (w_1,\ w_2,\ \ldots,\ w_n) \pmod{26} \\ \hline c = & (c_1,\ \ c_2,\ \ \ldots,\ c_n), \end{array}$$

qui est finalement retransformé en lettres. Ainsi, la numérisation du message

$$PLUSDEMUNITIONS$$

donne

$$(15,11,20,18,3,4,12,20,13,8,19,8,14,13,18)$$

on calcule

$$\begin{array}{rl} & (15,\ 11,\ 20,\ 18,\ 3,\ 4,\ 12,\ 20,\ 13,\ 8,\ 19,\ 8,\ 14,\ 13,\ 18) \\ + & (10,\ 4,\ 10,\ 7,\ 6,\ 9,\ 6,\ 9,\ 6,\ 8,\ 8,\ 5,\ 11,\ 5,\ 7) \pmod{26} \\ \hline & (25,\ 15,\ 4,\ 25,\ 9,\ 13,\ 18,\ 3,\ 19,\ 16,\ 1,\ 13,\ 25,\ 18,\ 25) \end{array}$$

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
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
6	38	32	4	8	30	36	34	39	31	78	72	70	76	9	79	71	58	2	0	52	50	56	54	1	59

TABLE 2.2 – Codage préliminaire de l'alphabet dans le chiffre du Che

qui donne $ZPEZJNSDTQBNZSZ$, lorsque réécrit en terme de lettres.

L'idée erronée de Vernan était que les nombres de la liste w apparaissent de façon complètement aléatoire. Cependant, on peut montrer qu'un espion qui connaîtrait une relativement petite portion du texte original (en fait, $p+q-1$ lettres) serait en mesure de complètement décoder le message.

Le chiffre du Che

Che Guevara (1928-1967)

(Voir : *Pour la science*, juillet-oct. 2002, p. 115) Lorsqu'en 1967, l'armée bolivienne captura et exécuta le révolutionnaire Che Guevara, les militaires trouvèrent sur son corps un papier montrant comment il préparait les messages qu'il transmettait à Fidel Castro. Le Che utilisait essentiellement le chiffre de Vernam. Les lettres du message en clair étaient d'abord transformées en nombres selon la règle de substitution du tableau 2.2. Comme nous allons le voir à la section 3.2, cette première procédure ne procure pas de véritable protection. Les chiffres du message sont alors découpés en blocs de cinq chiffres. Ce sont les lignes supérieures que l'on voit sur le document de la Figure 2.1. La ligne du milieu est la clé. C'est une suite aléatoire (sans structure) de chiffres de longueur égale au message chiffré de la première ligne. La ligne du bas de chaque groupe de trois lignes est obtenue en additionnant sans retenue les chiffres des 2 premières lignes. Puis il faut faire la substitution inverse du tableau 3.5 pour traduire les chiffres en lettres. Ceci constitue le message codé. Bien entendu, pour décoder, Fidel Castro devait avoir les clés de codage en sa possession.

2.10 Quelques notions mathématiques

Un zest de modulo

Dans notre exploration des systèmes cryptographiques, nous avons souvent été amenés à calculer «circulairement». La situation est tout à fait analogue au calcul des heures de la journée, des jours de la semaine, ou des mois de l'année. Typiquement, on se demande quel sera le jour de la semaine (dimanche, lundi, etc.) dans 75 jours, si nous sommes aujourd'hui

FIGURE 2.1 – Un message codé par Che Guevara

un jeudi. Vous pouvez certainement imaginer des questions semblables en ce qui concerne les heures de la journée (avec le système des 24 heures), ou les mois de l'année.

Dans tous ces cas, c'est la même approche mathématique, dont nous aurons d'ailleurs besoin pour discuter plus clairement de nombreuses questions de cryptographie. L'idée consiste simplement à «compter» en revenant à 0 lorsqu'on atteint un certain seuil m donné à l'avance ; 24 pour les heures de la journée, 7 pour les jours de la semaine, ou encore 12 pour les mois de l'année.

Mathématiquement, le contexte dans lequel on se situe est l'ensemble des nombres

$$0, 1, 2, 3, \ldots, m-1$$

qu'on désigne par $\mathbb{Z}_m$. Si un nombre n est plus grand que $m-1$, on le ramène à un nombre de l'ensemble $\mathbb{Z}_m$, en prenant simplement le reste de la division de n par m. On apprend dès la petite école que la division d'un entier par un autre entier peut être juste ou non. Ainsi, on a $12 \div 3 = 4$ et $11 \div 4 = 2 + 3/4$. On dit alors que 3 est le reste de la division de 11 par 4. En général, la *division* d'un entier n par un entier m donne un *quotient* q, avec

un *reste* r :

$$\frac{n}{m} = q + \frac{r}{m}.$$

Ici, $q \cdot m$ est le plus petit multiple de m qui est inférieur à n, et la valeur du reste r, se situe entre 0 et $m - 1$. On a donc :

$$n = q \cdot m + r, \qquad \text{avec} \quad 0 \leq r < m.$$

De façon imagée, on peut représenter ceci comme

c'est-à-dire que, pour $n = 104$ et $m = 3$, on aura $q = 34$ et $r = 2$:

$$\frac{104}{3} = 34 + \frac{2}{3}.$$

Autrement dit,

$$104 = 34 \cdot 3 + 2.$$

Lorsqu'on divise par 10, les valeurs possibles pour les restes sont

$$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}.$$

En général, on dit que deux nombres entiers a et b sont *congrus modulo* m, s'ils ont le même reste entre 0 et $m-1$ après division par m. Dans les contextes mathématiques, on écrit ceci

$$a \equiv b \pmod{m}. \tag{8}$$

La phrase $a \equiv b \pmod{m}$ se lit «a est congru à b modulo m». Dans un contexte informatique, «mod» est décrit comme une fonction et $(a \bmod m)$ calcule précisément le reste de la division par m. À la mode[5] informatique, la situation décrite par l'équation (8) s'écrit donc

$$(a \bmod m) = (b \bmod m).$$

Le lecteur vérifiera facilement que $20 \equiv 26 \pmod{6}$.

Pour l'instant, tout ceci n'est qu'un ensemble de notations. Les choses deviennent intéressantes lorsque en arrive à calculer modulo m. Autrement dit, on cherche à additionner, multiplier, soustraire, et parfois diviser, des éléments de $\mathbb{Z}_m$; de telle sorte que le résultat reste toujours dans $\mathbb{Z}_m$. Dans le cas de l'addition et de la multiplication, cela consiste simplement à

5. Sans jeu de mot.

effectuer l'opération habituelle, puis à calculer le reste du résultat pour la division par m. Ainsi, 9 fois 15, modulo 18, donne 9 ; puisqu'on calcule que $9 \cdot 15 = 135 \equiv 9 \pmod{18}$. On écrit ceci

$$9 \cdot 15 \equiv 9 \pmod{18}.$$

Un principe mathématique (qui nécessite une vérification que nous ne donnerons pas ici) affirme que, dans les calculs d'expressions complexes modulo m, on peut remplacer n'importe quel résultat intermédiaire par sa valeur modulo m. Un bon exemple (qui nous servira plus tard) est le calcul de la *puissance*. Ainsi on arrive à calculer rapidement à la main $2^{100} \pmod{10}$, simplement comme suit [6] :

$$\begin{aligned}
2^{100} &\equiv (2^4)^{25} \pmod{10} \\
&\equiv (16)^{25} \pmod{10} \\
&\equiv 6^{25} \pmod{10} \\
&\equiv 6 \cdot (6^2)^{12} \pmod{10} \\
&\equiv 6 \cdot 36^{12} \pmod{10} \\
&\equiv 6 \cdot 6^{12} \pmod{10} \\
&\equiv 6 \cdot (6^2)^6 \pmod{10} \\
&\equiv 6 \cdot 6^6 \pmod{10} \\
&\equiv 6 \cdot (6^2)^3 \pmod{10} \\
&\equiv 6^4 \pmod{10} \\
&\equiv 6 \pmod{10}
\end{aligned}$$

Il faut bien convenir que ceci est plus facile, et plus rapide, que de calculer que

$$2^{100} = 1267650600228229401496703205376,$$

pour ensuite prendre le reste de la division par 10. Pour la *soustraction* modulo m, une simple substitution transforme celle-ci en addition. En effet, $m - a$ joue exactement le rôle de $-a$ modulo m, puisque

$$a + (m - a) \equiv 0 \pmod{m}.$$

Le point ici, est que $m - a$ est un nombre dans $\mathbb{Z}_m$, quand a est entre 1 et $m - 1$. Si on remplace l'un par l'autre, on évite les nombres négatifs. Autrement dit, $m - a$ est *l'inverse additif* de a, modulo m. Pour calculer $a - b$, modulo m, on se ramène donc à calculer $a + (m - b)$, modulo m.

Très certainement l'opération la plus intéressante, mais aussi la plus délicate, est la division. Celle-ci n'est pas toujours possible. Comme pour la soustraction, on cherche à se ramener à

6. Rappelons à ce sujet les lois sur les exposants : $x^k x^n = x^{k+n}$ et $(x^k)^n = x^{k\,n}$.

une multiplication. L'essentiel est de savoir calculer a^{-1} dans les entiers modulo m, puisque

$$\frac{b}{a} = b \cdot \frac{1}{a} = b \cdot a^{-1}.$$

Si on peut trouver a^{-1}, dans les entiers modulo m, on dit que a est *l'inversible*, et que a^{-1} est l'inverse multiplicatif de y. Par exemple, on a

$$3 \cdot 9 = 27 \equiv 1 \pmod{26},$$

d'ou $3^{-1} \equiv 9 \pmod{26}$. En calculant tous les produits possibles, on trouve

a	1	3	5	7	9	11	15	17	19	21	23	25
a^{-1}	1	9	21	15	3	19	7	23	11	5	17	25

(9)

Il n'est pas toujours possible d'inverser modulo m. Cependant, la situation est facile lorsque m est un nombre premier comme 29. En effet, dans ce cas tous les nombres entre 1 et $m-1$ sont inversibles. Nous allons voir au chapitre **??** pourquoi c'est le cas, et comment effectuer ce calcul en général. Lorsque m est petit (< 100), il suffit d'effectuer quelques multiplications pour obtenir la table des inverses.

Retour sur le chiffrement par décalage

L'arithmétique modulaire que nous venons de décrire permet de mathématiser le chiffrement par décalage en le formalisant comme une addition dans $\mathbb{Z}_{26}$. On peut alors reformuler le code de César de la façon suivante. Supposons que nous remplacions les 26 lettres de l'alphabet par les entiers $0, 1, 2, \ldots, 25$ dans l'ordre habituel. C'est donc dire que $A \mapsto 0$, $B \mapsto 1$, et ainsi de suite jusqu'à $Z \mapsto 25$. Le codage de César qui décale chaque lettre de 3 positions peut maintenant être défini par la fonction $f : \mathbb{Z}_{26} \to \mathbb{Z}_{26}$

$$f(x) \equiv x + 4 \pmod{26}.$$

Ainsi, tout comme César, on obtient

$$\begin{aligned} A \mapsto 0 &\xrightarrow{f} 4 \mapsto D \\ B \mapsto 1 &\xrightarrow{f} 4 \mapsto E \\ &\vdots \\ Z \mapsto 25 &\xrightarrow{f} 3 \mapsto C \end{aligned}$$

Un soupçon de matrices

Nous allons travailler ici avec des *matrice* 2×2, c'est-à-dire les tableaux de nombres de deux lignes et deux colonnes, avec une petite excursion au cas plus général des matrices $2 \times k$. On considère aussi le cas des *vecteur*, qui sont simplement des matrices 2×1, c'est-à-dire avec une seule colonne. On ainsi les matrices

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \begin{pmatrix} 3 & 10 & 0 \\ 7 & -1 & 2 \end{pmatrix}.$$

Bien entendu la notion de matrice est plus générale, admettant m lignes et k colonnes. Mais la situation est assez amusante avec 2 lignes, et cela nous suffira pour l'instant. Nous allons aussi nous restreindre au cas où les *entrées* sont des entiers (ou parfois des nombres rationnels). Cela peut sembler un peu évident, mais deux matrices ne peuvent être égales que si elles ont la même forme et les mêmes entrées aux mêmes endroits. Ainsi, on ne peut avoir l'égalité

$$\begin{pmatrix} x & z \\ y & t \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

que si on a les 4 égalités

$$\begin{aligned} x &= a, \quad z = b, \\ y &= c, \quad t = d. \end{aligned}$$

Ce qui nous intéressera surtout, c'est le *produit de matrice*, qui prend les formes suivantes dans notre contexte :

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a\,x + b\,y \\ c\,x + d\,y \end{pmatrix}, \tag{10}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x & s \\ y & t \end{pmatrix} = \begin{pmatrix} a\,x + b\,y, & a\,s + b\,t \\ c\,x + d\,y, & c\,s + d\,t \end{pmatrix} \tag{11}$$

et plus généralement

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 & x_2 & \cdots & x_k \\ y_1 & y_2 & \cdots & x_k \end{pmatrix} = \begin{pmatrix} a\,x_1 + b\,y_1, & a\,x_2 + b\,y_2 & \cdots & a\,x_k + b\,y_k \\ c\,x_1 + d\,y_2, & c\,x_2 + d\,y_2 & \cdots & c\,x_k + d\,y_k \end{pmatrix} \tag{12}$$

Attention, le produit de matrices n'est pas comme le produit usuel de nombres. En particulier, les deux produits

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x & s \\ y & t \end{pmatrix} \qquad \begin{pmatrix} x & s \\ y & t \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{13}$$

sont presque toujours **différents**.

La *matrice identité*

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

joue un rôle analogue à celui du nombre 1 pour la multiplication des nombres. Ainsi, on a toujours

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x & s \\ y & t \end{pmatrix} = \begin{pmatrix} x & s \\ y & t \end{pmatrix} \tag{14}$$

Certaines matrices 2×2 ont un *inverse*. Par exemple l'inverse de

$$\mathbf{T} = \begin{pmatrix} 11 & 13 \\ 5 & 6 \end{pmatrix}$$

est la matrice

$$\mathbf{T}^{-1} = \begin{pmatrix} 6 & -13 \\ -5 & 11 \end{pmatrix}.$$

Cela signifie que

$$\mathbf{T}^{-1}\mathbf{T} = \begin{pmatrix} 6 & -13 \\ -5 & 11 \end{pmatrix} \begin{pmatrix} 11 & 13 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

D'autres matrices n'ont pas d'inverse, comme

$$\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}.$$

Il n'est pas difficile de montrer qu'une matrice

$$\mathbf{T} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

a un inverse si et seulement si son *déterminant*, $ad - bc$, est différent de zéro. L'inverse de $\mathbf{T}$ est alors donné par la formule

$$\mathbf{T}^{-1} = \begin{pmatrix} \dfrac{d}{ad-bc} & \dfrac{-b}{ad-bc} \\ \dfrac{-c}{ad-bc} & \dfrac{a}{ad-bc} \end{pmatrix} \tag{15}$$

Le cocktail des matrices modulo m

Pour nos applications, entre autres dans le cadre du système de Hill, nous avons besoin de calculer modulo m les produits de matrices, et les inverses de matrices. Pour le produit,

cela va presque de soi, il suffit de faire passer au modulo chaque entrée du résultat. Mais pour l'inverse, c'est un peu plus délicat. La formule (15) fonctionne tel quel pour les calculs modulo m, il faut cependant pouvoir «diviser par le déterminant». Une matrice n'est donc inversible que si son déterminant est inversible modulo m.

2.11 Chiffrement affine

On a déjà vu comment interpréter le chiffrement par décalage via le calcul modulo 26. Le *chiffrement affine* généralise ce genre de calcul. En effet, chaque lettre x de l'alphabet (considérée comme un nombre entre 0 et 25) est transformée en la lettre y donnée par

$$y = (a\,x + b \bmod 26).$$

Pour qu'on puisse décoder, il faut pouvoir calculer x à partir de y. Un petit calcul donne

$$x = (a^{-1}(y - b) \bmod 26),$$

Ceci montre qu'on ne peut utiliser un chiffrement de la forme

$$x \mapsto a\,x + b$$

que si a est inversible modulo m (voir Exercice 2.7). On dit alors que $f(x) = (a\,x+b \bmod 26)$ est un codage *admissible*. On utilise souvent des alphabets avec plus de lettres dans les chiffres affines. Il est alors plus facile de choisir un alphabet avec un nombre premier de lettres. On montrera au Chapitre **??** qu'un nombre a n'est inversible modulo m que si le plus grand commun diviseur de a et$m$ est 1. On dit alors que a et m sont *relativement premiers*.

Puisque chaque lettre est systématiquement remplacée par la même lettre, les chiffres affines correspondent à un cas spécial de chiffre par substitution mono alphabétique.

2.12 Exercices

2.1. (Koblitz) Dans les babillards électroniques, il est d'usage, lorsqu'on veut afficher un message offensant ou vulgaire, de le coder par décalage. Il est alors facile pour ceux qui le désirent de décoder le message. Dans un congrès international de chirurgiens, une équipe américaine affiche le message codé suivant.

BCIGOJCBGQCIGIIBGCIFWFSOI
QIZRIBQVSJOZSHZOBBSSGIWJO
BHSWZSHOWHSZIDFSGWRSBH

Déchiffrez (sans vous en offenser) ce message codé.

2.2. (ADFGVX) Le message Allemand intercepté par les Français, puis décrypté par Painvin, était

FGAXA XAXFF FAFVA AVDFA GAXFX
FAFAG DXGGX AGXFD XGAGX GAXGX
AGXVF VXXAG XDDAX GGAAF DGGAF
FXGGX XDFAX GXAXV AGXGG DFAGG
GXVAX VFXGV FFGGA XDGAX FDVGG A

Sachant que la grille trouvée par Painvin est celle en (7), et que la permutation est

$$(12, 6, 18, 15, 4, 1, 3, 16, 10, 8, 19, 14, 11, 7, 9, 2, 5, 21, 17, 20, 13),$$

décoder le message ci-haut (écrit en Allemand, natürlich).

2.3. (Arithmétique modulaire) Supposons qu'aujourd'hui nous soyons mardi et que ni cette année ni l'an prochain ne soient une année bissextile. Quel jour de la semaine serons-nous dans exactement un an ?

2.4. (Arithmétique modulaire) Un jeu d'enfant consiste à placer 5 points $\{a, b, c, d, e\}$ à distances égales sur un cercle puis à essayer de dessiner une étoile à 5 sommets en ne visitant chaque sommet qu'une seule fois chacun comme sur la figure 2.2

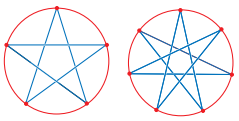


FIGURE 2.2 – Étoiles à 5 et 7 sommets

a) Trouvez toutes les façons de dessiner l'étoile à 5 sommets en partant du sommet A.
b) Pouvez-vous dessiner une étoile à 6 sommets de cette façon ? Pourquoi ?
c) Pouvez-vous dessiner une étoile à 8 sommets de cette façon ? Pourquoi ?
Solution.

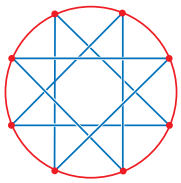


FIGURE 2.3 – Étoile à 8 sommets

2.5. (Chiffre affine) Codez le message en convertissant les lettres en nombres, en appliquant la fonction de codage affine donnée, et en reconvertissant les nombres en lettres.

LA MARIÉE IRA MAL

a) $f(x) = (x + 3 \mod 26)$ (le codage de César)

b) $f(x) = (3x + 11 \mod 26)$

2.6. (Arithmétique modulaire) Verifier qu'on a bien la table (9), pour l'inversion modulo 26.

2.7. (Chiffre affine) Codez les mots $ANNA$ et $NANA$ avec la fonction affine

$$f(x) = (6x + 5 \mod 26).$$

Qu'observez-vous ? Pouvez-vous expliquer ?

2.8. (Chiffre affine) Les fonctions de codage affines suivantes sont-elles admissibles ? dites pourquoi.

a) $(3x + 10 \mod 26)$, b) $(11x + 20 \mod 26)$, c) $(10x + 20 \mod 26)$.

2.9. (Chiffre affine) Trouvez (si possible) la fonction de décodage de chacune des fonctions de codage affine

a) $(3x + 10 \mod 26)$, b) $(11x + 20 \mod 26)$, c) $(10x + 20 \mod 26)$.

2.10. (Chiffre affine) Le message suivant a été crypté avec le chiffre affine $(7\,x+5 \mod 26)$:

$$EHTCFIFLFSVHEFBZPUJB.$$

Décryptez le.

2.11. (Chiffrement affine) Combien existe-t-il de chiffrements affines admissibles de la forme :

$$x \mapsto (a\,x + b \bmod 26)$$

Autrement dit, de combien de façons peut-on choisir a et b, pour que le codage puisse être décodé correctement (Voir l'exercice 2.7).

2.12. On a encodé un bigramme d'un texte en clair, à l'aide d'un codage de Hill, en utilisant la matrice

$$\mathbf{M} = \begin{pmatrix} 2 & 3 \\ 7 & 8 \end{pmatrix}.$$

Le résultat obtenu est CF. Retrouver le bigramme en clair.

2.13. (Chiffre de Hill) Dans le but d'augmenter la difficulté de cryptanalyse de votre codage, vous décidez d'utiliser le chiffre de Hill dans un alphabet à 29 symboles en appliquant d'abord la matrice

$$\begin{pmatrix} 3 & 11 \\ 4 & 15 \end{pmatrix} \pmod{29}$$

puis la matrice

$$\begin{pmatrix} 10 & 15 \\ 5 & 9 \end{pmatrix} \pmod{29}$$

a) Encodez le message ENVOYEZ

b) Expliquez comment déchiffrer un texte codé par l'application de deux matrices successives.

2.14. (Chiffre de Vernam) Coder le message suivant selon le chiffre de Vernam

NE DITES PAS A DIEU CE QU'IL DOIT FAIRE[7]

en utilisant la clé aléatoire suivante :

WAWPVRQQMBSRFVSHVBHDPVTLDDPMQS

Chapitre 3

Cryptanalyse des systèmes classiques

3.1 Introduction

Pour briser un cryptosystème, un opposant cherche a obtenir deux éléments d'information :

1. Quel est le type de système de codage utilisé ? et,
2. Quelle est la clé d'encodage utilisée ?

Bien entendu, son travail est simplifié (mais certainement pas terminé) s'il connaît le type de système utilisé. Nous allons presque toujours supposer que ce type est connu. Cela est assez réaliste, puisqu'en pratique un usager de la cryptographie tend à n'utiliser qu'un seul type de système. Avec le temps cette information finit par circuler. Cette hypothèse de travail est appelée le *principe de Kerckhoffs*[1]. Ce principe consiste à affirmer que la sécurité d'un système de chiffrement ne devrait pas être fondée sur le secret de la procédure utilisée, mais essentiellement sur le secret de la clé. Du point de vue de l'utilisateur, c'est un bon principe de prudence. S'il croit que l'opposant peut difficilement briser son système, même s'il en connaît la nature (mais pas la clé secrète), l'utilisateur considérera que son système est sur.

Nous avons déjà vu que les codes par décalages sont faciles à briser de ce point de vue. Le but de la cryptanalyse est, soit de montrer qu'on peut briser un système donné, soit de montrer que le système est impossible à briser. Ce sont les questions que se posent naturellement, chacun pour ses raisons, autant l'utilisateur d'un système, qu'un opposant qui cherche à

1. Selon Auguste Kerckhoffs, *La cryptographie militaire*, 1883.

mettre à jour les secrets de cet utilisateur. Les informations dont peut disposer un opposant sont diverses. Elles dépendent en quelque sorte de ses qualités d'espion. La situation de base consiste à supposer que l'opposant à intercepter un ou plusieurs messages codés. Mais il se pourrait aussi qu'il ait en plus intercepté le message en clair de quelques un de ces messages codés.

Nous allons commencer notre exploration de la cryptanalyse en essayant de décoder des textes que nous savons avoir été codés par substitution de l'alphabet.

3.2 Cryptanalyse des systèmes mono alphabétiques

Nous nous attaquons d'abord à la cryptanalyse des chiffrements mono alphabétiques, c'est-à-dire les chiffrements obtenus par une permutation des lettres de l'alphabet. Cette analyse est basée sur la fréquence d'apparition des lettres dans le texte en clair. Cela est particulièrement facile pour les chiffres à la César, et nous en avons déjà discuté. De même, pour les chiffres affines, il y a des méthodes spéciales qui sont très efficaces. Cependant, comme il s'agit d'un cas spécial d'une substitution d'alphabet, on peut aussi utiliser les méthodes que nous allons maintenant discuter.

On a déjà vu qu'il y a

$$403,291,461,126,605,635,584,000,000$$

permutations possibles des lettres d'un alphabet à 26 lettres. Chacune de ces permutations constitue une clé possible pour un chiffre par substitution. Il semble donc désespéré de vouloir la retrouver. On peut cependant s'attaquer très efficacement à n'importe quel chiffre par substitution, en exploitant le fait que le message codé devra forcément respecter la «forme» du texte original, qu'on suppose ici français.

Une première observation est qu'en français, toutes les lettres n'ont pas la même fréquence d'apparition. Ainsi, il y a en général bien plus de lettres [2] e que de lettres z. Dans un chiffrement par substitution, la lettre e est toujours remplacée par la même lettre, de même que le z d'ailleurs. Il y a donc de fortes chances que la lettre, que l'on retrouve le plus fréquemment dans le texte chiffré, corresponde au codage du e. Le tableau 3.1 (et le tableau 3.3 de l'Appendice 3.8 sous une forme plus visuelle) donne la distribution de fréquence des lettres d'un texte français «typique». Elle a été calculée à partir de la distribution de

2. Pour les besoins de la présentation de cette section, on utilise les lettres minuscules pour le texte en clair, et les majuscules pour le texte codé. Certains textes seront partiellement décodés, on y trouvera donc un mélange de minuscules et de majuscules.

lettres dans l'Encyclopedia Universalis [3]. Évidemment le résultat pourrait varier un peu avec le choix des textes de référence, mais pour des textes assez longs les distributions se ressemblent (Attention tout de même aux textes comme celui de l'exercice 3.1).

	a	b	c	d	e	f	g	h	i	j	k	l	m
%	8,40	1,06	3,03	4,18	17,26	1,12	1,27	0,92	7,34	0,31	0,05	6,01	2,96
	n	o	p	q	r	s	t	u	v	w	x	y	z
%	7.13	5.26	3.01	0.99	6.55	8.08	7.07	5.74	1.32	0.04	0.45	0.30	0.12

TABLE 3.1 – Fréquences d'apparition des lettres dans un texte français

On y observe que certaines lettres ont des fréquences très semblables. À la lumière de notre observation sur le fait que le calcul de fréquences peut varier un peu, selon le choix des textes de références, il est naturel de procéder aux regroupements suivants. Après le e, qui est nettement la plus fréquente, les 5 lettres

$$a,\ s,\ i,\ n,\ t$$

ont des fréquences assez proches allant (en ordre décroissant) de $8,40\%$ à $7,07\%$. Elles sont donc difficiles à distinguer au moyen d'un simple calcul de fréquence. Cependant elles se démarquent du groupe suivant composé des lettres

$$r,\ l,\ u,\ o$$

avec des fréquences allant de 6,55% à 5,26%. On a ensuite le groupe

$$d,\ c,\ p\ ,m$$

avec des fréquences allant de 4,18% et 2,96%, et enfin

$$b,\ f,\ g,\ h,\ j,\ k,\ q,\ v,\ w,\ x,\ y,\ z$$

toutes de fréquence de moins de 1,32%. Pour raffiner notre analyse, et distinguer entre elles les lettres d'un même groupe, on étudie la distribution de fréquence des groupes de deux lettres (appelé *bigrammes*). La distribution des bigrammes dans un texte français est donnée à l'Appendice 3.8. Les bigrammes les plus fréquents sont

$$es,\ de,\ le,\ en,\ re,\ nt,\ on,\ er,\ te,\ el,\ \ldots$$

Ainsi, la lettre s du deuxième groupe groupe en fréquence, ressort ici clairement en association avec la lettre e. Par opposition, la lettre a apparaît rarement en association avec

3. Voir http ://www.apprendre-en-ligne.net/crypto/menu/

la lettre e. Une autre distribution intéressante est celle des lettres doubles, dont les plus fréquentes sont

$$ee,\ ss,\ ll,\ tt,\ nn,\ mm,\ rr,\ pp,\ ff,\ \cdots$$

On peut ensuite passes à l'étude de la distribution des *trigrammes* les plus fréquents

$$ent, les, ede, des, que, ait, \ldots$$

et ainsi de suite. L'étude des «motifs» comme *ede* est parfois très efficace.

Illustrons comment procéder à une analyse de fréquence. On commence par observer la distribution de fréquence

X	R	Y	V	M	B	J	G	C	U	N	A	Q	L	...
242	109	104	104	102	99	94	88	80	69	54	43	35	32	...

pour les lettres du texte à déchiffré, ici celui de la Figure 3.1, qu'on sait avoir été codé par une substitution mono alphabétique. Comme le X apparaît beaucoup plus souvent que

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

FIGURE 3.1 – Un texte codé par une substitution mono alphabétique

toutes les autres lettres, on suppose qu'il correspond au e dans le texte en clair. La seconde lettre en fréquence est le R. Le R devrait correspondre au s du texte clair, puisque que le bigramme le plus fréquent est XR (49 fois) et que le doublet RR apparaît assez fréquemment (7 fois). On peut ensuite tenir compte de la fréquence des bigrammes, surtout ceux de la forme $_X$ et $X_$:

XR	GX	XY	VX	XJ	NX	XV	CY	VC	...
49	37	34	32	29	26	25	23	23	...

et des doubles lettres :

XX	LL	RR	JJ	GG	YY	II	QQ
14	7	7	6	5	4	3	2

pour guider nos prochains choix. Il est efficace, à partir de maintenant, de remplacer les lettres codées par celle que nous pensons devoir leur être substituées. Le Y et le V ont tous deux la même troisième plus grande fréquence, et celle de M est très proche. D'autre part, après e et s, les lettres a, n et t sont les plus fréquentes en français, toutes avec des taux d'apparitions assez proches. Cependant, toujours dans des textes français, on a peu souvent le bigramme ea ; de plus, les bigrammes aa et ae sont assez rare. Hors, dans notre texte codé, eY est dans les bigrammes les plus fréquents, ce qui est d'ailleurs le cas de en en français. On est donc poussé à choisir n comme substitut de Y. Nous sommes maintenant dans la situation suivante :

$$\begin{array}{ccc} X & R & Y \\ \downarrow & \downarrow & \downarrow \\ e & s & n \end{array}$$

La lettre M du texte codé possède des caractéristiques semblables au a du français, en ce qui concerne sa fréquence et son implication dans des bigrammes. On pose donc encore $M \mapsto a$. En observant que le trigramme XYJ est le plus fréquent dans notre texte chiffré, tout comme ent dans textes en clair, on choisit $J \mapsto t$. Cela semble confirmé par le fait que la double lettre JJ apparaît avec une fréquence semblable à celle de tt en français. On pose encore $B \mapsto i$, puisque tout comme i, c'est la sixième lettre la plus fréquente, et que les doubles lettres BB et ii sont tous peu fréquente dans leurs contextes respectifs. Par une analyse semblable, on choisit encore $V \mapsto r$. On en est donc rendu à

$$\begin{array}{ccccccc} X & R & Y & M & J & B & V \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ e & s & n & a & t & i & r \end{array}$$

Le début de notre texte partiellement décodé est donc :

e, n, A, e, t, i, n, s, t, a, n, t, a, Q, Q, a, r, U, r, e, n, t, G, e, s, N, C, i, W, t, s, N, U, n, e, L, a, i, n, N, P, C, L, L, e, e, t, i, G, s, e, A, r, i, D, i, r, e, n, t, e, n, I, a, A, e, N, U, A, a, n, N, e, G, a, F, r, e, s, U, r, G, e, Q, G, a, t, r, e, N, U, L, U, r, N, U, Q, a, G, a, i, s, r, C, E, a, G, G, e, r, C, i, D, i, t, A, e, t, t, e, Q, a, r, t, i, e, N, e, L, a, i, n, K, U, i, e, A, r, i, D, a, i, t, a, G, C, r, s, G, e, r, C, i, A, P, a, n, W, e, a, N, e, A, C, U, G, e, U, r, s, e, s, Q, e, n, s, e, e, s, G, e, I, I, r, a, E, e, r, e, n, t, G, e, s, S, C, i, n, t, U, r, e, s, N, e, s, e, s, r, e, i, n, s, s, e, N, e, G, i, e, r, e, n, t, e, t, s, e, s, W, e, n, C, U, Z, s, e, P, e, U, r, t, e, r, e, n, t, G, U, n, G, a, U, t, r, e, G, e, r, C, i, A, r, i, a, a, D, e, A, I, C, r, A, e, Q, C, U, r, I, a, i, r, e, D, e, n, i, r, G, e, s, L, a, W, i, A, i, e, n, s, G, e, s, A, P, a, G, N, e, e, n, s, e, t, G, e, s, a, s, t, r, C, G, C, W, U, e, s, G, e, r, C, i, Q, r, i, t, G, a, Q, a, r, C, G, e, e, t, N, i, t, a, U, Z, s, a, W, e, s, N, e, F, a, F, E, G, C, n, e, t, C, U, t, P, C, L, L, e, K, U, i, G, i, r, a, A, e, t, t, e, e, A, r, i, t, U, r, e, e, t, L, ...

À ce stade, on peut presque lire le texte. Cela suggère d'autres substitutions comme $A \mapsto c$, $Q \mapsto p$ et $U \mapsto u$. Chaque nouvelle substitution rend le texte plus clair et suggère de nouvelles substitutions. Finalement le texte en clair est :

en cet instant apparurent les doigts d'une main d'homme et ils écrivirent en face du candélabre sur le plâtre du mur du palais royal le roi vit cette partie de main qui écrivait alors le roi changea de couleur ses pensées l'effrayèrent les jointures de ses reins se délièrent et ses genoux se heurtèrent l'un l'autre le roi cria avec force pour faire venir les magiciens les chaldéens et les astrologues le roi prit la parole et dit aux sages de Babylone tout homme qui lira cette écriture et me fera connaître son explication revêtira la pourpre mettra le collier d'or a son cou et comme troisième dans le royaume il commandera alors vinrent tous les sages du roi mais ils ne purent pas lire l'écriture et faire connaître au roi l'explication le roi Balthasar fut donc très effraye la couleur de son visage changea et ses grands furent bouleverses la reine en raison des paroles du roi et de ses grands vint dans la salle du festin la reine prit la parole et dit que le roi vive éternellement que tes pensées ne t'effrayent pas et que ton visage ne change pas de couleur il y a dans ton royaume un homme qui possède en lui l'esprit des dieux saints pendant les jours de ton père une lumière un discernement et une sagesse comme la sagesse des dieux furent trouves en lui et le roi nabuchodonosor ton père l'établit comme chef des devins des magiciens des chaldéens et des astrologues parce qu'un esprit supérieur une intelligence un discernement l'explication des songes l'interprétation des énigmes la solution des problèmes furent trouves en lui en Daniel a qui le roi avait donne le nom de Beltshassar que daniel soit donc appelé et il fera connaître l'explication.

Tout ce processus peut être grandement automatisé assez facilement. De simples outils informatiques, alliés interactivement avec le décodeur, rendent le tout rapide et aisé.

3.3 L'écriture automatique

Pour mieux comprendre pourquoi l'analyse de fréquence est aussi efficace, il est amusant de considérer la différence entre l'écriture au hasard, et les textes de la littérature française. D'autre part, cette comparaison nous amènera naturellement à une approche générale au problème de la cryptanalyse via la théorie de l'information [4]. Le mathématicien (et ensuite politicien) Émile Borel [5] suggère l'image très colorée suivante :

> *.. Concevons qu'on ait dressé un million de singes à frapper au hasard sur les touches d'une machine à écrire et que, sous la surveillance de contremaîtres illettrés, ces singes dactylographes travaillent avec ardeur dix heures par jour avec un million de machines à écrire de types variés. Les contremaîtres illettrés rassembleraient les feuilles noircies et les relieraient en volumes. Et au bout d'un an, ces volumes se trouveraient renfermer la copie exacte des livres de toute nature et de toutes langues conservés dans les plus riches bibliothèques du monde. Telle est la probabilité pour qu'il se produise pendant un instant très court, dans un espace de quelque étendue, un écart notable de ce que la mécanique statistique considère comme le phénomène le plus probable...*

Depuis, cette image a été reprise par plusieurs sous toute sorte de formes. Nos amis les anglophones remplacent souvent les oeuvres littéraires françaises par les sonnets de Shakespeare. Des auteurs comme Borges l'ont aussi adaptée à leur imaginaire particulier.

Cependant, la très très grande majorité du travail de ces singes ne contiendra que des mots vides de sens, de toute sorte de longueurs, avec une ponctuation pour le moins anarchique. Pour s'en faire une idée, voici 100 mots de longueur 4 générés tout à fait au hasard. À chaque étape, une lettre a la même probabilité de 1/26 d'être choisie.

4. Ceci sera discuté au chapitre 5.

5. Dans un article intitulé *Mécanique Statistique et Irréversibilité*, paru en 1913 dans la revue J. Phys. 5e série, vol. 3, pp.189-196.

```
KHMC ITAY TJUY SNLB GAHQ ANXC WYZB YPSX VAQG APZM
DEQA RNAV QGXE WVUT TRIE UKPE RVYC FQNJ INEG NSQN
WLWL YUVF UJNM PUNM ZDLT MFYK FTMW WWSK MRNS MLSG
VNCE EVMH OYLW SVAX JOJH FXGP FZIX DXQX WKRB GYFK
ROGX HPYQ TGEN XPEI LQJB UQSK FXMR AHOZ DVCH HXCO
JAMR HTDW SWAJ MUMF YLNL DPPB MMRZ INJF KRIW GJQJ
OXTH STSJ KMPG NVKC HOLJ FTOY MPKJ RJDI RELB ZEOY
GUAB WTAY CLLC EBYY AUUX OGPZ CTRG IWVD NLCZ BXHX
FULR EPKU FGSH EFHB YAMS NMMA ZSHM AUPD YZFQ OLNA
YWUG LORW ESEU FVOE XIHW UHEK DTYS XTCF AZCP MBXM
```

Aucun de ces mots n'a de ressemblance avec un mot connu. Si on permet plutôt aux singes de choisir les lettres en respectant la distribution de fréquence des lettres de l'anglais (tentant de reproduire un sonnet de Shakespeare), on engendre des mots comme :

```
NEIO YSRS XRLP HSDS NHXB FNOL YJUJ OKHH EMOH AKRY
LQFX TNNK OWHT WRVB UJVX MVFJ PHBC EPTT FEJQ UYCZ
HXAV ZSXP OEHT RKLL HDCK YPIB MOUU ADCW NERW SLDU
VBJB DLRU ECCT PENZ DVGH AENA XKJB QBMK BPVS FKPT
ICBZ FJNW LEMA ENIP MHUR QIJA DYSH LOND CHJI MUFS
VZRJ DGVN KKMY ARWX WFQY LBQF RXSX FYCS DWHA JZJH
VWLG ZDJJ TFYE GKEK IECH PVIE BTDB ZRFL HOZG CMBF
CHOC PWZC ROPR GZYN USLA UYRS CLSR NXCL FPUP SMHN
NKUY XMRU XPMO SMSU FJYX SUSG BOHJ IBTI HKTC BKSE
SGIF GVGF ANEF FRTE AJTE SOTU SZJB AORU LPWH LGLQ
```

Ceux-ci ressemblent un peu plus aux mots anglais. Un seul mot est vraiment intelligible, mais en français plutôt qu'en anglais. Pour aider encore plus les singes, on fait en sorte qu'après avoir choisi la première lettre selon la distribution de l'anglais, les lettres subséquentes sont choisies selon la distribution de fréquence des bigrammes de l'anglais. On obtient des mots qui sont encore plus proches des mots anglais :

```
PENT FOPT BELA SEAL CRER CRAT AMAL ATME BANO TINI
FTOS BEIE SENT FRES EMET AREA PRAR TRIR ATIO PRST
YENG PITH BORE LATI MTES TINE RESE SORS TEAN PELE
INDE OSHO ONIO UNOU REAT AREN GANE WRAC LEST CURE
NDIN WINT TISE TINE TOWE WEER NDES ASER ITHT IONT
OALO THAN FITH MONA OMON TENT THIN POFE FITH ONAS
TORO TONE STHE SAER ITHE TUNT THIN ARAS RIAT ATHA
ATIR PONT GENT TETE SMER TINE WENE HETT ICOT BETA
ANEM WSUT FUNS OLES ONDA TONT NSTH TORI TIEN NDOT
TIEA PRCE ATIN HARN GANE TTOU TINE MAIS AREN WNIE
```

Les mots de l'anglais et du français (et autres langues) sont donc très structurés. De plus, dans un texte, on sait bien que les phrases sont organisées selon les règles grammaticales. On est donc bien loin de lettres choisies au hasard, et c'est exactement ce qu'exploite le cryptanalyste pour briser les systèmes de codage par substitution. Nous reviendrons sur cette façon de voir les choses au chapitre sur la théorie de l'information.

3.4 Cassage du chiffre de Vigenère

Charles Babbage (1791-1871)

Le système cryptographique de Vigenère résista pendant trois siècles jusqu'à ce que le mathématicien britannique Charles Babbage mette au point une méthode de décryptage vers 1854. Cette découverte assura à l'Angleterre une suprématie militaire lorsqu'elle déclara la guerre à la Prusse cette même année. La difficulté vient du fait que le codage est *poly alphabétique*, c'est-à-dire que la même lettre peut être remplacée par diverses lettres.

Trouver la longueur de la clé

L'idée de Babbage est fondée sur l'observation que si le mot clé est de longueur p, alors deux lettres à distance p dans le texte en clair subissent le même décalage. Si on regroupe ensemble toutes les lettres obtenues en faisant des sauts de longueur p dans le texte chiffré, la distribution de fréquence des lettres de cet ensemble sera la même que celle de l'ensemble des lettres correspondantes dans le texte en clair et que dans tout texte de la langue française. On pourra donc utiliser l'analyse e fréquence sur des sous-ensembles de lettres comme pour la substitution mono alphabétique.

La première étape pour briser un texte codé avec le chiffre de Vigenère consiste donc à trouver la longueur p du mot clé. On essaie diverses valeurs pour cette longueur p, en comparant chaque fois l'histogramme des fréquences des lettres aux rangs

$$1,\ p+1,\ 2p+1,\ 3p+1,\ \ldots$$

dans le texte codé, à la fréquence des lettres du français pour différentes valeurs de p. Une mauvaise valeur de p donne généralement un histogramme plutôt uniforme contrairement à la fréquence des lettres dans un texte français. On a de grandes chances d'avoir la bonne valeur de p lorsque l'histogramme des fréquences observées ressemble à une version décalée de l'histogramme des fréquences des lettres du français. On obtient alors la première lettre du mot clé puisqu'elle correspond au décalage de l'histogramme. Maintenant qu'on connaît p, la suite est facile. Il suffit de trouver le décalage pour les lettres aux rangs

$$2,\ p+2,\ 2p+2,\ 3p+2,\ \ldots$$

puis pour les lettres aux rangs

$$3,\ p+3,\ 2p+3,\ 3p+3,\ \ldots$$

et ainsi de suite.

Un exemple de cassage

Essayons de décoder le texte suivant chiffré avec le système de Vigenère.

```
SBAGF PLLGA ABVPN LVHLE GKAGU SEICF SICYQ NGSHT UHRWR
HRJWL LWGZT BPHVT KAIEE OOBUP JNUAE FGUTV GATXO HILQU
LTELT MGTAB UVNBU VLXGH UIKLD WGSAV QSLBP LEGDV RWWYE
 WWIOB USAFC PSHPX UTNSA BGUTV JVILK YANUZ IEGZC ACAS
```

On commence par déterminer la longueur p de la clé, en progressant dans les longueurs à partir de $p = 1$. Pour $p < 4$, on trouve des distributions de fréquences avec peu de «variabilité». Soudain, pour $p = 4$, on constate que les lettres de rang 1, 5, 9, 13 etc, ont la distribution de fréquence donnée en rouge dans la partie gauche de la figure 3.2. Cette distribution est semblable à une version décalée de l'histogramme en bleu, qui est celui du français. On constate aussi qu'il suffit de le décaler de deux positions vers la gauche, comme dans la partie droite de la figure 3.2, pour le superposer à l'histogramme en bleu de la meilleure façon possible. Si la similarité satisfaisante, alors la première lettre du mot clé est C. On passe alors aux lettres de rang 2, 6, 10, etc ; et un décalage de 7 positions

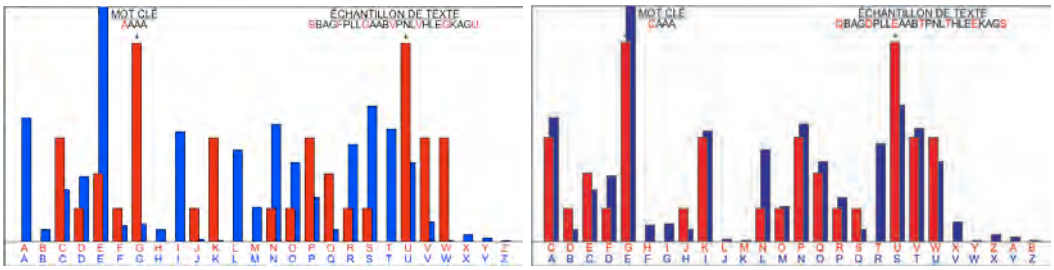


FIGURE 3.2 – Déterminer la longueur de la clé, et le premier décalage.

semble donner la plus grande similitude avec l'histogramme bleu. La seconde lettre du mot clé semble donc être H. On continue ainsi pour trouver enfin le mot clé *CHAT*, en même que le texte clair :

quand il s'était installé dans le pays par quel hasard par quel
destin avait-il choisi justement cette maison-là
cette maison isolée au pied de la colline en bordure du bois
la maison qu'allaient choisir aussi les chats

3.5 L'indice de coïncidence

Si on veut comparer de façon plus rigoureuse l'histogramme du français et l'histogramme obtenu en faisant des sauts de$k$ lettres, on utilise l'indice de coïncidence. Cet indice donne une valeur numérique à la similitude des deux histogrammes. Ce concept fut mis au point par le cryptologue américain William Friedman [6] qui le publia en 1920, et il est utilisé pour déterminer la longueur du mot clé dans un codage de Vigenère. Pour une explication de la définition qui suit, voir la section 4.6. Dans un texte quelconque de n lettres, on compte le nombre de répétitions de chaque lettre :

$$\begin{array}{c} n_A = \text{ nombre de } A \text{ dans le texte} \\ n_B = \text{ nombre de } B \text{ dans le texte} \\ n_C = \text{ nombre de } C \text{ dans le texte} \\ \cdots \\ n_Z = \text{ nombre de } Z \text{ dans le texte} \end{array}$$

On calcule *l'Indice de coïncidence* simplement par la formule :

$$IC := \frac{n_A(n_A-1) + n_B(n_B-1) + \ldots + n_Z(n_Z-1)}{n(n-1)} \tag{1}$$

À titre d'exemple, calculons l'indice de coïncidence du texte :

Un enfant n'a pas d'aversion pour la laideur de sa mère

le nombre de lettres dans cette phrase est $n = 43$, le nombre de a est 7, le nombre de b et de c est zéro, le nombre de d est 3, etc. L'indice de coïncidence est donc

$$IC = \frac{(7\times 6) + 0 + 0 + (3\times 2) + \ldots}{43\times 42} = 0,070$$

6. William Friedman et son épouse Elizabeth sont connus pour avoir réfuté la théorie selon laquelle Francis Bacon serait l'auteur des pièces de William Shakespeare

Les spécialistes de l'analyse de textes français ont calculé que *l'indice de coïncidence moyen en français* est

$$IC_f = 0,074.$$

L'indice de coïncidence d'un texte écrit en français est plus grand que l'indice de coïncidence d'un texte où chaque lettre est choisie aléatoirement et où toutes les lettres ont la même probabilité d'apparition. Un petit calcul (voir section 4.6) nous permet d'obtenir facilement l'indice de coïncidence moyen d'un texte aléatoire

$$IC_a = 0,038$$

On voit ainsi qu'il existe une différence appréciable entre l'indice de coïncidence d'un texte français et l'indice de coïncidence d'un texte purement aléatoire. C'est cette différence entre l'indice de coïncidence d'un texte français et celui d'un texte aléatoire qui est exploité par les cryptanalystes. Voici quelques propriétés intéressantes de l'indice de coïncidence.

Observations sur l'indice de coïncidence

1. Pour tout chiffre mono alphabétique, la somme des distributions de fréquence des lettres n'est pas perturbée par le codage et l'indice de coïncidence est le même pour le texte codé que pour le texte en clair.

2. Si l'indice de coïncidence d'un codage d'un texte français chiffré est beaucoup plus petit que $0,074$, le chiffre est probablement poly alphabétique.

Test de Friedman

On peut utiliser l'indice de coïncidence pour déterminer la longueur de la clé dans un texte, codé selon le chiffre de Vigenère, de la façon suivante. Pour déterminer la longueur de la clé d'un chiffre poly alphabétique à partir d'un texte chiffré , on calcule d'abord l'indice de coïncidence de chacun des sous-ensembles de lettres suivants du texte chiffré :

1. l'ensemble de toutes les lettres du texte
2. l'ensemble des lettres en position 1,3, 5, ..., dans le texte
3. l'ensemble des lettres en position 1,4, 7, ..., dans le texte

⋮

k. l'ensemble des lettres en position 1,$k+1$, $2k+1$, ..., dans le texte

⋮

Si l'ensemble, considéré à la k-ième étape, est celui pour lequel l'indice de coïncidence est le plus élevé, alors on choisit k comme longueur du mot clé. Ainsi, pour trouver la longueur du

RFITG RJIAY FILRU YNIFT THGPQ VMRPX FCOAF VJJJV PZNKE
YGWTW FVEMJ ISRVP JJSBN HFKSS TESTA VUUZZ SRVVT LVRPX
JETRT VNKOV TIGII GCRSZ QHGSZ HUVXM JEDEC MJETN UCYIO
HXIWR VREPJ LRFCV RVSBW WFESY GYWJA EOIXA IAXMY VLRUM
SXEAK IZISR VPJJO HXVNV RFUTJ TINNM XKEFF IXZNQ WWYII
RUHFI MROIS KQHKW JKRBW ZJETR PXJIR VVSNI EOTMY RNAKU
ZVOHS YNMIR PHWRI RPXFJ YGTSZ MEECW JDEGV VJVNE CTUFR
GCZJT MBK

FIGURE 3.3 – Un texte codé par le chiffre de Vigenère

mot-clé du texte de la figure 3.3, codé avec le système de Vigenère, on calcule les indices de coïncidence pour les différents sous-ensembles correspondant à chaque étape, pour obtenir les résultats donnés au tableau 3.2. La clé est donc probablement de longueur 6.

Période	1	2	3	4	5	**6**	7	8	...
IC	0,0457	0,0471	0,0689	0,0497	0,0534	0,1222	0,0342	0,0415	...

TABLE 3.2 – Indices de coïncidence pour les diverses étapes

3.6 Briser un codage de Hill

Supposons qu'on ait obtenu un message codé avec un codage de Hill, et qu'une partie du texte en clair correspondant à ce message est connue. Par exemple, on a déduit que la fin de ce message est $HITLER$. Nous allons montrer qu'avec peu d'information, on peut retrouver la matrice d'encodage d'un chiffre de Hill et ainsi décoder tout le reste du message.

Le message codé suivant a été intercepté et on sait qu'il a été codé avec un code de Hill modulo 29.

!IEJYXJTORPAN !OEJSXB !FSIBUSN,LQQT !

Pour déchiffrer le message, on cherche à reconstruire la matrice de codage $\mathbf{T}$ à partir de l'information à notre disposition. Comme on sait que les dernières lettres du message en clair correspondent à la salutation fanatique de nos adversaires, les 4 dernières lettres sont

$TLER$. Les bigrammes codés QQ et $T!$ correspondent au bigrammes en clair TL et ER. C'est donc dire que la matrice $\mathbf{T}$ est telle que,

$$\mathbf{T}\begin{pmatrix}19 & 4\\ 11 & 17\end{pmatrix} = \begin{pmatrix}16 & 19\\ 16 & 27\end{pmatrix} \tag{2}$$

puisque la numérisation des bigrammes en clair TL et ER, donne $\begin{pmatrix}19\\11\end{pmatrix}$ et $\begin{pmatrix}4\\17\end{pmatrix}$; et celle des bigrammes codés QQ et $T!$, donne $\begin{pmatrix}16\\16\end{pmatrix}$ et $\begin{pmatrix}19\\27\end{pmatrix}$. On peut «isoler» la matrice $\mathbf{T}$, dans l'équation (2), pour obtenir

$$\mathbf{T} = \begin{pmatrix}16 & 19\\ 16 & 27\end{pmatrix}\begin{pmatrix}19 & 4\\ 11 & 17\end{pmatrix}^{-1}$$

parce que la matrice qui l'accompagne est inversible modulo 29. En effet, on calcule

$$\begin{pmatrix}19 & 4\\ 11 & 17\end{pmatrix}^{-1} \equiv \begin{pmatrix}9 & 3\\ 1 & 22\end{pmatrix} \pmod{29}$$

On calcule donc que

$$\mathbf{T} = \begin{pmatrix}18 & 2\\ 26 & 4\end{pmatrix},$$

en ensuite son inverse modulo 29 :

$$\mathbf{T}^{-1} = \begin{pmatrix}6 & 26\\ 19 & 27\end{pmatrix}.$$

On peut alors décoder le message au complet, pour trouver

$$WE,ARE,READY,TO,ATTACK,HEIL,HITLER$$

3.7 Exercices

3.1. Que remarquez-vous d'étrange dans l'extrait suivant du texte *La disparition* de Georges Perec :

> *Tout avait l'air normal, mais tout s'affirmait faux. Tout avait l'air normal, d'abord, puis surgissait l'inhumain, l'affolant. Il aurait voulu savoir où s'articulait l'association qui l'unissait au roman : sur son tapis, assaillant à tout instant son imagination, l'intuition d'un tabou, la vision d'un mal obscur, d'un quoi vacant, d'un non-dit : la vision, l'avision d'un oubli commandant tout, où s'abolissait la raison : tout avait l'air normal mais...*

3.8 Appendice : Fréquences de n-grammes

Les tableaux de cet appendice donnent les distributions de fréquences de n-grammes dans «les textes[7] français». Pour $n = 1$, c'est la distribution de fréquence des lettres ; $n = 2$, celle des bigrammes ; et $n = 3$, celle de trigrammes. En plus de donner un tableau partiel pour la distribution de fréquence des doubles lettres, on donne un tableau comparatif pour la distribution de fréquences des lettres entre diverses langues qui utilisent le même alphabet. Dans les tableaux 3.6 et 3.7, on trouve sur une même ligne tous les bigrammes qui commencent par la lettre située au début de cette ligne.

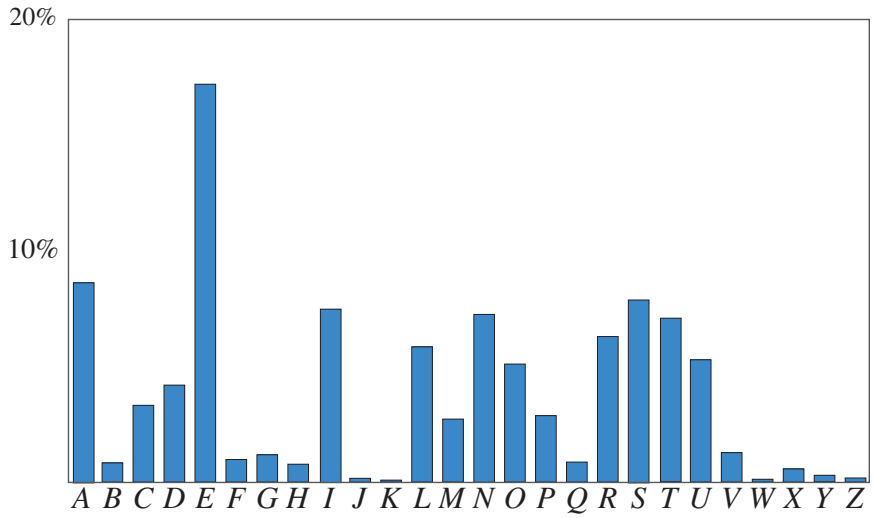


TABLE 3.3 – Le «profil» des fréquences des lettres du français

7. Les fréquences de ce tableau ont été prélevées dans un texte français de 100 000 lettres composé d'un texte de Gustave Flaubert (20 600 lettres), de Jules Vernes (19 438 lettres) et de trois articles de l'Encyclopedia Universalis. réf. : http ://www.apprendre-en-ligne.net/crypto/menu/

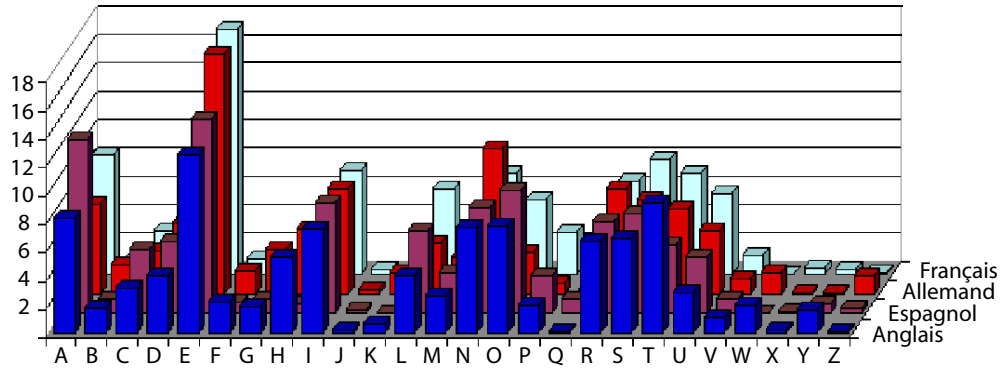


TABLE 3.4 – Comparaison des fréquences de lettres

	ES	*DE*	*LE*	*EN*	*RE*	*NT*	*ON*	*ER*	*TE*	*EL*
/100,000	3318	2409	2366	2121	1885	1694	1646	1514	1494	1382
	AN	*SE*	*ET*	*LA*	*AI*	*IT*	*ME*	*OU*	*EM*	*IE*
/100,000	1378	1377	1307	1270	1255	1243	1099	1086	1056	1030
	ED	*NE*	*TI*	*UR*	*QU*	*EC*	*AR*	*IS*	*RA*	*TA*
/100,000	998	985	984	980	975	917	905	897	896	881

TABLE 3.5 – Les 30 bigrammes les plus fréquents en français.

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	31	242	392	208	48	135	232	37	1255	32	7	663	350
B	158	2	1	2	130	1	2	0	132	4	10	181	1
C	312	0	73	19	765	2	2	411	209	3	5	124	5
D	427	1	8	24	2409	2	5	25	378	3	0	14	21
E	616	176	917	998	782	258	209	67	179	96	8	1382	1056
F	181	1	1	8	180	118	1	1	190	0	0	43	1
G	135	1	10	9	408	4	3	3	69	6	4	74	10
H	267	5	4	1	285	0	0	0	149	3	0	3	4
I	176	85	203	172	1030	114	115	6	49	14	0	798	181
J	76	0	0	0	100	0	0	0	2	0	0	0	0
K	8	0	0	0	6	0	3	0	6	0	0	0	10
L	1270	14	22	58	2366	25	14	39	512	4	1	647	18
M	510	152	11	11	1099	0	1	1	302	0	0	7	243
N	405	30	438	785	985	124	222	24	316	17	7	89	68
O	6	83	88	101	46	32	115	7	452	14	3	184	391
P	671	1	3	21	441	5	1	136	119	0	0	377	2
Q	2	0	3	0	1	0	0	1	0	0	0	1	3
R	896	53	168	302	1885	46	96	5	583	11	3	292	181
S	809	85	306	735	1377	151	73	83	565	36	0	453	192
T	881	25	166	515	1484	52	19	64	984	28	3	331	70
U	168	87	165	162	781	40	83	4	534	41	3	302	128
V	277	0	1	0	502	0	0	0	288	0	0	1	0
W	11	1	1	0	3	0	0	2	8	0	0	0	0
X	35	14	37	36	68	8	7	5	57	0	0	21	15
Y	63	0	7	7	59	3	4	0	0	0	0	13	8
Z	8	0	2	6	49	3	1	0	1	1	0	11	4

TABLE 3.6 – Fréquence des bigrammes se terminant par a–M

	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	1378	17	412	44	905	409	613	599	301	2	6	69	12
B	1	146	1	3	187	29	16	44	3	0	0	4	0
C	1	677	11	7	100	14	142	132	2	0	0	11	0
D	5	231	4	6	134	64	3	406	4	1	0	5	0
E	2121	136	699	190	1514	3318	1307	761	258	11	125	15	60
F	1	213	1	2	106	12	1	61	0	0	0	1	0
G	103	47	5	1	197	12	23	81	1	0	0	2	0
H	17	107	0	3	18	5	0	42	0	1	0	7	0
I	797	524	75	215	400	897	1243	11	190	1	40	0	4
J	0	91	0	0	0	0	0	42	0	0	0	2	0
K	3	9	0	0	5	1	0	0	0	0	0	3	0
L	41	281	69	47	16	126	42	369	14	0	0	15	1
M	4	334	201	2	10	10	8	52	1	0	0	3	0
N	249	303	130	82	55	846	1694	114	109	0	1	19	20
O	1646	8	175	19	491	126	109	1086	28	9	4	62	4
P	4	505	125	1	363	31	65	140	1	0	0	1	0
Q	0	0	1	0	1	0	0	975	0	0	0	0	0
R	88	520	82	51	176	386	445	183	77	1	1	21	5
S	107	521	496	191	137	702	578	343	92	1	6	30	10
T	40	363	268	96	668	404	269	270	41	4	6	18	3
U	516	19	184	15	980	591	469	14	177	1	264	8	4
V	0	167	0	0	81	0	0	11	0	0	0	0	0
W	0	3	0	1	0	4	0	0	0	0	0	2	0
X	3	7	56	11	3	15	35	2	18	0	4	0	0
Y	5	15	14	0	10	75	9	2	4	0	0	0	0
Z	2	15	4	1	0	3	1	0	7	4	0	0	2

TABLE 3.7 – Fréquence des bigrammes se terminant par N–Z

	EE	*SS*	*LL*	*TT*	*NN*	*MM*	*RR*	*PP*	*FF*	*CC*	*GG*	*II*	*AA*	*DD*	*UU*
/100 000	782	702	647	269	249	243	176	125	118	73	63	49	31	24	14

TABLE 3.8 – Fréquence des lettres doublées en français.

	ENT	*LES*	*EDE*	*DES*	*QUE*	*AIT*	*LLE*	*SDE*	*ION*	*EME*
/100 000	900	801	630	609	607	542	509	508	477	472
	ELA	*RES*	*MEN*	*ESE*	*DEL*	*ANT*	*TIO*	*PAR*	*ESD*	*TDE*
/100 000	437	432	425	416	404	397	383	360	351	350

TABLE 3.9 – Les 20 trigrammes les plus fréquents en français.

Chapitre 4

Probabilités

4.1 La roulette des probabilités

Notre collègue et ami, Adriano Garsia de UCSD [1], a eu l'idée lumineuse de rendre faciles et accessibles les diverses notions de la théorie des probabilités par l'introduction systématique de la *roulette*. L'efficacité de cette approche vient en partie du fait que c'est un objet avec lequel nous sommes déjà familiers, que ce soit dans le contexte des casinos ou encore des jeux télévisés. D'autre part, en généralisant un tout petit peu la roulette du casino, on obtient un puissant instrument pour l'analyse des situations probabilistes. Comme illustré à la Figure 4.1, on s'imagine donc une roulette comme un disque solide, dont on convient que la circonférence est de longueur un. Cette roue peut pivoter horizontalement autour d'un axe. La friction de la roulette sur son axe est si petite que la moindre impulsion lui fait faire un grand nombre de tours avant de s'arrêter en un endroit qui, pratiquement, est aléatoire. On place une flèche de référence près de la circonférence de la roue sur le plateau qui la supporte. Dans les applications, la roue est «marquée» sur certains segments de sa circonférence. On s'intéresse, lorsque la roue s'arrête, au cas où la flèche de référence se trouve face à un segment marqué. Il semble normal d'admettre que la *probabilité* que cela soit le cas est p, si p est la longueur de ce segment. Ainsi, à la Figure 4.1, on a marqué deux régions, l'une avec «oui», l'autre avec «non». Dans ce qui suit, nous allons voir plusieurs variantes de cette idée de marquage.

1. University of California San Diego

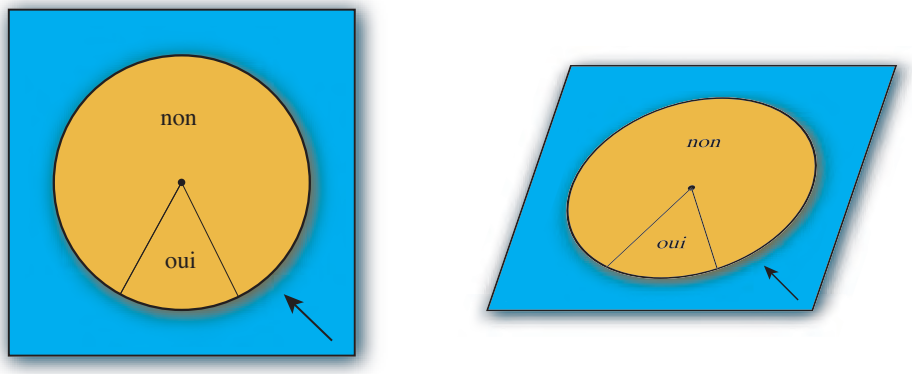


FIGURE 4.1 – Une roulette typique

4.2 Exemples autour du lancer de deux dés

Voici quelques exemples pour nous familiariser avec les concepts et le vocabulaire des probabilités. On se situe dans le contexte où on lance deux dés qui ne sont pas pipés. Les 36 résultats possibles sont illustrés à la Figure 4.2. Transposons d'abord ce lancer de deux dés

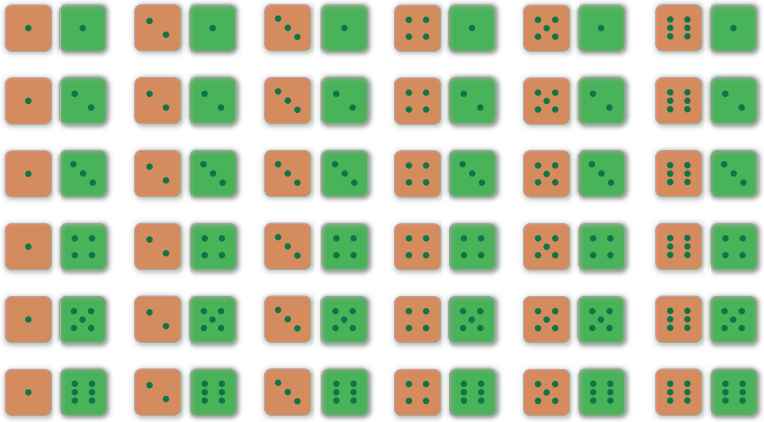


FIGURE 4.2 – Les résultats possibles du lancer de deux dés

au contexte des roulettes. On se construit la roulette de la Figure 4.3, où le résultat du premier dé est sur le disque central, et le résultat du second dé est distribué sur le pourtour

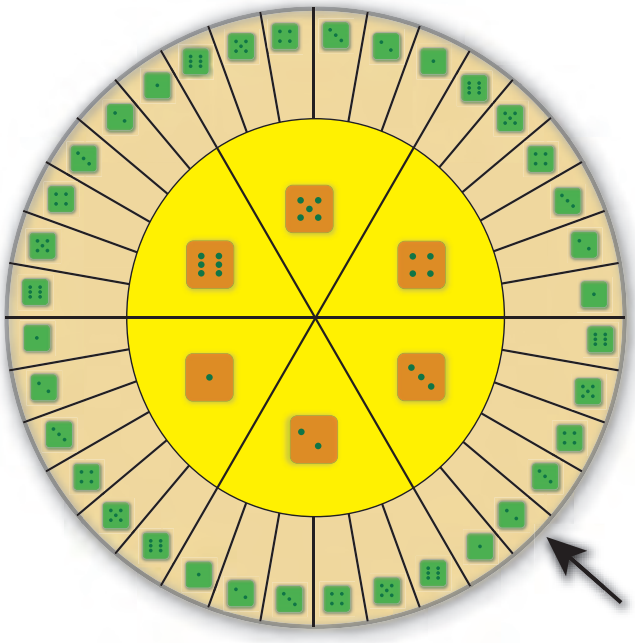


FIGURE 4.3 – Une roulette pour le lancer de deux dés

de la roulette, selon les divers résultats du premier dé. Le fait que les dés ne soient pas pipés correspond au fait que les subdivisions sont toutes égales. Lorsque la roulette cesse de tourner, on se retrouve face à un résultat qui correspond au lancer de deux dés. À la Figure 4.3, c'est le couple $(2, 3)$:

On s'intéresse ici à la somme des valeurs des dés. La fonction qui associe à chaque paire de dés, la somme de leurs valeurs, est un exemple de ce qu'on appelle techniquement une *variable aléatoire*, ici dénotée X. Les valeurs possibles de cette variable aléatoire sont, dans notre cas, les nombres entiers entre $2 = 1 + 1$ et $12 = 6 + 6$. Sous forme compacte, on écrit $X = 8$, pour signifier qu'on s'intéresse au fait que la somme des deux dés est 8. Ces valeurs n'ont pas toutes les mêmes probabilités de se produire, puisque le nombre de façons d'obtenir ces sommes varie. Dans la foulée de nos notations précédentes, on dénote $P(X = 8)$ la probabilité que la somme X soit égale à 8. C'est un nombre entre 0 et 1, qui correspond ici a la somme des longueurs de tous les arcs pour lequel la somme des deux dés donne 8.

Pour mieux comprendre la situation, on modifie notre roulette en y ajoutant, sur le pourtour, la somme correspondant au résultat correspondant. On obtient ainsi la roulette de la Figure 4.4. En comptant le nombre d'arcs (tous de longueur 1/36) correspondants à une

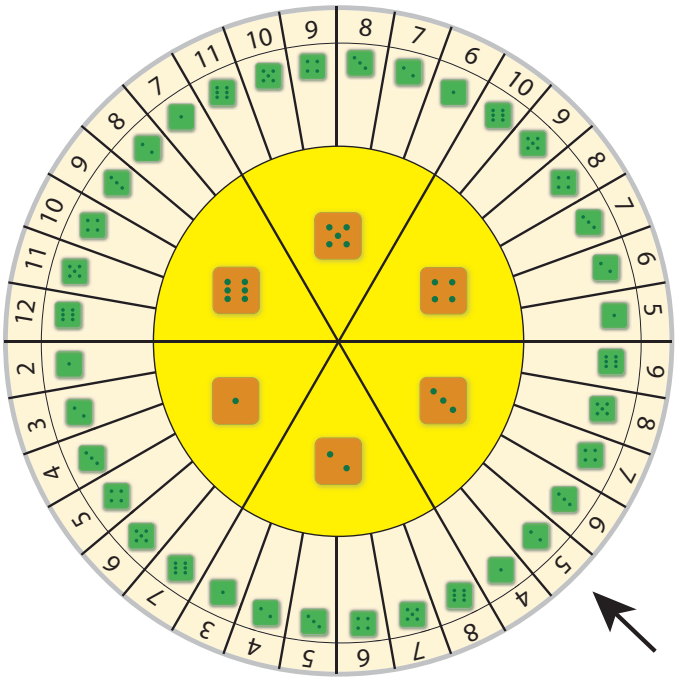


FIGURE 4.4 – Une roulette correspondant à la somme de deux dés

somme donnée, on obtient les probabilités suivantes pour les diverses valeurs possibles de la somme :

k	2	3	4	5	6	7	8	9	10	11	12
P(X=k)	1/36	2/36	3/36	4/36	5/36	6/36	5/35	4/36	3/36	2/36	1/36

Si on ne s'intéresse qu'à une roulette qui ne donne que la somme, sans tenir compte du résultat des dés comme tels, on peut redessiner le tout en regroupant les portions d'arcs qui donnent la même somme. On obtient alors la roulette de la Figure 4.5. Si on répète souvent *l'expérience aléatoire* de faire tourner cette roulette 4.5, on obtient une succession de réponses

$$10, 8, 7, 8, 5, 9, 10, 7, 8, 9, 6, 7, 3, 8, 5, 8, 6, 8, 5, 7, 6,$$
$$9, 8, 8, 11, 8, 12, 8, 8, 3, 5, 9, 10, 9, 12, 1, 6, 10, 8, 3$$

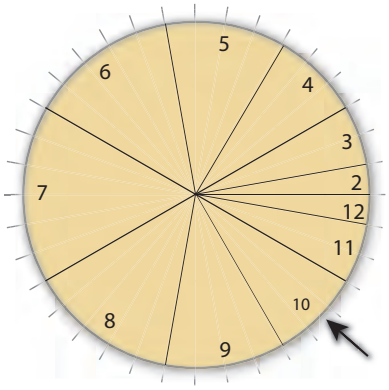


FIGURE 4.5 – La roulette de la somme

La question qu'on se pose est de savoir qu'elle est la moyenne de ces réponses, dans notre cas c'est $7,45$. Autrement dit, on cherche à savoir qu'elle est la somme de deux dés en moyenne. La valeur théorique de cette moyenne correspond à ce qu'on appelle *L'espérance mathématique* de la variable X. Elle est obtenue par le calcul suivant

$$\begin{aligned} E(X) &= 2 \times \frac{1}{6} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36} \\ &\quad + 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} \\ &= 7 \end{aligned}$$

C'est la moyenne de la somme pour un grand nombre de lancers des deux dés.

4.3 Le jargon des probabilités

Nous pouvons passer maintenant à une description plus détaillée du concept de probabilité et des notions associés. Nous appliquerons ensuite ces notions à une analyse du jeu de «craps», pour aider nos amis férus de la fréquentation des casinos du Nevada. Puis, nous donnerons une courte explication de l'indice de coïncidence. Cependant, notre véritable motivation est de se préparer pour notre discussion de la théorie de l'information.

Expérience aléatoire

Une *expérience aléatoire* est une activité, pas nécessairement scientifique (mais précise) qui produit des résultats qui dépendent du hasard, et dont on sait précisément décrire l'ensemble des résultats possibles (qu'on suppose fini dans cette discussion). Cet ensemble de résultats possibles s'appelle l'*espace échantillonnal*, et on le note ici Ω. Ainsi, tirer au hasard une lettre dans la phrase suivant constitue une expérience aléatoire.

toutes tes attitudes tentent de tromper ta tante et ses tiers (1)

L'espace échantillonnal de cette expérience est l'ensemble

$$\{a,e,i,m,n,o,p,r,s,t,u\}$$

des lettres qui apparaissent dans la phrase.

Nous pouvons toujours transposer une expérience aléatoire au contexte des roulettes. Les résultats possibles correspondent aux régions marquées sur le tour de la roulette. Par exemple les 36 secteurs de la roulette de la Figure 4.3, qui correspondent aux 36 possibilités de résultats du lancer de deux dés.

Événement simple

Dans le contexte où on a l'intention de procéder à une expérience aléatoire, on s'intéresse à certaines propriétés des résultats possibles. Pour discuter plus clairement de ces questions, on introduit la notion *d'événement*. Un événement est tout simplement un autre nom pour un sous-ensemble quelconque de l'espace échantillonnal Ω. Pour notre exemple ci-haut, on peut considérer l'événement qui correspond à piger une voyelle :

$$V = \{a, e, i, o, u\}.$$

Si on pige une lettre au hasard dans la phrase (1), on peut obtenir ou non une voyelle. Si c'est le cas, on dit aussi que l'événement V s'est produit. Les événements sont désignés ici par des lettres majuscules $A, B, \ldots$ Un exemple particulier est *l'événement simple*, $A = \{x\}$, pour lequel le sous-ensemble n'est formé que d'un seul des résultats possibles de l'expérience. En étirant [2] un peu la définition d'événement simple, on dit parfois qu'un événement simple «est» le résultat correspondant, confondant ainsi $\{x\}$ avec x. Ainsi, pour notre expérience

2. Ce qui n'est mathématiquement pas tout à fait correct, mais habituel.

précédente, choisir un e est un événement simple. Bien sûr, choisir une voyelle n'est pas un événement simple.

Dans le contexte des roulettes, un événement correspond à un ensemble de régions marquées. Si ces régions marquées sont contiguës, on peut décrire l'événement en marquant une pointe de tarte de la roue vers l'intérieur, comme l'événement «obtenir un 2 sur le premier dé». Cet événement est représenté par la pointe de tarte qui correspond aux six résultats

$$(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),$$

c'est-à-dire le sixième de tarte situé au bas de la roulette.

Probabilité

Nous avons introduit la notion d'événement pour clarifier la discussion, voilà maintenant comment exploiter ceci. Pour un événement donné, par exemple piger une voyelle, on se demande qu'elles soient nos chances de succès. C'est ce qu'on appelle la *probabilité* que l'événement se produise. Ainsi, la probabilité de piger une voyelle dans notre phrase de référence est de 18/51, parce qu'il y a 18 voyelles sur les 51 lettres de la phrase. Plus généralement, la probabilité $P(A)$ qu'un événement se produise Dans une expérience aléatoire, où tous les résultats ont la même chance de se produire, on calcule $P(A)$ comme

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}.$$

Le résultat est donc toujours un nombre entre 0 et 1. Plus généralement, si les résultats possibles d'une expérience :

$$\{a_1, a_2, \ldots, a_k\}$$

ont des probabilités respectivement égales à $P(a_i)$. Ces probabilités doivent des nombres entre 0 et 1, connues d'une certaine façon, et on a toujours

$$P(a_1) + P(a_2) + \ldots + P(a_k) = 1.$$

La *probabilité d'un événement* A est alors définie comme

$$P(A) := \sum_{a \in A} P(a).$$

Autrement dit, c'est la somme des probabilités des résultats qui sont dans l'ensemble A. En particulier, on a $P(\Omega) = 1$, ce qui correspond à dire bêtement qu'un des résultats possibles se produira.

Sur une roulette, la probabilité d'un événement est la somme des longueurs d'arcs des régions marquées qui correspondent à cet événement.

Variable aléatoire

L'étape suivante, dans notre rapide survol de la théorie des probabilités, correspond à calculer certaines caractéristiques des résultats possibles d'une expérience. On peut ainsi penser à mesurer la taille d'une personne choisie au hasard dans la classe. Une *variable aléatoire*, X, est tout simplement une fonction qui associe à chaque événement simple, $a \in \Omega$, un nombre $X(a)$. Les variables aléatoires servent à décrire, de façon compacte, mais claire, certains événements particuliers. Si k est l'une des valeurs possibles de la variable aléatoire X, on peut ainsi écrire,

$X = k$, pour désigner l'événement qui se produit exactement lorsque le résultat, a, de l'expérience est tel que $X(a) = k$,

$X \geq k$, pour désigner l'événement qui se produit exactement lorsque le résultat, a, de l'expérience est tel que $X(a) \geq k$,

etc

On peut donc écrire un énoncé du type :

«La probabilité qu'une personne choisie au hasard soit plus grande que 6 pieds»

simplement comme

$$P(T \geq 6),$$

si on a convenu que T désigne la variable aléatoire qui consiste à mesurer la taille.

La variable aléatoire X qui donne la somme de deux dés a déjà été considérée dans notre discussion.

Espérance d'une variable aléatoire

Pour une variable aléatoire, on s'intéresse souvent à la *valeur moyenne* prise par cette valeur aléatoire. C'est par exemple la taille moyenne des individus d'une population. C'est exactement ce que mesure *l'espérance mathématique*, $E(X)$, d'une variable aléatoire X. Si l'ensemble des valeurs que peut prendre X est

$$\{x_1, x_2, \ldots, x_k\},$$

alors calcul l'espérance de X comme suit

$$E(X) = x_1 P(X = x_1) + x_2 P(X = x_2) + \cdots + x_k P(X = x_k).$$

L'interprétation de cette espérance est la suivante. Si l'on répète l'expérience aléatoire un grand nombre de fois, et qu'on calcule la moyenne des valeurs x_i obtenues, alors nous devrions «espérer» obtenir un résultat moyen près de $E(X)$. Considérons ainsi la variable aléatoire X qui donne la fréquence des lettres dans notre phrase (1), et dont les valeurs apparaissent au tableau 4.1. L'espérance de cette variable aléatoire est donc :

	a	*d*	*e*	*i*	*m*	*n*	*o*	*p*	*r*	*s*	*t*	*u*
X=	3	2	11	2	1	3	2	1	3	6	15	2

TABLE 4.1 – Fréquence d'apparition des lettres dans la phrase (1).

$$\begin{aligned} E(X) &= X(a) \cdot P(a) + X(e) \cdot P(e) + X(o) \cdot P(o) \cdots + X(t) \cdot P(t) \\ &= 8,37 \end{aligned}$$

ce qui signifie que $8,37$ est la fréquence moyenne d'une lettre pigée au hasard dans cette phrase.

Probabilité conditionnelle

Lors de la réalisation d'une expérience aléatoire, il est possible qu'on accumule des informations (privilégiées !) sur les résultats éventuels. Autrement dit, on découvre que le résultat fera certainement partie d'un certain sous-ensemble précis B de l'ensemble des résultats possibles. Dans notre langage récemment introduit, cela correspond à dire qu'on sait que l'événement B se produit avec certitude. On cherche alors à savoir qu'elle est la probabilité qu'un certain autre événement A se produise. C'est cette notion que permettent de décrire les *probabilités conditionnelles*.

Pour illustrer, posons-nous la question de savoir si la somme X des deux dés est strictement plus grande que 6, sachant que le résultat du 1^{er} dé est inférieur ou égal à 3. En terme de variable aléatoire, on écrit $X > 6$, pour signifier que la somme est supérieure à 6. Le résultat du premier dé (le brun) est une autre variable aléatoire désignée B. La condition énoncée ci-haut est donc que $B \leq 3$. En terme des roulettes, on peut comprendre le phénomène de la façon suivante. Supposons qu'on sait que la roulette à été trafiquée de façon à ne jamais s'arrêter sur une certaine portion précise de sa circonférence. On a illustré cette idée à la figure 4.6, en voilant la portion du bas de la roulette. On considère donc qu'il est impossible

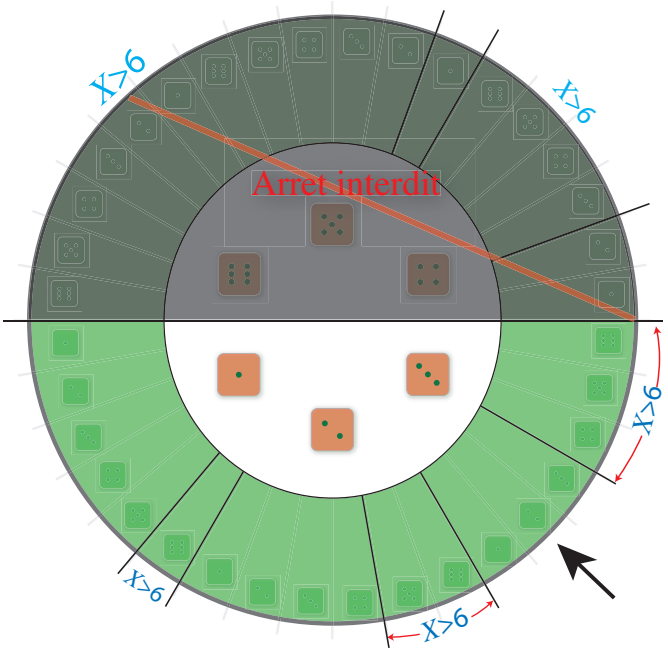


FIGURE 4.6 – La roulette d'un résultat conditionnel

de s'arrêter dans cette portion. Figure 4.6. Le calcul de la probabilité conditionnelle se fait alors de la façon suivante. On calcule la probabilité d'obtenir une somme supérieure à 6, tout ayant au plus 3 sur le premier dé, ce qui est dénoté $P(X > 6 \text{ et } B \leq 3)$. La probabilité conditionnelle, dénotée $P(X > 6 \mid B \leq 3)$, est définie comme

$$\begin{aligned} P(X > 6 \mid B \leq 3) &= \frac{P(X > 6 \text{ et } B \leq 3)}{P(B \leq 3)} \\ &= \frac{6/36}{1/2} \\ &= \frac{1}{3}. \end{aligned}$$

En d'autres mots, on a une chance sur trois d'obtenir une somme des dés supérieure à 6, si on sait que la valeur du premier dé est 1, 2 ou 3.

En général, pour des événements quelconques A et B, on dénote $P(A|B)$, la probabilité conditionnelle de A, étant donné B. Elle se calcule simplement comme le rapport

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \tag{2}$$

où $A \cap B$ est l'intersection des événements A et B. Si l'événement correspondant à cette intersection se produit, c'est que à la fois A et B se produisent. Ainsi, pour la phrase (1), on peut calculer la probabilité de choisir l'une des lettres qui apparaissent dans le mot «*attitudes*», sachant qu'on a choisi une lettre qui apparaît dans le mot «*toutes*». Pour ce faire, on considère les événements

$$A : \text{ choisir une lettre qui appartient au mot «}attitudes\text{»},$$

et

$$B : \text{ choisir une lettre qui appartient au mot «}toutes\text{»}.$$

Autrement dit, $A = \{a, t, i, u, d, e, s\}$, $B = \{t, o, u, e, s\}$, et en conséquence

$$A \cap B = \{t, u, e, s\}.$$

La probabilité de choisir une lettre du mot «*attitudes*» sachant que cette lettre appartient au mot «*toutes*» est donc

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{34}{36}.$$

Indépendance

Lorsque la réalisation d'un événement B influence la probabilité de réalisation d'un autre événement A, on dit que A est *dépendant* de B. Sinon, on dit que A et B sont *indépendant*. Plus précisément, l'événement A est indépendant de B exactement lorsque

$$P(A) = P(A|B). \tag{3}$$

Ce qui dit exactement que la probabilité que A se produise est la même que la probabilité que A se réalise, sachant que B est réalisé. Ainsi, il y a dépendance entre les événements consistants à choisir une lettre dans «*attitudes*» et choisir une lettre dans «*toutes*», puisqu'on calcule que

$$P(A) = \frac{35}{51} \qquad \text{tandis que} \qquad P(A|B) = \frac{34}{36}.$$

Il est important de souligner que s'il y a dépendance entre A et B au sens ci-dessus, cela ne signifie pas du tout qu'il y ait un quelconque lien de cause à effet entre A et B. Ce n'est qu'une notion probabiliste. On peut reformuler tout simplement la notion d'indépendance, en utilisant la définition de probabilité conditionnelle (2), on obtient alors que A et B sont indépendants si et seulement si

$$P(A \cap B) = P(A) \cdot P(B) \tag{4}$$

On adapte le concept d'indépendance aux contextes des variables aléatoires de la façon suivante. On dit qu'une variable X est *indépendante* d'une autre variable Y si et seulement si la connaissance de la valeur qu'a X ne change pas la probabilité que Y prenne une certaine valeur. Plus précisément pour toute valeur a, que peut prendre X, et toute valeur b, que peut prendre Y, on a

$$P(X = a \text{ et } Y = b) \;=\; P(X = a) \cdot P(Y = b)$$

Nous allons encore une fois illustrer toutes ces notions dans le contexte du jeu de craps. C'est un jeu un peu complexe, dont la bonne ou mauvaise compréhension peut influencer votre avenir financier.

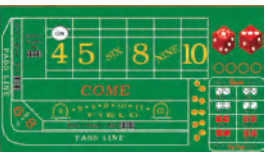


Planche de craps

4.4 Le jeu de craps

Le jeu de craps est un jeu de dés qui se pratique dans les casinos. Il en existe différentes variantes, et nous n'allons en étudier qu'une version un peu simplifiée. Voici la description des règles du jeu. Il y a un joueur appelé le *lanceur* qui lance les (deux) dés, et parie contre la maison (le casino). On permet aux *observateurs* de parier sur le résultat du lancer des dés. Pour jouer une ronde de craps, le lanceur lance les dés une première fois, et on calcule la somme des deux dés.

1. Si le résultat est 7 ou 11, le lanceur gagne.
2. Si le résultat est 2, 3 ou 12, le lanceur perd.
3. Si le résultat est 4, 5, 6, 8, 9 ou 10, le joueur relance les dés jusqu'à ce que
   (a) un 7 apparaisse et le lanceur perd,
   (b) le résultat du 1$^{\text{er}}$ lancer réapparaît et le lanceur gagne.

Lors d'un gain, le lanceur gagne simplement l'équivalent de sa mise (en la récupérant bien sûr).

Essayons maintenant de décrire tout cela en terme de roulettes. Remarquons qu'au cours d'une ronde, le joueur produit plusieurs nombres de façon aléatoire. Les plus significatifs correspondent aux variables aléatoires suivantes :

$$\begin{aligned} U &= \text{le résultat du premier lancer} \\ V &= \text{le résultat du lancer qui détermine le gagnant} \\ X &= \begin{cases} 1 & \text{si le lanceur gagne,} \\ 0 & \text{si le lanceur perd} \end{cases} \end{aligned}$$

Lorsque $U = 2,\ 3,\ 7,\ 11$ ou 12, la ronde est terminée aussitôt après le premier lancer. Si $U = V$, alors

$$X = \begin{cases} 1 & \text{si } U = 7 \text{ ou } 11 \\ 0 & \text{si } U = 2,\ 3,\ \text{ou } 12 \end{cases}$$

Si $U = 4,\ 5,\ 6,\ 8,\ 9$ ou 10, alors le jeu continue avec les différentes possibilités de l'organigramme de la figure 4.7

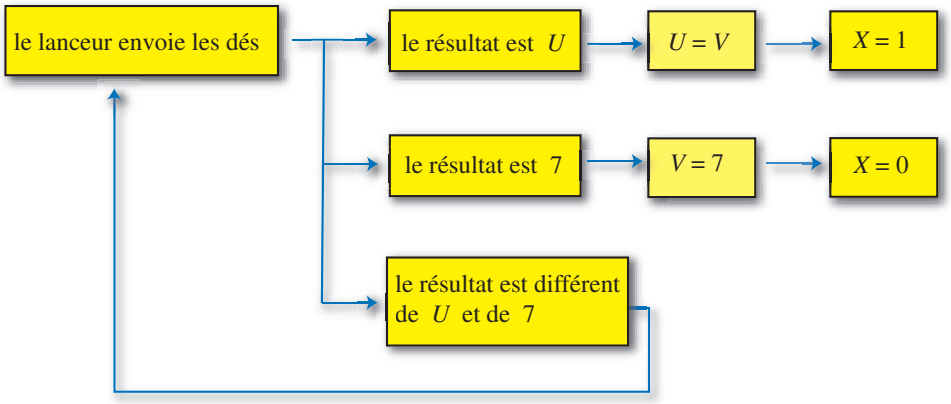


FIGURE 4.7 – Organigramme du jeu de Craps après le premier lancer, $U = \{4, 5, 6, 8, 9, 10\}$

L'ensemble des 17 possibilités pour les valeurs de (U, V, X) sont

$$\begin{array}{ccccccc} (2,2,0) & (3,3,0) & (4,4,1) & (5,5,1) & (6,6,1) & (7,7,1) & (8,8,1) \\ (4,7,0) & (5,7,0) & (6,7,0) & (8,7,0) & (9,9,1) & (10,10,1) & (11,11,1) \\ (2,12,0) & (9,7,0) & (10,7,0) & & & & \end{array}$$

Notre objectif est de construire une roulette dont le résultat à une impulsion représente une ronde de craps. Nous allons représenter les 17 triplets (U, V, X) possibles par 17 régions sur la roulette en modifiant la roulette de la figure 4.4. On observe que, pour $U = 2,\ 3,\ 7,\ 11$ ou 12, les valeurs de V et X sont imposées par la valeur de U. Pour $U = 4,\ 5,\ 6,\ 8,\ 9$ ou 10, il y a deux situations possibles pour les couples (V, X). Ainsi, pour $U = 4$, la valeur de V est déterminée en tournant la roulette 4.5 jusqu'à ce qu'on obtienne une valeur de 4 ou de 7. Autrement dit, on force le résultat à être un 4 ou 7 , ce qui correspond à une probabilité conditionnelle. Les probabilités respectives de 4 et 7 sont alors de $1/3$ et $2/3$. De façon analogue, pour $U = 5$, la valeur de V doit être 5 ou 7 avec des probabilités respectives

2/5 et 3/5. Et ainsi de suite pour les valeurs $U = 6,\ 8,\ 9$. Notre roulette (voir figure 4.8) est constituée de trois cercles concentriques ; le plus petit cercle pour les valeurs de U, le moyen pour les valeurs de V, et le plus grand cercle pour les valeurs de X .

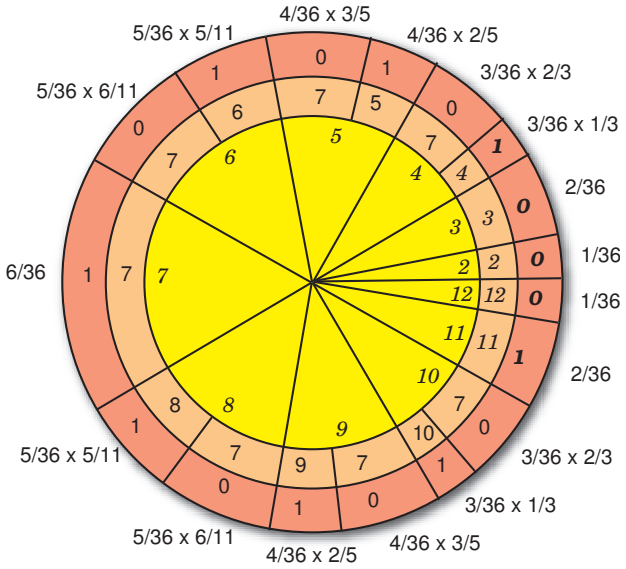


FIGURE 4.8 – Une roulette représentant le jeu de craps

On peut lire sur cette roulette toute l'analyse des probabilités pour le jeu de craps. Pour calculer la probabilité que le lanceur gagne, on additionne les longueurs d'arcs étiquetés pour obtenir

$$\begin{aligned} P(X=1) &= \frac{2}{36} + 2\cdot\left(\frac{1}{3}\cdot\frac{3}{36}\right) + 2\left(\frac{2}{5}\cdot\frac{4}{36}\right) + 2\left(\frac{5}{11}\cdot\frac{5}{36}\right) + \frac{6}{36} \\ &= \frac{244}{495} = 0,4929 \end{aligned}$$

ce qui est anormalement élevé dans un casino. L'espérance de gain du lanceur est donc

$$\mathrm{E}(\text{gain}) = 1 \times 0,4929 + (-1) \times 0,5071 = -0,0142$$

Ce qui signifie qu'à chaque ronde de jeu, le joueur perd en moyenne 1,4 sou par dollar parié.

4.5 Probabilité totale

Comme on vient de le voir, les résultats d'une expérience sont parfois obtenus à la suite de plusieurs étapes aléatoires. Pour trouver la probabilité de réalisation d'un de ces résultats , il faut suivre tous les chemins menant à ce résultat dans l'arbre des possibilités et calculer la somme des probabilités associées à chaque chemin. Illustrons ce processus par un exemple. On a trois urnes appelées A, B et C contenant des pièces de 5 cents et de 10 cents de la façon décrite à la figure 4.9

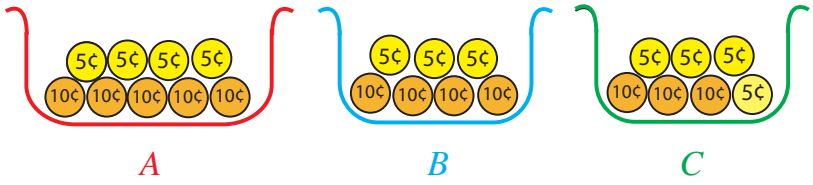


FIGURE 4.9 – Trois urnes contenant des pièces de 5 cents et 10 cents.

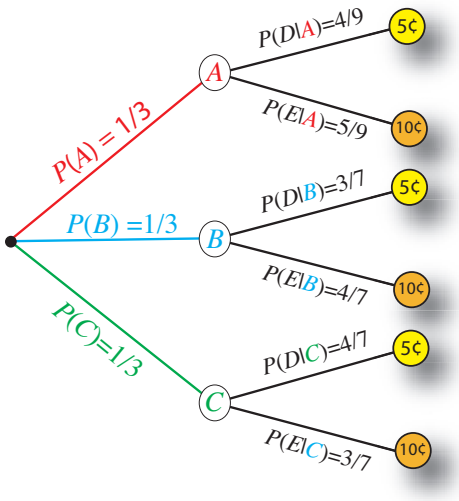


FIGURE 4.10 – L'arbre des possibilités d'un tirage

On choisit une urne au hasard pour y piger une pièce. Le résultat dépend donc de deux expériences aléatoires : l'expérience qui choisit l'urne et l'expérience qui choisit la pièce. Pour calculer, par exemple, quelle est la probabilité de choisir une pièce de 5 cents, on

considère l'arbre des possibilités de choix d'une urne et d'une pièce tel que représenté à la figure 4.10. Il y a trois chemins possibles pour obtenir une pièce de 5 cents. On donne à chaque arête une probabilité, et la probabilité d'un chemin est le produit des probabilités de chaque arête qui le compose. La probabilité d'obtenir un 5 cents est alors la somme des probabilités de tous les chemins qui se terminent en un sommet correspondant à 5 cents. Ainsi, on a

$$\begin{aligned} P(D) &= P(A)\cdot P(D|A) + P(B)\times P(D|B) + P(C)\times P(D|C) \\ &= \frac{1}{3}\cdot\frac{4}{9} + \frac{1}{3}\cdot\frac{3}{7} + \frac{1}{3}\cdot\frac{4}{7} = \frac{13}{27} \end{aligned}$$

où D désigne l'événement «choisir un 5 cents», et E l'événement «choisir un 10 cents». Ce genre de situation est fréquente, et on dit avoir fait un calcul de *probabilité totale*.

4.6 Explication de l'indice de coïncidence

Pour un texte quelconque de longueur n, l'indice de coïncidence est la probabilité de tirer deux fois la même lettre dans le texte considéré. Pour calculer cette probabilité, on procède comme suit. Posons

$$\begin{gathered} n_A = \text{ nombre de } A \text{ dans le texte} \\ n_B = \text{ nombre de } B \text{ dans le texte} \\ n_C = \text{ nombre de } C \text{ dans le texte} \\ \cdots \\ n_Z = \text{ nombre de } Z \text{ dans le texte} \end{gathered}$$

La probabilité $P(AA)$ que les deux lettres tirées au hasard soient des A, est obtenue en divisant le nombre total de façons de tirer deux des A du texte, par le nombre total de façons de tirer deux lettres quelconques. On a donc

$$P(AA) = \frac{\binom{n_A}{2}}{\binom{n}{2}} = \frac{n_A(n_A-1)}{n(n-1)} \tag{5}$$

La probabilité que les deux lettres tirées soient deux fois la même lettre peut donc s'obtenir en additionnant les probabilités de tirer deux fois la lettre A, deux fois la lettre B, ..., deux fois la lettre Z. Si nous notons IC l'indice de coïncidence, on obtient donc la formule :

$$\begin{aligned} IC &= P(AA) + P(BB) + \ldots + P(ZZ) \\ &= \frac{n_A(n_A-1) + n_B(n_B-1) + \ldots + n_Z(n_Z-1)}{n(n-1)} \end{aligned}$$

Si à chaque pige toutes les lettres ont la même probabilité 1/26 d'être tirées, alors on constate que $P(AA) = (1/26)^2$, et de même pour toutes les autres lettres. Il découle de la formule que l'indice de coïncidence des textes au hasard est 1/26.

4.7 Exercices

4.1. On choisit au hasard un nombre dans l'ensemble $\{1, 2, 3 \ldots, 15\}$.

a) Quelle est la probabilité que le nombre choisi soit divisible par trois si on sait qu'il est impair ?

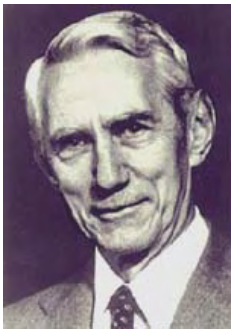
b) Ces deux événements sont-ils indépendants ?

4.2. Dans une population de lapins, il y a 2/5 de mâles et 3/5 de femelles. De plus, 5% des mâles et 3% des femelles sont albinos. Quel est le pourcentage d'albinos dans cette population ?

Chapitre 5

La théorie de l'information

On est parfois amené à conclure, par l'expérience, qu'un système cryptographique semble bien plus sûr qu'un autre. Par exemple, on peut avoir trouvé comment briser le premier de deux systèmes, sans être parvenu à briser le second. Cependant, rien n'assure que quelqu'un d'autre, avec plus de chance ou de finesse d'esprit, ne saura trouver que le dernier système est en fait encore plus facile à briser ? Dans ce contexte, il est certainement préférable d'avoir une approche plus objective à ce genre de comparaison, surtout si le fait de garder secrètes nos informations codées est d'une grande importance. Pour arriver a faire une comparaison rigoureuse entre systèmes de codage, nous allons utiliser les outils de la *théorie de l'information*, mise au point par le mathématicien Claude Shannon vers 1947. Cette théorie s'articule autour de la notion d'*entropie*, via laquelle nous allons pouvoir mesurer la sécurité d'un cryptosystème.



Claude Shannon
(1916–2001)

5.1 Entropie et incertitude

La notion d'entropie viens de la physique, plus spécifiquement de la partie de la physique qu'on appelle thermodynamique. Elle consiste en une mesure de l'état de désordre d'un système d'atomes ou de molécules. L'entropie augmente lorsque le système évolue vers un état de plus grand désordre, et elle diminue si le système évolue vers un état plus ordonné. Stephen Hawking, dans son livre *Une brève histoire du temps*, explique qu'une tasse déposée sur une table est dans un état élevé d'ordre et possède une entropie faible. Par opposition, une tasse brisée sur le plancher est plus en désordre et son entropie est donc considérée comme plus élevée. Une des lois fondamentales de la thermodynamique affirme que l'entropie (d'un système fermé) augmente toujours avec le temps. À la blague, on pourrait donc dire

que le proverbe «tant va la cruche à l'eau qu'à la fin elle se brise» exhibe une compréhension intuitive de cette loi.

Pour illustrer, considérons un jeu de cartes qui, au départ, est bien ordonné, avec les cartes allant de l'as au roi de coeur, puis de l'as au roi de carreau, de l'as au roi de pique, et enfin de l'as au roi de trèfle. Son entropie est alors considérée comme très faible. Lorsqu'on brasse (au hasard) les cartes, l'entropie (ou désordre) du jeu augmente. Autrement dit, les probabilités des divers mélanges possibles se rapprochent toutes de plus en plus d'une valeur commune, qui est de

$$\frac{1}{52!} = \frac{1}{80658175170943878571660636856403766975289505440883277824000000000000}$$

La seconde loi de la thermodynamique affirme que jamais l'écart, entre l'une de ces probabilités, et la valeur $1/52!$ n'ira en augmentant. Comme illustré à la figure 5.1, la situation est semblable pour les molécules d'un gaz dans une boîte. Avec le temps, elles auront tendance à ce répartir dans toute la boîte, et non à rester confinées dans la moitié de gauche.

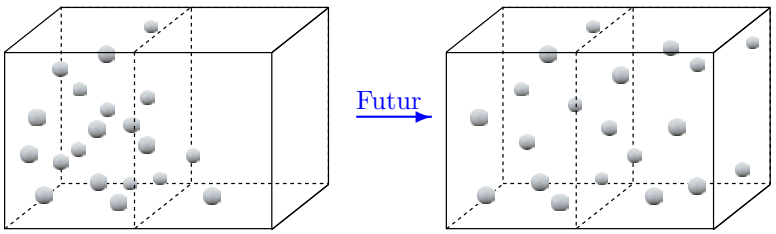


FIGURE 5.1 – L'entropie augmente avec le temps.

Ludwig Boltzmann
(1844 – 1906)

Issu de la thermodynamique, d'abord proposée par Rudolf Clausius en 1865, puis améliorée par Ludwig Boltzmann en 1872, la notion d'entropie a été transformée par Shannon pour servir de fondement à sa théorie de l'information. Cette théorie est maintenant utilisée dans de nombreux contextes. En particulier, la notion d'entropie, selon Shannon, a été adaptée à la cryptographie pour mesurer le «désordre» d'un cryptosystème. Ainsi, nous allons voir que : plus l'entropie d'un cryptosystème est élevée, plus celui-ci est difficile à briser.

Dans le contexte de la théorie de l'information, les concepts d'entropie et d'incertitude interviennent dans des situations où l'on désire comparer entre elles l'*incertitude* d'expériences aléatoires. Pour illustrer, considérons les expériences aléatoires qui consistent à

- A – Lancer (idéalement) une pièce de monnaie,
- B – Lancer un dé (non pipé) à 6 faces,
- C – Piger une carte (parfaitement au hasard) dans un paquet de 52 cartes.

De toute évidence, on a plus de chances de prédire correctement le résultat de l'expérience A, que de prédire celui des expériences B ou C. Nous dirons qu'il y a plus d'incertitude sur le résultat des expériences B ou C, que sur le résultat de l'expérience A. C'est cette incertitude que l'entropie va permettre de mesurer.

Contexte des expériences avec résultats équiprobables

Avant de donner une définition «rigoureuse» d'entropie, nous allons chercher à comprendre comment Shannon à pu être mené à une telle définition. Nous allons considérer certaines expériences aléatoires E simples, pour motiver la façon de procéder au calcul de l'entropie $H(E)$ de l'expérience en question. On dit qu'une expérience aléatoire E est *équiprobable*, si ses k résultats possibles :

$$e_1, e_2, \ldots, e_k,$$

ont tous la même probabilité de $P(e_i) = 1/k$. C'est la situation du lancer (idéal) d'une pièce de monnaie ($k = 2$), d'un dé ($k = 6$), etc. Comme on l'a remarqué plus haut, on s'attend à ce que la mesure $H(E)$, de l'entropie de E, soit plus grande plus le nombre k est élevé. Il semble aussi tout à fait raisonnable de s'attendre à ce que deux expériences aléatoires équiprobables avec k résultats possibles devraient avoir très exactement la même entropie. Autrement dit,

1) *l'entropie d'une expérience aléatoire équiprobable, avec k résultats possibles, est une fonction de k seulement. De plus, la valeur de $H(E)$ croit avec k.*

Bien entendu, lorsqu'il n'y a qu'un seul résultat possible ($k = 1$), il n'y a aucune incertitude sur le résultat de l'expérience. On doit donc poser

2) *$H(E) = 0$, dans le cas où $k = 1$.*

Pour déterminer plus précisément le comportement de la fonction d'entropie, on considère deux expériences aléatoires *indépendantes* E et F. C'est-à-dire deux expériences telles qu'une information connue sur l'issue de l'une n'affecte pas la probabilité des résultats de l'autre. On concocte alors une nouvelle expérience aléatoire, appelée *produit* de E et F, et noté $E \cdot F$. Plus précisément, c'est l'expérience plus complexe qui consiste à faire les deux expériences E et F. Peu importe si on fait ces deux expériences en même temps ou l'une après l'autre, puisqu'elles sont indépendantes. Par exemple, on lance un dé et on tire simultanément l'une des 52 cartes d'un paquet. Les résultats de $E \cdot F$ sont les couples (e, f) de résultats respectifs, e de E, et f de F. Si E est une expérience équiprobable avec k résultats possibles, et F en est une avec n résultats possibles, alors $E \cdot F$ est une expérience équiprobable avec $k\,n$ résultats possibles. L'observation cruciale est ici que

3) *Si E et F sont indépendantes, alors l'entropie de $E \cdot F$ est la somme de l'entropie*

de E et de celle de F. En formule :

$$H(E \cdot F) = H(E) + H(F). \tag{1}$$

Pour voir la justesse de ce principe, on raisonne comme suit. Supposons que, lors de l'expérience $E \cdot F$, E et F se réalisent l'une après l'autre. On commence par prédire le résultat de E avec une incertitude de $H(E)$. Puis on prédit le résultat de F, avec incertitude $H(F)$. Comme les deux expériences sont indépendantes, il semble naturel de considérer que l'incertitude de $E \cdot F$ est l'incertitude de E cumulée à celle de F. Autrement dit, c'est la somme de $H(E)$ et de $H(F)$. Bien entendu, il ne s'agit pas ici d'une preuve mathématique, mais plutôt d'un raisonnement plausible qui nous guide pour la mise en place d'une bonne définition.

L'intérêt des trois observations ci-haut est qu'elles nous poussent à choisir une fonction très particulière pour calculer l'entropie. En effet, dans le répertoire des fonctions mathématiques usuelles, une seule répond aux trois critères énoncés, à savoir la fonction $\log_a(k)$, puisqu'on a bien que

1) $\log_a(k)$ est une fonction dont la valeur croît avec k,
2) $\log_a(1) = 0$, et
3) $\log_a(k\,n) = \log_a(k) + \log_a(n)$.

Cette troisième condition correspond bien à notre troisième principe, puisque l'expérience $E \cdot F$ admet $k\,n$ résultats possible. Il ne nous reste plus qu'un petit détail à ajuster, à savoir le choix d'une base pour le logarithme. C'est-à-dire, le choix de a. Dans le contexte de la cryptographie, ou en informatique en général, il est tentant (et on ne résistera pas à la tentation) de choisir $a = 2$. Autrement dit, on considère que notre unité de base pour mesurer l'incertitude correspond à une expérience aléatoire équiprobable avec 2 résultats possibles. On a donc décidé ainsi que le fait d'apprendre le résultat de cette expérience nous donne exactement un «bit» d'information. On pose donc,

$$H(E) := \log_2(k), \tag{2}$$

pour E une expérience équiprobable avec k résultats possibles.

Résultats non équiprobables

Nous n'avons pas encore terminé de dégager la définition complète d'entropie. En effet, l'équation (2) ne nous permet que de calculer dans le cas d'expériences équiprobables. Pour étendre la définition à toutes les expériences aléatoires, on commence par mesurer la contribution à l'entropie de chaque résultat possible d'une expérience équiprobable. Ainsi,

pour E admettant k résultats équiprobables, on est amené à penser que chaque résultat e contribue également à l'incertitude de E, d'une valeur égale à

$$H(e) := \frac{1}{k} H(E). \tag{3}$$

Autrement dit, l'entropie a été très équitablement répartie entre les divers résultats possibles. Pour la suite de notre discussion, il sera très utile d'écrire l'égalité (4) sous la forme équivalente suivante

$$H(e) := p \log_2(1/p), \tag{4}$$

où p est la probabilité de l'événement e. Ici, il faut penser que $1/p = k$. L'avantage très net de la formule [1] ci haut est qu'elle a maintenant un sens pour n'importe quelle expérience aléatoire. Ainsi cette formule (4) s'exporte naturellement vers les expériences à résultats non équiprobables, pour enfin donner la définition générale suivante de l'entropie

Définition d'entropie de Shannon. Si E est une expérience aléatoire qui admet $e_1, e_2, \ldots, e_k$ comme résultats possibles, avec les probabilités respectives $p_1, p_2, \ldots, p_k$, alors l'*entropie* de E, notée $H(E)$, est donnée par la formule

$$H(E) := p_1 \log_2 1/p_1 + p_2 \log_2 1/p_2 + \ldots + p_k \log_2 1/p_k \tag{5}$$

Pour illustrer, on peut calculer l'entropie d'un texte de n lettres où, comme à la section 3.3, chaque lettre est choisie aléatoirement selon la distribution des lettres dans un texte français. On considère donc d'abord l'expérience aléatoire E qui consiste à choisir une lettre de l'alphabet, en pigeant les lettres dans une urne où chaque lettre apparaît dans les mêmes proportions que dans un texte français typique. Après calcul, on obtient :

$$\begin{aligned} H(E) &= p_A \log_2 1/p_A + p_B \log_2 1/p_B + \ldots + p_Z \log_2 1/p_Z \\ &= 3.95 \end{aligned}$$

avec

$$p_A = 0.084, \quad p_B = 0.011, \quad p_C = 0.03, \quad p_D = 0.042, \quad p_E = 0.173, \ \ldots$$

Le choix d'un texte de n lettres consiste à répéter cette expérience n fois, de façon indépendante. Nos principes décrits ci-haut entraînent alors que l'entropie de l'expérience consistant à produire un texte T, de n lettres, est

$$H(T) = 3.95\, n. \tag{6}$$

Il est tout naturel de définir l'entropie d'une variable aléatoire X, comme étant

$$H(X) := p_1 \log_2 1/p_1 + p_2 \log_2 1/p_2 + \ldots + p_k \log_2 1/p_k,$$

1. Qu'on appelle parfois *entropie locale*.

où p_i est maintenant la probabilité que le résultat de X soit a_i, $1 \leq i \leq k$. Ainsi, l'entropie de la variable aléatoire qui donne la somme de deux dés se calculera comme suit :

$$\begin{aligned} H(X) &= p_2 \log_2 1/p_2 + p_3 \log_2 1/p_3 + p_4 \log_2 1/p_4 + p_5 \log_2 1/p_5 + p_6 \log_2 1/p_6 + p_7 \log_2 1/p_7 \\ &\quad + p_8 \log_2 1/p_8 + p_9 \log_2 1/p_9 + p_{10} \log_2 1/p_{10} + p_{11} \log_2 1/p_{11} + p_{12} \log_2 1/p_{12} \end{aligned}$$

avec les probabilités données par la roulette 4.4. Comme

$$\begin{aligned} &p_2 = p_{12} = \tfrac{1}{36}, && p_3 = p_{11} = \tfrac{2}{36}, && p_4 = p_{10} = \tfrac{3}{36}, \\ &p_5 = p_9 = \tfrac{4}{36}, && p_6 = p_8 = \tfrac{5}{36}, && p_7 = \tfrac{6}{36}, \end{aligned}$$

on obtient

$$\begin{aligned} H(X) &= 2\left(\frac{36}{1}\log_2\frac{36}{1}\right) + 2\left(\frac{36}{2}\log_2\frac{36}{2}\right) + 2\left(\frac{36}{3}\log_2\frac{36}{3}\right) \\ &\quad + 2\left(\frac{36}{4}\log_2\frac{36}{4}\right) + 2\left(\frac{36}{5}\log_2\frac{36}{5}\right) + \left(\frac{6}{36}\log_2\frac{36}{6}\right) \\ &= 3.27 \end{aligned}$$

5.2 Propriétés de l'entropie

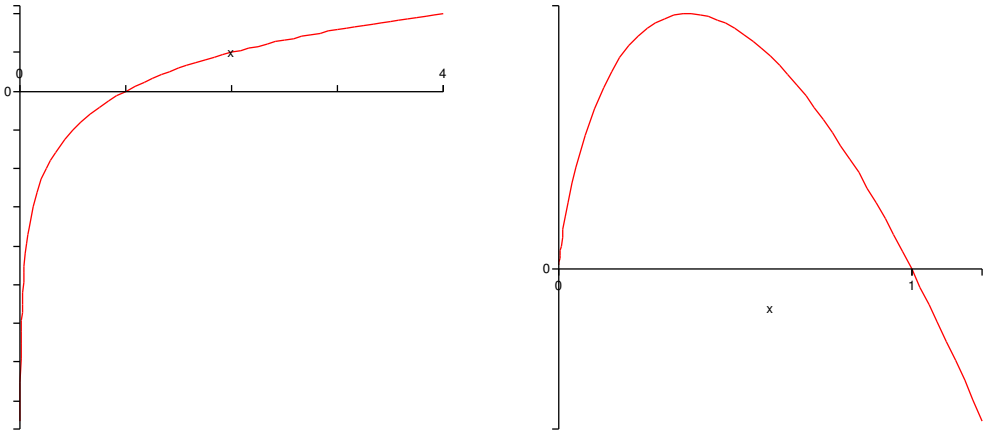
Pour clarifier certains aspects de la définition d'entropie, nous allons maintenant discuter certaines des propriétés simples des fonctions $\log_2(x)$ et $x \log_2(x)$. Pour mettre en lumière certaines de ces propriétés, il est très certainement utile de considérer le graphe de ces fonctions (voir Figure 5.2). Ceci permet de constater que l'entropie locale

$$H(e) = p \log_2(1/p),$$

d'un événement e de probabilité $p = 0$, doit être égale à 0, puisque la fonction

$$f(x) = x \log_2 1/x$$

tend vers zéro lorsque x tend vers zéro. D'autre part, si e est un événement certain pour l'expérience E, alors on a forcément $H(E) = 0$. En effet, lorsqu'une expérience aléatoire E admet un résultat certain, alors tous les autres résultats ont forcément une probabilité égale à 0. Comme l'entropie de E est la somme des entropies locales, qui sont dans ce cas toutes égal à 0, on déduit que $H(E) = 0$. En fait, une expérience E possède un événement certain si et seulement si $H(E) = 0$. C'est donc dire que l'entropie 0 correspond très exactement au cas où le résultat de l'expérience est assuré d'avance.



a) $\log_2(x)$. b) $x\,\log_2(1/x)$.

FIGURE 5.2 – Les graphes de $\log_2(x)$ et $x\log_2(1/x)$.

Avec des arguments simples, on peut montrer la valeur la plus grande possible, pour l'entropie d'une variable aléatoire avec k valeurs possibles, est exactement celle qui correspond aux cas où ces valeurs sont équiprobables. On trouve alors que

$$0 \leq H(X) \leq \log_2(k). \tag{7}$$

Ainsi, il y a 26^n textes possibles de longueur n, écrits avec les 26 lettres de l'alphabet. On trouve donc que l'entropie d'un texte de longueur n doit se situer entre 0 et

$$\log_2(26^n) = n\,\log_2(26) \simeq 4.7\,n.$$

5.3 Quantité d'information et entropie conditionnelle

Quantité d'information

Nous allons maintenant expliquer en quoi l'entropie d'un texte mesure la densité d'information de ce texte. L'idée est ici que plus un texte est dense en contenu, moins il est possible de le remplacer par un texte plus court qui contient la même information. Pour mieux développer ce point de vue, imaginons que nous ayons réussi à condenser au maximum l'information contenue dans un texte sous la forme (typique en informatique) d'une suite

de 0 et de 1. Chacune de ces unités d'information est appelée un *bit*. Le nombre de ces bits serait alors considéré comme donnant la *quantité d'information* contenue dans le texte en question.

On peut aussi parler de la quantité d'information obtenue par le fait d'apprendre le résultat d'une expérience aléatoire. Ainsi, si on veut transmettre le résultat du lancer d'une pièce de monnaie, la façon la plus compacte correspond à écrire le nombre 0 pour pile, et 1 pour face. Cette information correspond donc à un bit d'information, tout comme l'entropie de l'expérience en question. En fait, on arrive facilement à la conclusion que cette notion de quantité d'information possède les mêmes propriétés fondamentales que l'entropie. Nous allons donc identifier les deux notions, et la discussion qui suit va mettre en évidence en quoi cette identification est bien fondée.

Pour illustrer le fait que l'entropie mesure la quantité d'information, considérons l'expérience qui consiste à piger une boule au hasard dans un boulier contenant des boules numérotées de 1 à 16. On cherche à savoir quel est le nombre minimum de questions qui devraient être posées (en moyenne), à une personne qui a pigé de façon cachée une telle boule, pour arriver à déterminer le numéro X de la boule cachée. On suppose ici que les seules réponses possibles à nos questions sont soit oui, soit non. Ainsi, chaque réponse nous apporte au plus un bit d'information. L'entropie de E est

$$H(E) = \log_2 16 = 4.$$

On interprète cette valeur comme la quantité d'information nécessaire pour déterminer à coup sûr le résultat de l'expérience E, il faut donc 4 bits d'information pour déterminer le résultat de E. On doit donc poser au moins 4 questions. Ce qui est encore plus frappant, c'est qu'il est possible de trouver les bonnes questions à poser en se guidant sur le fait qu'on cherche à obtenir, avec chaque question, la plus grande quantité d'information possible. On a vu à la section 5.2 que l'entropie (quantité d'information) est maximale dans le contexte de résultats équiprobables. Autrement dit, c'est lorsque les réponses à la question sont équiprobables qu'on obtient le plus d'information. Ceci nous amène à procéder comme suit. Avec notre première question, on cherche à couper l'ensemble des valeurs possibles, en deux sous-ensembles égaux. On pose donc à notre interlocuteur la question :

Est-ce que le numéro de la boule est supérieur à 8 ($X > 8$) ?

Les deux réponses possibles sont équiprobables, et la réponse réduit les possibilités à la moitié du nombre initial. La formulation de la seconde question dépend évidemment de la réponse à la première question. Mais l'idée demeure encore de couper en deux le nombre de résultats possibles. Il suffit de continuer à appliquer ce principe jusqu'à avoir circonscrit la

valeur cherchée. Le tableau suivant schématise les possibilités pour la seconde question, et les suivantes :

$$X>8\ ?\begin{cases}\text{oui}, & X>12\ ?\begin{cases}\text{oui}, & X>14\ ?\begin{cases}\text{oui}, & X>15\\ \text{non}, & X>13\end{cases}\\ \text{non}, & X>10\ ?\begin{cases}\text{oui}, & X>11\\ \text{non}, & X>9\end{cases}\end{cases}\\ \text{non}, & X>\ 4\ ?\begin{cases}\text{oui}, & X>\ 6\ ?\begin{cases}\text{oui}, & X>7\\ \text{non}, & X>5\end{cases}\\ \text{non}, & X>\ 2\ ?\begin{cases}\text{oui}, & X>3\\ \text{non}, & X>1\end{cases}\end{cases}\end{cases}$$

Clairement la réponse à la dernière question détermine chaque fois la valeur de X.

De façon générale, si k est l'entropie d'une expérience aléatoire avec résultats équiprobables, alors le nombre minimum de questions permettant d'identifier le résultat spécifique de l'expérience est aussi égal à k. En particulier, si le nombre de ces résultats équiprobables possibles est n, alors le nombre de questions est au moins $\log_2 n$. Bien entendu, si ce nombre n'est pas un entier, on doit l'arrondir vers le haut.

Entropie conditionnelle

Pour bien articuler notre discussion, nous allons maintenant introduire la notion d'entropie conditionnelle. Celle-ci permet de mettre en lumière la quantité d'information obtenue en apprenant quel est le résultat spécifique d'une expérience aléatoire, à propos duquel on avait déjà une certaine idée. Ainsi, si X et Y sont deux variables aléatoires, et si les valeurs possibles pour X sont $a_1, a_2, \ldots, a_k$, et celles pour Y sont $b_1, b_2, \ldots, b_n$; alors, l'entropie conditionnelle de Y étant donné qu'on connaît la valeur de X, notée $H(Y|X)$, se calcule via la formule

$$H(Y|X) := p_1\, H(Y|X=a_1) + p_2\, H(Y|X=a_2) + \ldots + p_k\, H(Y|X=a_k), \tag{8}$$

où p_i est la probabilité que la valeur de X soit a_i ; et où, comme en (9), on a

$$H(Y|X=a) := q_1 \log_2 1/q_1 + \cdots + q_n \log_2 1/q_n, \tag{9}$$

avec q_j égal à la probabilité que $Y=f_j$ sachant que $X=a$. En formule :

$$q_j = P(Y=f_j|X=a).$$

Un petit calcul supplémentaire montre que pour des variables aléatoires indépendantes, on a

$$H(Y|X) = H(Y). \tag{10}$$

On interprète l'entropie conditionnelle comme mesurant le gain d'information obtenu en moyenne lorsque, connaissant la valeur de la variable aléatoire X, on apprend qu'elle est la valeur de la variable aléatoire Y. Un exemple simple est le suivant. Considérons l'expérience aléatoire E qui consiste à prédire les conditions météorologiques pour la journée de demain. Admettons que les résultats possibles soient

a) Ensoleillée , avec probabilité 0.3.

b) Partiellement couvert , avec probabilité 0.5.

c) Partiellement couvert et pluvieux , avec probabilité 0.2.

Si Y est la variable aléatoire qui vaut 0 s'il ne pleut pas demain, et 1 si il pleut, et si X est la variable aléatoire qui vaut 1 si demain est partiellement couvert, et 0 sinon ; alors $H(Y|X) = 0.86$ mesure l'incertitude sur la possibilité de pluie demain, sachant que le temps sera partiellement couvert.

Entropie de paires de variables aléatoires

Nous allons sous peu avoir à discuter de la quantité d'information obtenue lorsqu'on apprend la valeur prise par deux (ou plus) variables aléatoires. L'entropie associée est encore ici définie selon les mêmes principes généraux, avec des détails techniques auxquels il n'est pas nécessaire de s'attarder trop longtemps. Pour simplifier notre présentation, convenons d'écrire p_{ij}, pour la probabilité que X prenne la valeur a_i en même temps que Y prend la valeur b_j. L'entropie de $H(X,Y)$ est :

$$\begin{aligned} H(X,Y) \quad = \quad & p_{11}\log_2 1/p_{11} + \cdots + p_{1n}\log_2 1/p_{1n} + \\ & p_{21}\log_2 1/p_{21} + \cdots + p_{2n}\log_2 1/p_{2n} + \\ & \cdots \\ & p_{k1}\log_2 1/p_{k1} + \cdots + p_{kn}\log_2 1/p_{kn}. \end{aligned} \tag{11}$$

Après un petit calcul (méticuleux), on trouve les jolies formules

$$H(X,Y) = H(X) + H(Y|X) \tag{12}$$

$$= H(Y) + H(X|Y) \tag{13}$$

Une observation importante (qui découle de notre discussion) est que

$$H(X,Y) \leq H(X) + H(Y), \tag{14}$$

avec égalité exactement lorsque X et Y sont des variables aléatoires indépendantes.

5.4 Systèmes cryptographiques et théorie de l'information

Nous allons maintenant aborder la cryptanalyse du point de vue de la théorie de l'information. La question qui se pose est de savoir dans quelle mesure la connaissance d'un message chiffré donne de l'information sur le message clair, en supposant que le système utilisé est connu. Afin d'étudier les qualités du système cryptographique étudié, on s'imagine que l'envoi de messages est une expérience aléatoire, qui fait intervenir trois variables aléatoires M, K et C. La première M donne le message à envoyer, puis la valeur de K correspond à la clé d'encodage choisit, et enfin C donne le message codé résultant. La situation est donc la suivante. On a

1) Le résultat de la variable aléatoire M est l'un des messages clairs possibles

$$\{m_1, m_2, \ldots, m_n\}$$

qui est choisit par l'envoyeur avec une certaine probabilité $P(M = m_i) = p_i$.

2- La variable aléatoire K donne l'une des clés possibles :

$$\{k_1, k_2, \ldots, k_s\},$$

selon le système choisit, et q_j est la probabilité que celle-ci soit k_j.

3- Enfin C donne le message codé résultant, et la probabilité d'obtenir le message codé c se calcule comme suit

$$P(c) = \sum_{f_{k_i}(m_j)=e} p_i\, q_j. \tag{15}$$

Ici, la somme s'effectue sur tous les couples (k_i, m_j) constitués d'une clé k_i et d'un message clair m_j qui donne le message codé c. Autrement dit, on considère toutes les façons d'obtenir le message codé c, à partir de divers messages clairs selon le choix d'une clé.

On supposera que l'envoyeur et le receveur choisissent une clé de codage avec un mécanisme aléatoire qui est indépendant du choix du message à coder. Autrement dit, les variables M et K sont supposées indépendantes. Il en résulte (voir formule (10)) que

$$H(K|M) = H(K). \tag{16}$$

D'autre part, lorsqu'on connaît la clé, on peut récupérer le message clair à partir du message codé. En conséquence, la quantité d'information

$$H(M) + H(K)$$

globalement obtenue lorsqu'on connaît la valeur des deux variables aléatoires indépendantes M et K ; est donc égale à la somme

$$H(C) + H(K|C)$$

de la quantité d'information, $H(C)$, qui est obtenue lorsqu'on apprend la valeur de C, et de la quantité d'information supplémentaire, $H(K|C)$, obtenue lorsqu'on découvre la clé. En formule, on a

$$H(C) + H(K|C) = H(K) + H(M). \tag{17}$$

Sous une forme un peu cachée, cette identité ne fait qu'affirmer que la connaissance du message codé ne contribue aucune information supplémentaire lorsqu'on connaît déjà le message clair et la clé. Elle nous sera utile.

Lorsqu'on peut réussir à récupérer la clé à partir de la simple analyse d'un message codé, on en conclut qu'aucune information nouvelle n'est obtenue si on nous procure effectivement la clé. Autrement dit, on répondra à l'espion qui nous offre d'acheter cette clé que cette information ne nous intéresse plus, puisque nous l'avons déjà. Du point de vue de l'entropie, cela correspond à la situation

$$H(K|C) = 0 \tag{18}$$

puisque l"entropie $H(K|C)$ donne l'incertitude sur la valeur de la clé, quand on connaît le texte codé.

Il est naturel de considérer que la quantité d'information transmise est l'entropie $H(M)$ du message en clair. Autrement dit, notre discussion sur la théorie de l'information nous assure que, quelque soit son approche, un cryptanalyste doit absolument obtenir au moins $H(M)$ bits d'information pour récupérer toute l'information contenue dans le message M. Nous allons supposer qu'il ne connaît, au départ, que le message chiffré C. Bien entendu, il sait aussi quel est le système cryptographique utilisé, mais pas la clé.

Intuitivement, un système cryptographique est parfaitement sûr si la connaissance du message chiffré n'apporte aucune information sur le message clair. En terme de la fonction d'entropie, ceci revient simplement à dire qu'un système cryptographique est *parfait* si

$$H(M|C) = H(M). \tag{19}$$

Pour continuer notre discussion de la problématique du cryptanalyste, soulignons que l'entropie $H(K|C)$ mesure la quantité d'information moyenne qu'on peut obtenir sur la clé secrète quand on connaît le texte codé, et $H(K)$ est la quantité d'information minimum qui est nécessaire (en principe) pour retrouver K. D'une même façon, $H(M|C)$ est la quantité d'information moyenne qu'on peut obtenir sur le message en clair, à partir de la connaissance du message codé correspondant. Même si ces mesures n'expliquent pas comment réaliser

ces éventuels tours de force, nous pouvons maintenant discuter des mérites théoriques d'un système cryptographique, d'une façon objective. L'étude des systèmes par substitution, développée à la section suivante, illustre bien tout ceci.

5.5 Systèmes par substitution mono alphabétique

Supposons qu'on code par substitution mono alphabétique des messages clairs (en français) ayant n caractères. Si on tient compte de la distribution des lettres dans un texte français on a vu à l'équation (6) qu'on peut estimer que

$$H(M) \leq 3.95\, n$$

Si on tient compte, en plus, des statistiques sur les bigrammes on peut faire un meilleur estimé, et calculer qu'on a certainement

$$H(M) \leq 3.2\, n$$

Dénotons X_i la variable aléatoire qui donne la i-ième lettre du message clair. L'expérience de Shanon, consistant à mesurer le nombre moyen d'essais nécessaires pour deviner la prochaine lettre d'un texte français, permet de montrer que pour $n > 15$, on a apparemment

$$H\left(X_{n+1} \mid X_1 \cdot X_2 \cdot \cdots \cdot X_n\right) \leq 1.4.$$

Si on considère ce dernier estimé avec un soupçon de scepticisme, on peut se fixer plutôt comme valeur

$$H(M) = 2\, n \tag{20}$$

Rappelons que pour un codage par substitution mono alphabétique, il y a

$$26! = 403291461126605635584000000$$

clés possibles. En supposant que le choix d'une clé est équiprobable, on trouve que

$$H(K) = \log_2 26! = 91.69 \tag{21}$$

Supposons aussi que le choix du message codé est équiprobable, alors on a

$$H(C) = \log_2(26^n) = n \log_2 26 \approx 4.7\, n \tag{22}$$

En utilisant (17), et les estimés (20), (21) et (22) ci-dessus, on obtient

$$\begin{aligned} H(K|C) &= H(K) - H(C) + H(M) \\ &= 91.69 - 4.7\, n + 2\, n. \end{aligned} \tag{23}$$

Autrement dit, pour réussir à briser un système mono alphabétique, il faut s'arranger pour obtenir un message codé de longueur n de façon à ce que $H(K|C) = 0$. En substituant cette valeur dans la formule (23), on trouve que la longueur n nécessaire est telle que

$$91,69 = 2.7\,n, \quad \text{et on trouve que} \quad n \approx 33,96.$$

En d'autres mots, un cryptogramme de 34 lettres, pour un système de codage par substitution mono alphabétique, contient en moyenne toute l'information nécessaire à récupérer la clé d'encryptage. On peut donc, en principe, reconstruire la clé à partir de ce cryptogramme, cependant le mode d'emploi n'est pas fournit.

5.6 Exercices

5.1. Deux urnes contiennent chacune 20 boules. La première urne contient 10 boules blanches, 5 noires et 5 rouges. La seconde urne contient 8 boules blanches, 8 noires et 4 rouges. On tire une boule au hasard de chaque urne. Quelle est l'expérience dont l'issue est la plus incertaine ?

5.2. Les observations météo montrent qu'au fil des ans la probabilité qu'il pleuve en un certain endroit le 15 juin est 0,4 et la probabilité qu'il n'y ait pas de précipitation est 0,6. La probabilité qu'il pleuve en ce même endroit le 15 novembre est 0,65, la probabilité qu'il neige est 0,15 et la probabilité qu'il n'y ait pas de précipitation est 0.2. Pour lequel de ces deux jours le temps est-il plus incertain.

5.3. Considérons l'expérience E qui consiste a choisir une lettre au hasard dans un texte français. Calculer l'entropie de E.

5.4. Dans le milieu médical, on sait que 2 personnes sur 100 sont atteintes d'une certaine maladie. Pour reconnaître les malades, on utilise un test qui est toujours positif quand le patient est malade, mais qui est aussi souvent positif que négatif quand le patient est sain. Soit E le résultat au test de dépistage et F l'expérience qui détermine si le patient est malade.

a) Calculer l'entropie de F

b) Calculer $H(F|E)$

c) Est-ce que l'expérience E diminue l'incertitude sur l'expérience F ? Autrement dit, est-ce que le résultat du test nous aide à savoir si une personne est atteinte de la maladie ?

5.5. On pige successivement et sans remise deux boules numérotées dans un boulier contenant 100 boules numérotées de 1 à 100. Combien de questions doit-on poser pour être certain de connaître les nombres x et y sur les boules si les seules réponses données sont oui et non.

5.6. Détermination de fausse monnaie. On a 25 pièces de monnaie de la même valeur, 24 d'entre elles ont le même poids et une est fausse et a un poids légèrement inférieur.

a) Combien de pesées sur une balance à plateaux doit-on faire sans utiliser de poids, pour déterminer quelle pièce est fausse.

b) Expliquer comment on doit faire le nombre minimal de pesées pour déterminer la fausse pièce.

5.7. *Les menteurs et les honnêtes gens.* Les habitants d'une ville A disent toujours la vérité et les habitants de la ville voisine B mentent toujours. Un étranger arrive dans la région et connaît la réputation de ces deux villes, mais il ne sait pas dans quelle ville il se trouve. Pour le découvrir, il arrête un passant et l'interroge, mais les passants ne répondent que par oui ou non et les habitants d'une ville visitent souvent la ville voisine.

a) Quel est le nombre minimum de questions que l'étranger doit poser pour savoir dans quelle ville il se trouve ?

b) Peut-on à l'aide d'une seule question déterminer dans quelle ville est débarqué l'étranger et dans quelle ville habite la personne interrogée.

Chapitre 6

Cryptographie moderne

6.1 Introduction

Les cryptosystèmes, que nous avons étudiés jusqu'à maintenant, nécessitent une clé secrète connue seulement de l'envoyeur et de son correspondant. On dit que ce sont des systèmes à *clé privée* (ou *secrète*). Nous avons vu, dans les chapitres précédents, que ces systèmes sont vulnérables parce qu'il est généralement possible de découvrir la clé à partir des messages codés (avec un peu d'ingéniosité), sauf lorsqu'on n'utilise une clé qu'une seule fois et que cette clé est au moins aussi longue que le message à envoyer. Cette dernière possibilité est cependant difficile à mettre en place, puisqu'elle nécessite un canal parallèle de transmission pour les clés secrètes. Dans un système à clé privée, la connaissance détaillée de la méthode d'encodage est considérée comme équivalente à la connaissance à la méthode de décodage. Autrement dit, on peut facilement calculer l'une à partir de l'autre.

Merkle, Hellman et Diffie, (1977)

En 1976, Diffie, Hellman et Merkle proposent d'élaborer un nouveau type de cryptosystème : la cryptographie à *clé publique*. Leur idée consiste à élaborer des cryptosystèmes autour de fonctions d'encodage pour lesquelles on ne peut découvrir la fonction de décodage qu'au prix d'un calcul qui est beaucoup trop exigeant pour se faire dans un temps raisonnable, même avec une banque d'ordinateurs des plus puissants, sauf si on connaît une certaine information (une clé) secrète. De telles fonctions sont appelées des *fonctions* à *sens unique* (ou *piège*). Comme, dans ce cas, on n'a pas à craindre qu'un adversaire décode nos messages en connaissant la fonction (ou clé) de codage, on peut rendre publique la clé de codage. Autrement dit, la clé de codage est connue de tous, et la clé de décodage n'est connue que de celui qui doit recevoir des messages. Tout en étant convaincus de la possibilité de mettre au point de tels systèmes, Diffie, Hellman et Merkle n'ont alors aucun système explicite à

Rivest, Shamir et Adelman, (1977)

proposer. Ils ont cependant ouvert la porte à toute une série de systèmes de chiffrements à clé publique. Dans ces nouveaux chiffrements, la symétrie du codage et du décodage est rompue par l'utilisation de fonctions à sens unique.

Apparemment, le premier exemple de systèmes à clé publique effectif est proposé en 1977 par Rivest, Shamir et Adelman, dans le contexte de l'arithmétique modulaire. Cependant, la même année, les services secrets britanniques autorisèrent le mathématicien Clifford Cocks à révéler qu'entre 1969 et 1975, lui et James Ellis avaient déjà mis au point le même système. Plusieurs autres systèmes ont depuis vit le jour. Les plus connus sont

- *Le chiffrement RSA* : basé sur la difficulté de la factorisation des grands entiers.
- *Le chiffrement de El Gamal* : basé sur la difficulté de résoudre le problème du logarithme discret dans un corps fini.
- *Les systèmes sur les courbes elliptiques* : basé sur la difficulté de certains calculs sur les courbes elliptiques. $\mathbb{Z}_p n$.

Pour être à même de décrire notre premier système à clé publique, introduisons quelques notions de base de la théorie des nombres.

6.2 Éléments de théorie des nombres

La décomposition en facteur des nombres entiers est un problème qui est (pour l'instant) très difficile à résoudre même avec des ordinateurs très puissants. C'est cette difficulté qui est exploitée dans le premier système à clé publique que nous allons étudier. Pour articuler notre discussion, nous aurons donc besoin de certaines notions provenant de la partie des mathématiques qui étudie les nombres entiers, les opérations sur ceux-ci, et leurs propriétés. C'est ce qu'on appelle la *théorie des nombres*. Dans ce chapitre, nous n'introduisons que les concepts de théorie des nombres qui sont essentiels à notre développement. Nous verrons à la section suivante de quelle façon chacun de ceux-ci intervient en cryptographie.

Divisibilité et nombres premiers

En premier lieu, nous devons clarifier la notion de divisibilité d'entiers, et la notion de nombre premier, qui lui est fortement reliée. On dit que b *divise* a, ou que a est *divisible* par b, si le reste de la division de a par b est 0. Dans cette situation, b est appelé *diviseur* de a. Ainsi, 3 est un diviseur de 12, mais n'est pas un diviseur de 7. Plus généralement, b divise a, exactement quand on peut écrire a sous la forme

$$a = q \cdot b,$$

pour un certain entier q. Tout entier b (≥ 1) est divisible par au moins les deux entiers 1 et b. Les entiers qui n'ont pas d'autres diviseurs que 1 et eux-mêmes sont les *nombres premiers*. Les petits nombres premiers sont

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, \ldots$$

et on peut prouver qu'il y en a une infinité.

Les nombres premiers sont les blocs de base de la *factorisation* des nombres entiers. Plus précisément, tout entier positif se décompose de façon unique comme produit de nombres premiers écrits en ordre croissant. Ainsi,

$$\begin{aligned} 100 &= 2 \cdot 2 \cdot 5 \cdot 5 = 2^2 5^2 \\ 641 &= 641 \\ 999 &= 3 \cdot 3 \cdot 3 \cdot 37 = 3^3 37 \\ 1024 &= 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^{10} \end{aligned}$$

Cependant, le problème qui consiste à trouver cette factorisation est en général très difficile calculatoirement. Par exemple, en utilisant une vaste banque d'ordinateurs et les meilleures méthodes actuellement connues, on a réussit à factoriser en 5 mois le nombre

$$\begin{aligned} &3107418240490043721350750035888567930037346022842 7 \\ &2754572016194882320644051808150455634682967172328 6 \\ &7824379162728380334154710731085019195485290073377 2 \\ &4822783525742386454014691736602477652346609 \end{aligned}$$

comme produit des deux nombres premiers

$$\begin{aligned} &1634733645809253848443133883865090859841783670033 0 \\ &9231218111085238933310010450815121211816751157 9 \end{aligned}$$

et

$$\begin{aligned} &1900871281664822113126851573935413975471896789968 \\ &5154936666385390880271038021044989571912614655 71 \end{aligned}$$

Cependant, on ne sait toujours pas trouver la factorisation du nombre

$$\begin{aligned} &7403756347956171282804679609742957314259318888923128 9 \\ &0849362326389727650340282662768919964196251178439958 9 \\ &4330502127585370118968098286733173273108930900552505 1 \\ &1687706329907239638078671008609696253793465056379635 9 \end{aligned}$$

pour laquelle un prix de $30,000 est d'ailleurs offert par la compagnie RSA Laboratories [1]. On ne connaît pas de bonne méthode de factorisation pour l'instant, et une majorité de chercheurs pense qu'il n'y a pas de solution efficace [2] à ce problème. Bien qu'il est intéressant de remarquer que si n n'est pas premier, alors n possède un facteur premier qui est plus petit ou égal à sa racine carrée $\sqrt{n}$, cela ne permet pas de réduire le problème à une taille réalisable.

Plus grand commun diviseur

Afin de construire notre système de codage, nous aurons aussi absolument besoin de savoir calculer le *plus grand commun diviseur* de deux (très grands)entiers a et b. Il s'agit là du plus grand entier qui divise à la fois a et b. Il est noté $\text{pgcd}(a, b)$, ou parfois simplement (a, b). Cette définition est un peu problématique lorsque a et b sont tous deux égaux à 0, puisque tout nombre divise 0, mais nous poserons simplement que $\text{pgcd}(0, 0) := 0$. Puisque la définition entraîne déjà que

$$\text{pgcd}(a, 0) = \text{pgcd}(0, a) = a,$$

lorsque a n'est pas zéro, le fait d'avoir ainsi choisit 0 comme valeur de $\text{pgcd}(0, 0)$ permet d'éviter d'avoir des exceptions. C'est donc un choix judicieux du point de vue de la recherche de la simplicité, rien de plus.

Si les décompositions en facteurs respectives de a et de b sont connues, on peut facilement calculer le plus grand commun diviseur de a et b. C'est d'ailleurs la méthode souvent enseignée à l'école. Le plus grand commun diviseur cherché est tout simplement la partie commune des factorisations. Ainsi, puisque

$$4200 = 2^3 \cdot 3 \cdot 5^2 \cdot 7, \qquad \text{et} \qquad 10780 = 2^2 \cdot 5 \cdot 7^2,$$

le plus grand commun diviseur de 4200 et 10780 est

$$2^2 \cdot 5 \cdot 7 = 140.$$

Lorsqu'il n'y a pas de partie commune aux factorisations de a et b, le plus grand commun diviseur est 1. Ainsi, $\text{pgcd}(17, 22) = 1$. Deux entiers a et b qui n'ont que 1 comme diviseur commun sont dits *relativement premiers* . Cette méthode «scolaire», pour le calcul du plus grand commun diviseur de deux nombres, devient tout à fait impraticable «à la main» dès

1. Voir le site www.rsasecurity.com/rsalabs/ pour les détails et mises à jour concernant ce problème, et d'autres du même genre.

2. Il serait trop long d'expliquer correctement le véritable sens de cette affirmation, mais cela correspond plus ou moins à dire que les calculs sont hors de portée des ordinateurs modernes.

que ces deux nombres sont assez grands. Ainsi, il est peu probable qu'on trouve aisément que

$$\text{pgcd}(6874009, 2673157) = 1237$$

puisqu'il faudra un relativement long travail avant de trouver la factorisation en nombres premiers

$$2673157 = 1237 \cdot 2161,$$

et de tester que 1237 est bien facteur de 6874009. De plus, on a déjà souligné plus haut qu'il est généralement reconnu que la factorisation de grands entiers est «difficile» même en procédant avec finesse avec l'aide de puissants ordinateurs. Malgré cela, nous allons voir qu'il est possible de calculer très rapidement le plus grand commun diviseur de très grands nombres ; bien entendu sans les factoriser.

Euclide d'Alexandrie (Environ -350)

6.3 L'algorithme d'Euclide

Il est étonnant qu'on se restreigne encore à n'enseigner que la méthode de calcul du plus grand commun diviseur qui a été présentée ci-haut, surtout quand on sait qu'il y a plus de 2300 ans qu'on connaît une méthode bien plus simple et efficace, connue sous le nom *d'algorithme d'Euclide* ; d'autant plus que celle-ci ne nécessite que de simples divisions d'un entier par un autre (ce qui est aussi enseigné à l'école, n'est-ce pas). De plus, c'est l'un des premiers véritables algorithmes de l'histoire des mathématiques. En cette ère d'omniprésence des ordinateurs, dont le fonctionnement est bien sûr géré par des algorithmes, cela semble être une occasion idéale d'associer à l'enseignement de notions mathématiques typiques, l'illustration d'une composante essentielle de l'informatique. On pourrait enfin ajouter à ce commentaire éditorial que ces calculs ont une application très certaine en cryptographie, et souligner l'importance de celle-ci dans notre univers moderne fort épris de communications.

L'algorithme d'Euclide fonctionne de la façon suivante. Observons d'abord que

$$\text{pgcd}(a, a) = a.$$

De plus, puisque

$$\text{pgcd}(a, b) = \text{pgcd}(b, a),$$

on peut toujours se ramener au cas où $a > b$. L'idée principale, derrière l'algorithme d'Euclide, est d'exploiter le fait (justifié plus loin) que

$$\text{pgcd}(a, b) = \text{pgcd}(b, r), \tag{1}$$

où r est le reste de la division de a par b, c'est-à-dire qu'on a

$$a = q \cdot b + r,$$

avec r qui se situe entre 0 et $b-1$. Ainsi, on aura

$$\text{pgcd}(6874009, 2673157) = \text{pgcd}(2673157, 1527695)$$

puisque

$$6874009 = 2 \cdot 2673157 + 1527695.$$

Ce qu'il est important de remarquer ici, c'est qu'on a réussi ainsi à transformer le problème original en un problème semblable, mais plus «simple», tout simplement parce que les entiers impliqués sont maintenant plus petits. Pour achever notre calcul, il suffit de recycler cette idée jusqu'à ce que le problème ait une solution évidente. Illustrons ce processus en poursuivant avec notre exemple ci-haut. On a la suite des divisions successives

$$\begin{aligned} 2673157 &= 1 \cdot 1527695 + 1145462 \\ 1527695 &= 1 \cdot 1145462 + 382233 \\ 1145462 &= 2 \cdot 382233 + 380996 \\ 382233 &= 1 \cdot 380996 + 1237 \\ 380996 &= 308 \cdot 1237 + 0 \end{aligned}$$

À ces divisions correspond la succession d'égalités

$$\begin{aligned} \text{pgcd}(2673157, 1527695) &= \text{pgcd}(1527695, 1145462) \\ &= \text{pgcd}(1145462, 382233) \\ &= \text{pgcd}(382233, 380996) \\ &= \text{pgcd}(380996, 1237) \\ &= \text{pgcd}(1237, 0) \end{aligned}$$

Comme on a déjà remarqué qu'en général $\text{pgcd}(a, 0) = a$, notre calcul est terminé, et on peut conclure que

$$\text{pgcd}(6874009, 2673157) = 1237$$

Comme on le constate, le calcul est aisé et rapide, et ne nécessite pas de factorisation. Cette approche fonctionne toujours, et on peut garantir qu'elle ne nécessitera pas beaucoup d'étapes. Techniquement, pour ceux que cela intéresse, on peut montrer que le nombre d'étapes[3] requisent est au plus $\log_\varphi(a)$, où φ est le nombre d'or :

$$\varphi = \frac{1+\sqrt{5}}{2}.$$

Autrement dit, ce nombre d'étapes est de l'ordre de grandeur du nombre de chiffres qui apparaissent dans l'écriture de a en base 10.

3. La pire situation correspond au cas où b et a sont deux nombres de *Fibonacci* successifs.

Justification de l'approche d'Euclide [4]

Pour vérifier que la simplification d'Euclide (1) est justifiée, on remarque que pour tout nombre c qui divise a et b, on a forcément que c divise aussi

$$r = a - q \cdot b.$$

En effet, le fait que c divise a et b correspond à dire que

$$a = s \cdot c, \qquad \text{et} \qquad b = t \cdot c.$$

Mais alors

$$r = s \cdot c - q \cdot t \cdot c = (s - q \cdot t) \cdot c,$$

ce qui montre bien que c divise r. De façon tout à fait semblable, on vérifie que tout nombre qui divise b et r, divise aussi a. On en déduit donc que le plus grand commun diviseur de a et b divise r, et le plus grand commun diviseur de b et r divise a. La seule façon de réconcilier tout ceci est d'avoir précisément $\text{pgcd}(a, b) = \text{pgcd}(b, r)$.

6.4 Algorithme d'Euclide étendu

Nous allons voir maintenant, comment une simple modification de l'algorithme d'Euclide permet de calculer une expression, pour le plus grand commun diviseur d d'entiers a et b, sous la forme

$$d = b \cdot x - a \cdot y \tag{2}$$

avec certains entiers x et y. Par exemple, pour $\text{pgcd}(252, 198) = 18$, on obtient (selon la méthode décrite à la section suivante) l'expression

$$18 = 252 \cdot 4 - 198 \cdot 5.$$

Il en découle aussi une méthode pour calculer l'inverse b^{-1} d'un entier b modulo n, lorsque b et n sont relativement premiers. En effet, la formule (2) entraîne que le plus grand commun diviseur

$$\text{pgcd}(n, b) = 1$$

s'exprime comme

$$b \cdot x - n \cdot y = 1,$$

ce qui correspond à dire qu'on a trouvé x tel que

$$b \cdot x \equiv 1 \pmod{n}.$$

Autrement dit, $x = b^{-1}$ est l'inverse multiplicatif de b modulo n.

4. Pour les mordus.

Euclide étendu

On peut décrire la k-ième étape de l'algorithme d'Euclide, comme le calcul

$$r_{k-2} = q_k \cdot r_{k-1} + r_k, \tag{3}$$

du quotient, q_k, et du reste, r_k, pour la division de r_{k-2} par r_{k-1}. Ici r_{k-2} et r_{k-1} correspondent aux restes trouvés aux deux étapes précédentes. Bien entendu, l'algorithme démarre avec

$$r_0 = a, \qquad r_1 = b,$$

et la dernière étape est celle pour laquelle $r_m = 0$. Au fur et à mesure du déroulement de l'algorithme d'Euclide, nous allons calculer deux suites d'entiers

$$x_0,\ x_1,\ x_2,\ x_3,\ \ldots\ , x_{m-1}$$
$$y_0,\ y_1,\ y_2,\ y_3,\ \cdots\ , y_{m-1}$$

avec m correspondant, comme ci-dessus, au numéro de la dernière étape de calcul de l'algorithme d'Euclide. Les nombres cherchés x et y, afin d'obtenir (2), s'obtiennent alors respectivement comme $(-1)^m x_{m-1}$ et $(-1)^m y_{m-1}$. On débute le calcul des x_k et y_k en posant

$$x_0 = 0, \quad \text{et} \quad x_1 = 1$$
$$y_0 = 1, \quad \text{et} \quad y_1 = 0,$$

puis on construit chacun des termes suivant avec la recette :

i) $x_k = q_k \cdot x_{k-1} + x_{k-2}$

ii) $y_k = q_k \cdot y_{k-1} + y_{k-2}$,

On peut alors vérifier «récursivement» qu'avec les valeurs finales

$$x = (-1)^m x_{m-1}, \qquad \text{et} \qquad y = (-1)^m y_{m-1},$$

on a bien l'identité

$$b \cdot x - a \cdot y = d$$

où d est le plus grand commun diviseur de a et b. Il est à remarquer que, si le seul but du calcul est de trouver l'inverse multiplicatif de b modulo n, alors il n'est pas nécessaire de calculer la suite des y_k, seulement celle des x_k.

Un exemple

Calculons l'inverse de $b = 317$ modulo $n = 521$. Les étapes de l'algorithme d'Euclide correspondent aux divisions

$$\begin{array}{rcll} 521 &=& 1 \cdot 317 + 204 & q_2 = 1 \quad r_2 = 204 \\ 317 &=& 1 \cdot 204 + 113 & q_3 = 1 \quad r_3 = 113 \\ 204 &=& 1 \cdot 113 + 91 & q_4 = 1 \quad r_4 = 91 \\ 113 &=& 1 \cdot 91 + 22 & q_5 = 1 \quad r_5 = 22 \\ 91 &=& 4 \cdot 22 + 3 & q_6 = 4 \quad r_6 = 3 \\ 22 &=& 7 \cdot 3 + 1 & q_7 = 7 \quad r_7 = 1 \\ 3 &=& 3 \cdot 1 + 0 & \end{array}$$

On a donc $m = 8$. On obtient la suite des x_k, en utilisant pour les construire les valeurs q_k trouvées ci-haut :

$$\begin{aligned} x_0 &= 0 \\ x_1 &= 1 \\ x_2 &= 1 \cdot x_1 + x_0 = 1 \\ x_3 &= 1 \cdot x_2 + x_1 = 2 \\ x_4 &= 1 \cdot x_3 + x_2 = 3 \\ x_5 &= 1 \cdot x_4 + x_3 = 5 \\ x_6 &= 4 \cdot x_5 + x_4 = 23 \\ x_7 &= 7 \cdot x_6 + x_5 = 166 \end{aligned}$$

L'inverse multiplicatif cherché est donc

$$(-1)^8 x_7 = 166.$$

6.5 Exponentiation modulo n

Lorsqu'on calcule la suite des valeurs $(a^e \bmod n)$, pour $a = 2, 3, 4, \ldots$, avec e assez grand, on constate que les résultats apparaissent de façon très aléatoirement dans l'ensemble des entiers modulo n. Par exemple, la suite des nombres

$$(2^{17} \bmod 31), \quad (3^{17} \bmod 31), \quad (4^{17} \bmod 31), \quad \ldots \quad , \quad (29^{17} \bmod 31)$$

donne (dans cet ordre) les résultats très variables

$$4, 22, 16, 25, 26, 18, 2, 19, 7, 3, 11, 17, 10, 23, 8, 21, 14, 20, 28, 24, 12, 29, 13, 5, 6, 15, 9, 27$$

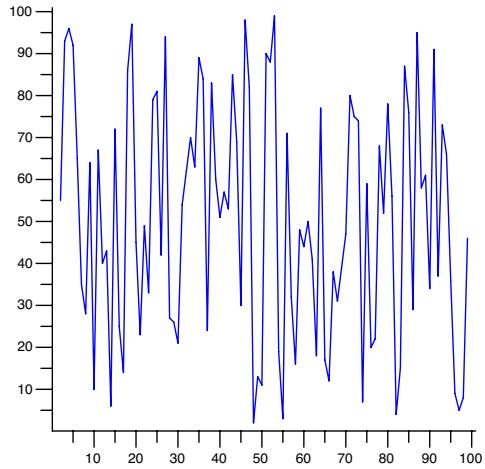


FIGURE 6.1 – Graphe de la fonction $(a^{37} \bmod 101)$.

Autrement dit, on a un comportement erratique du graphe de la fonction $(a^e \bmod n)$, comme illustré à la figure 6.1 avec $e = 37$ et $n = 101$. Ce phénomène est encore plus marquant si a, e et n sont de grands nombres entiers. À toute fin pratique, il paraît impraticable d'essayer de récupérer directement a à partir de $(a^e \bmod n)$, puisqu'à de petites variations dans la valeur de a, correspondent de grandes variations dans les valeurs correspondantes de $(a^e \bmod n)$. C'est ce phénomène qui est utilisé pour coder des messages dans la méthode développée par Rivest, Shamir et Adelman, dans leur système maintenant connu sous le nom RSA. D'autre part, la sécurité de leur système dépend de la difficulté de factoriser de grands entiers. Nous y reviendrons.

Pour l'instant, nous allons d'abord voir qu'on peut facilement calculer

$$(a^e \bmod n)$$

même lorsque a, e et m sont de très grands (plus de 100 chiffres !). L'idée est très simple, elle consiste à exploiter judicieusement les lois sur les exposants, et le fait qu'on calcule modulo n, pour conserver relativement petits les entiers à manipuler. Autrement dit, on veut éviter d'avoir trop d'étapes de calcul, et d'avoir une inflation galopante dans la taille des entiers

à manipuler. Cette inflation est apparente lorsqu'on calcule quelques puissances de 2 :

$$\begin{aligned}
2^{10} &= 1024 \\
2^{20} &= 1048576 \\
2^{30} &= 1073741824 \\
2^{40} &= 1099511627776 \\
2^{50} &= 1125899906842624 \\
2^{60} &= 1152921504606846976 \\
2^{70} &= 1180591620717411303424 \\
2^{80} &= 1208925819614629174706176 \\
2^{90} &= 1237940039285380274899124224 \\
2^{100} &= 1267650600228229401496703205376
\end{aligned}$$

surtout quand on pense que nous envisageons de calculer des puissances de loin plus grandes que 100, comme

$$2^{1446283347341906077815323861918008631842476492257561233594446287611332967286114578384292139}$$

qui est un nombre de plus de 10^{90} chiffres. La longueur d'un ruban sur lequel on tenterait d'écrire ce chiffre devrait correspondre à plusieurs fois la le diamètre de l'univers connu. Malgré cela, nous allons voir qu'il est facile de calculer très rapidement ce nombre, quand on travaille modulo un entier d'une centaine de chiffres. En particulier, nous allons voir qu'on peut même calculer à la main la valeur de

$$(2^{1024} \bmod 23),$$

bien que 2^{1024}, sans prendre de modulo, donne le nombre astronomique :

$$\begin{aligned}
&179769313486231590772930519078902473361797697894230657273430081157732 \\
&675805500963132708477322407536021120113879871393357658789768814416622 \\
&492847430639474124377767893424865485276302219601246094119453082952085 \\
&005768838150682342462881473913110540827237163350510684586298239947245 \\
&938479716304835356329624224137216
\end{aligned}$$

Pour calculer de grandes puissances modulo n, on procède comme suit. On observe d'abord que les règles de calculs pour les exposants sont aussi valables modulo n. En particulier, on a les identités

$$a^{2k} \equiv (a^2)^k \pmod{n}, \qquad \text{et} \qquad a^{2k+1} \equiv (a^2)^k \cdot a \pmod{n}, \tag{4}$$

et on peut choisir de remplacer a^2 par son reste modulo n, chaque fois que cela permet de simplifier le calcul. Illustrons avec le calcul de 2^{1024} modulo 23. On a

$$2^{1024} = (2^2)^{512} = (4^2)^{256} = (16^2)^{128},$$

et donc

$$2^{1024} \equiv 3^{128} \pmod{23},$$

puisque $16^2 = 256$ et $256 \equiv 3 \pmod{23}$. On peut continuer notre calcul en remarquant que

$$3^{128} = (3^2)^{64} = (9^2)^{32},$$

et, puisque $81 \equiv 12 \pmod{23}$, on a maintenant

$$2^{1024} \equiv 12^{32} \pmod{23}.$$

Le reste du calcul donne

$$\begin{aligned} 2^{1024} &\equiv (12^2)^{16} \pmod{23} \\ &\equiv (6^2)^8 \pmod{23} \\ &\equiv (13^2)^4 \pmod{23} \\ &\equiv (8^2)^2 \pmod{23} \\ &\equiv 18^2 \pmod{23} \\ &\equiv 2 \pmod{23} \end{aligned}$$

Lorsqu'en cours de route l'exposant est impair, on utilise la seconde règle en (4).

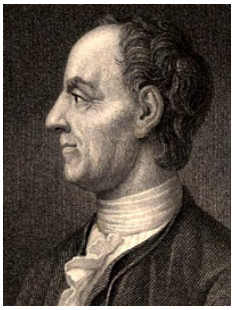
Pierre de Fermat (1601-1665)

Le théorème d'Euler-Fermat

Pour l'une des rares fois (sinon la seule) nous allons utiliser le terme «théorème» dans notre présentation. C'est là pour insister sur l'élégance d'une très jolie propriété des calculs d'exposants modulo un entier n. Un cas particulier de ce théorème a été d'abord obtenu par Pierre de Fermat , puis il a été généralisé par Leonhard Euler. Nous aurons besoin de la version plus générale d'Euler. Celle-ci passe par l'introduction d'en nouvelle fonction, la *fonction φ d'Euler*, qui est définie comme suit. Pour un entier n, on compte combien il y a d'entiers k, situés entre 1 et $n-1$, pour lesquels on a $\text{pgcd}(n,k) = 1$. Ce nombre est dénoté $\varphi(n)$. Ainsi, on a les valeurs suivantes, accompagnées ici d'une liste des entiers qui

sont comptés par $\varphi(n)$.

$$\begin{array}{rcll} \varphi(2) & = & 1 & \{1\} \\ \varphi(3) & = & 2 & \{1,2\} \\ \varphi(4) & = & 2 & \{1,3\} \\ \varphi(5) & = & 4 & \{1,2,3,4\} \\ \varphi(6) & = & 2 & \{1,5\} \\ \varphi(7) & = & 6 & \{1,2,3,4,5,6\} \\ \varphi(8) & = & 4 & \{1,3,5,7\} \\ \varphi(9) & = & 6 & \{1,2,4,5,7,8\} \\ \varphi(10) & = & 4 & \{1,3,7,9\} \end{array}$$


Leonhard Euler (1707-1783)

Il n'est pas difficile de conclure, à partir de la définition, que $\varphi(p) = p - 1$, quand p est un nombre premier. Nous serons particulièrement intéressés par le fait que

$$\varphi(p \cdot q) = (p-1) \cdot (q-1),$$

chaque fois qu'on a deux nombres premiers distincts p et q. Cette égalité s'obtient en observant que les seuls nombres d'entiers, entre 1 et $p \cdot q - 1$, qui ne sont pas relativement premier [5] à $p \cdot q$ sont exactement les nombres :

$$\begin{array}{l} p,\ 2\,p,\ 3\,p,\ \ldots\ ,(q-1)\cdot p \\ q,\ 2\,q,\ 3,q,\ \ldots\ ,(p-1)\cdot q \end{array}$$

Théorème 6.5.1 (Euler-Fermat)**.** *Si a est relativement premier à n, alors*

$$a^k \equiv a^\ell \pmod{n}$$

exactement lorsque

$$k \equiv \ell \pmod{\varphi(n)}.$$

Le cas particulier que nous utiliserons correspond à $n = p \cdot q$ (produit de deux nombres premiers distincts), dans la cas où $k = e \cdot f$, et $\ell = 1$. On a donc

$$a^{e \cdot f} \equiv a \pmod{n} \tag{5}$$

exactement quand

$$e \cdot f \equiv 1 \pmod{\varphi(n)}, \tag{6}$$

avec $\varphi(n) = (p-1) \cdot (q-1)$. Observons ici que l'équation (6) dit très exactement que f est l'inverse multiplicatif de e modulo $\varphi(n)$.

5. On dit que a et n sont relativement premiers si $\text{pgcd}(a,n) = 1$.

6.6 Le système RSA

Nous sommes maintenant prêts à décrire le système de cryptographie à clé publique introduit par Rivest, Shamir et Adelman. Dans ce système, chaque participant se construit une clé de la façon suivante :

1. Il commence par choisir [6] en secret 2 très grands nombres premiers p et q (avec au moins 100 chiffres chacun), et il calcule $n = p \cdot q$.
2. Il est donc à même de calculer la fonction $\varphi(n) = (p-1)(q-1)$. Il choisit (voir ci dessous) alors un entier e, au hasard entre 1 et $\varphi(n)$, qui est relativement premier à $\varphi(n)$.
3. Il peut alors calculer l'inverse multiplicatif $f = e^{-1}$, de e modulo $\varphi(n)$.
4. Enfin, le participant rend publique sa clé d'encodage, (n, e), et garde secrète la clé de décodage f.

En supposant que chaque participant ait réussi à réaliser ces étapes, on publie un annuaire donnant la clé (n, e) de chaque participant. Nous verrons qu'en toute probabilité, ces clés sont forcément distinctes.

Pour coder un message à l'intention d'un certain participant, on consulte cet annuaire de clés pour obtenir la valeur particulière de n et de e qui lui correspond. L'encodage procède de la manière suivante. On commence par découper le message à envoyer, en morceaux dont la longueur est plus petite que la moitié du nombre de chiffres dans n. On numérise un de ces morceaux en remplaçant chaque lettre par deux chiffres de la façon suivante

$$\texttt{a} \mapsto 10,\ \texttt{b} \mapsto 11, \texttt{c} \mapsto 12, \ldots$$

Par exemple, on a la numérisation

$$\texttt{bonjour} \mapsto 11242319243027.$$

Les morceaux du message sont ainsi devenus de grands entiers modulo n (parce que leur longueur est plus petite que n). L'encodage d'un morceau numérisé a se fait en calculant

$$b := (a^e \bmod n)$$

Pour décoder ce message, on cherche à récupérer a à partir de b. Comme nous allons mieux le voir plus loin, cela est une entreprise très difficile, sauf si l'on connaît f. Dans ce cas, il suffit en effet de calculer

$$\begin{aligned}(b^e \bmod n) &= ((a^e)^f \bmod n)\\ &= a,\end{aligned}$$

étant donné l'équation (5). On dé-numérise ensuite a pour récupérer le message envoyé.

6. Nous allons discuter plus loin comment réaliser cet exploit

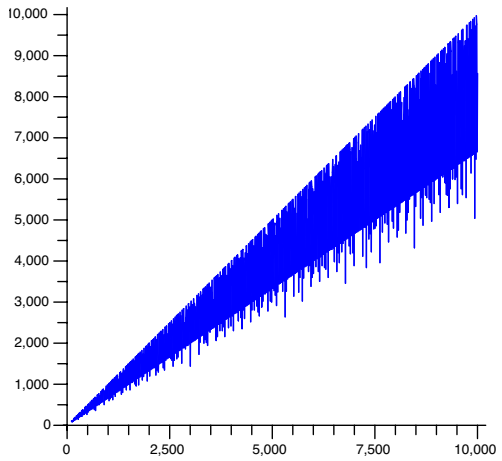


FIGURE 6.2 – Graphe de la fonction φ d'Euler.

6.7 Sécurité du système RSA

Pour avoir une meilleure idée de la sécurité du système RSA, nous allons en discuter certains aspects. Une discussion technique (omise ici) permet de montrer que les attaques les plus efficaces, contre le système RSA, sont équivalentes à la factorisation de l'entier n. Pratiquement, sans connaître cette factorisation, on ne peut pas calculer la valeur de $\varphi(n)$. Une indication de la difficulté de calculer directement $\varphi(n)$ est bien mise en évidence quand on considère le graphe de la fonction φ de la figure 6.2. On constate en effet que la valeur de la fonction oscille très rapidement, et avec de grands écarts. Autrement dit, on peut difficilement prédire la valeur de $\varphi(n)$, à partir des valeurs précédentes. Bien entendu, $\varphi(n)$ est très facile à calculer lorsque la factorisation de n est connue.

La situation pour φ est semblable au problème de la factorisation, en ce qu'il est impossible de prédire la factorisation d'un entier particulier à partir de la factorisation des entiers précédents. L'impression générale des chercheurs dans le domaine est que la factorisation demeurera un problème difficile nécessitant des calculs qui dépassent la capacité des ordinateurs actuels, si on espère une réponse dans un temps raisonnable. Cependant, on arrive lentement à repousser le seuil des entiers les plus gros qu'il est possible de factoriser, en quelques mois ou années, avec des méthodes de plus en plus fines. Il faut donc continuer à prendre les affirmations sur la sécurité de RSA avec un grain de sel (ou de sagesse). Pour

l'instant, si les nombres p et q, servant à construire la clé de RSA, sont choisi avec plus de 150 chiffres, la factorisation de $n = p \cdot q$ semble être complètement hors de portée des ordinateurs modernes.

6.8 Recherche de grands nombres premiers

Il nous reste à clarifier deux des éléments du processus de construction des clés, qui a été décrit à la section . Un premier élément consiste à savoir choisir au hasard un grand nombre. Ceci est tout simple, il suffit de choisir[7] successivement des chiffres entre 0 et 9 (sauf le premier qui est choisit entre 1 et 9) pour obtenir un nombre de la longueur voulue.

Pour la troisième étape de la construction de la clé, on doit choisir au hasard un nombre e, relativement premier à $\varphi(n) = (p-1)\cdot(q-1)$. Pour ce faire, on choisit d'abord m, un grand nombre au hasard plus petit que $\varphi(n)$. On calcule alors facilement $d = \text{pgcd}(m, \varphi(n))$, puis on pose

$$e := \frac{m}{d}.$$

Le résultat est forcément relativement premier à $\varphi(n)$, et cette partie de la construction est terminée.

La seule étape qui nous reste est plus délicate à réaliser. Elle consiste à trouver deux grands nombres premiers p et q. Avant de procéder, nous allons d'abord nous s'assurer qu'il est possible de trouver une grande quantité de nombres premiers de grandes longueurs. En fait, les deux aspects sont liés. C'est parce qu'il y a un beaucoup de nombres premiers, et qu'ils se retrouvent assez fréquemment, qu'il sera facile d'en trouver. L'énoncé central est que la fréquence d'apparition de nombres premiers, à proximité d'un entier n, est très précisément $1/\log(n)$. Autrement dit, $1/\log(n)$ est la probabilité d'obtenir un nombre premier, lorsqu'on choisit au hasard un nombre parmi ceux qui sont « proches» de n. Ainsi, on a environ une chance sur 500 de piger un nombre premier, lorsqu'on pige un nombre de 200 chiffres au hasard.

Test de primalité

La possibilité d'obtenir facilement de grands nombres premiers dépend de cette assez grande fréquence de nombres premiers, alliée au fait qu'il est possible de tester facilement (rapidement) si un grand nombre est premier. Supposons donc qu'on a une certaine procédure

7. On trouve des dés à 10 faces dans tous les bons magasins ... euh, peut-être.

pour tester si un nombre est premier, c'est ce que nous appellerons un *test de primalité*. Pour obtenir un grand nombre premier, on procède alors comme suit :

1) On choisit un nombre k (impair) au hasard de la longueur voulue.
2) Si k est déclaré premier par notre test, on a terminé.
3) Sinon, on ajoute 2 à k et on teste la primalité du résultat. Cette étape est répétée jusqu'à ce qu'on ai trouvé.

Le fait que ce processus fonctionne bien, résulte de ce qu'il y a fréquemment des nombres premiers. On peut même obtenir des résultats précis sur le nombre moyen d'étapes nécessaires avant de trouver un nombre premier. Tout peut se faire très rapidement dans la pratique, et en quelques secondes on obtient les nombres premiers nécessaires.

Il y a plusieurs tests de primalité allant de ceux qui sont faciles à décrire, mais qui ne fonctionnent pas toujours parfaitement, à ceux qui sont très précis, mais beaucoup plus délicats à décrire en détail. Pour donner une petite idée, nous allons en présenter un qui est probabiliste, et qui ne nécessite aucune nouvelle notion. Il est basé sur l'observation suivante. Si n n'est pas premier, alors (sauf pour de très rares nombres exceptionnels) il y a une chance sur 2 pour que

$$a^n \not\equiv a \pmod{n} \tag{7}$$

pour a entre 2 et $n-1$ (pour 1 on a toujours égalité). Rappelons que, si p est premier, on a toujours

$$a^p \not\equiv a \pmod{n}$$

pour a entre 2 et $p-1$. C'est en fait un cas spécial du théorème d'Euler-Fermat. Pour tester la primalité (ou non) d'un nombre k, on choisit au hasard un nombre entre 2 et $k-1$ et on calcule

$$(a^k \bmod n).$$

si le résultat n'est pas a, on est certain que k n'est pas premier. Si le résultat est bien a, on peut penser que k est premier avec probabilité de 1/2 de se tromper. L'idée simple, mais efficace est tout simplement de répéter ce test pour de nombreuses valeurs différentes de a. Si pour l'une d'entre elles on n'a pas l'égalité, on a la certitude que k n'est pas premier. Mais si on a égalité pour chacune des valeurs testées, alors on peut déclarer que k est premier, avec probabilité $(1/2)^m$, où m est le nombre de valeurs testées. Ainsi, si $m = 20$, la probabilité d'avoir raison est plus grande que 0.999999.

6.9 Logarithme discret

Nous allons maintenant décrire un autre système de cryptographie à clé publique, encore basé sur des calculs avec des entiers modulaires. Lorsqu'on travaille avec les nombres réels, le

calcul du nombre x qui fait en sorte que $y = b^x$, pour y et b donnés, correspond à trouver le *logarithme*, $\log_b(y)$. Dans le cas des entiers modulo un nombre premier, on peut considérer le même problème ; parce que les puissances successives b^x parcourent toutes les entiers ($\neq 0$) modulo p, lorsque x parcourt tous les entiers modulo $p-1$. Ainsi, avec $p = 11$, on a

y	1	2	3	4	5	6	7	8	9	10
$\log_6(y)$	0	9	2	8	6	1	3	7	4	5

On obtient ainsi la notion de *logarithme discret*. Ici, le mot discret sert à distinguer de la notion usuelle qu'on qualifie souvent de continue. Pour p un nombre premier, et b entre 2 et $p-2$, on peut reformuler le théorème d'Euler-Fermat comme

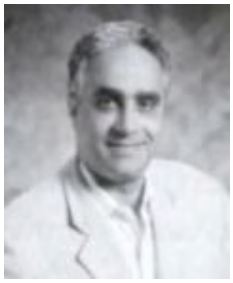
$$y \equiv b^x \pmod{p}$$

exactement lorsque

$$x \equiv \log_b(x) \pmod{p-1}.$$

Fixons au départ un grand nombre premier p, et on choisit un grand nombre b dans les entiers modulo p. Nous avons vu que, pour un x grand, on peut calculer rapidement une expression $b^x \pmod{p}$. Par contre, il est très difficile de calculer $\log_b y \pmod{p-1}$ pour de grands entiers y. Voici un cryptosystème qui utilise les propriétés du logarithme discret.

Le cryptosystème d'Elgamal


Taher Elgamal

Dans ce système, on suppose que les blocs de message clair sont numérisés dans les entiers modulo p. On commence par choisir un grand nombre premier p, et un nombre g (modulo p), qui sont tous deux connus de tous. L'utilisateur A choisit un grand nombre a (modulo $p-1$) qui sera sa clé secrète de décodage. La clé publique de A est le nombre $g^a \pmod{p}$. Pour envoyer un message m à A, l'utilisateur B choisit aléatoirement un grand entier k (modulo p), et il envoie à A la paire

$$(K, M), \quad \text{où} \quad K = (g^k \bmod p), \quad \text{et} \quad M = (m \cdot g^{a \cdot k} \bmod p).$$

Le receveur A, qui connaît la clé secrète a, récupère le message m à partir de cette paire de la façon suivante. Il calcule [8] d'abord $(K^{-a} \bmod p) = (g^{-a \cdot k} \bmod p)$, à partir du premier élément du couple reçu ; puis il multiplie M par ce résultat pour obtenir

$$\begin{aligned} M \cdot g^{-a \cdot k} &\equiv (m \cdot g^{a \cdot k}) \cdot g^{-a \cdot k} \pmod{p} \\ &\equiv m \cdot g^{a \cdot k - a \cdot k} \pmod{p} \\ &\equiv m \end{aligned}$$

8. Il ne faut pas oublié ici que les exposants se calculent modulo $p-1$. On peut donc remplacer le nombre négatif $-a$, par $p-1-a$.

Intuitivement, le message codé M envoyé à A est une version masquée de m obtenue par la multiplication par $g^{a\cdot k}$. Le nombre K, qui accompagne le message codé M, est un indice qui permet à A de retirer le masque. Cet indice $K = (g^k \bmod p)$ ne peut être utilisé que par quelqu'un qui connaît la clé a. Il semble que pour qu'un cryptanalyste puisse casser le cryptosystème de Elgamal, il doive retrouver la clé a à partir de la clé publique g^a. C'est donc dire qu'il aura trouvé une solution efficace au problème du calcul du logarithme discret. Les experts du domaine ont tendance à croire que c'est la seule possibilité. Plus précisément, on a la conjecture [9] de Diffie-Hellman :

Il est impossible de calculer $g^{a\cdot b}$ en ne connaissant que g^a et g^b, dans les entiers modulo p.

Pour illustrer le système, on s'imagine que A (Alice) et B (Bob) veulent mettre au point un cryptosystème d'Elgamal et choisissent le nombre premier $p = 2579$ et le nombre $g = 2$. Supposons que A choisit le nombre $a = 765$ comme clé secrète. La clé publique g^a, de A, est donc

$$2^{765} \equiv 949 \pmod{2579}$$

Pour encoder le message $m = 1299$, B choisit aléatoirement le nombre secret $k = 1824$ et envoi à A le couple

$$(2^{1824}, 1299 \cdot 2^{765\cdot 1824}) \equiv (2217, 2484) \pmod{2579}.$$

Pour décoder, puisque $-765 \equiv 1813 \pmod{2578}$, A calcule

$$\begin{aligned} 2484 \cdot 2217^{-765} &\equiv 2484 \cdot 2217^{1813} \pmod{2579} \\ &\equiv 2484 \cdot 2321 \pmod{2579} \\ &\equiv 1299 \pmod{2579}. \end{aligned}$$

Et A récupère le message en clair.

9. C'est le terme utilisé en mathématique pour désigner un énoncé pour lequel on a beaucoup de raisons de croire qu'il est vrai, mais pour lequel on n'a pas encore de preuve.

Chapitre 7

Pour les mordus

La recherche de systèmes cryptographique à clés publiques de plus en plus performants et sûrs, à mené les spécialistes du domaine à l'utilisation d'outils mathématiques de plus en plus sophistiqués. Pour illustrer cette tendance, nous allons expliquer une approche qui s'inspire des propriétés de certains objets géométriques : les *courbes elliptiques*. Nous ne présentons que l'information minimale nécessaire sur les courbes elliptiques pour pouvoir apprécier et développer les applications en cryptographie.

7.1 Courbes elliptiques

Le fil conducteur de notre approche consiste à traduire dans le contexte des calculs modulo un nombre premier, des constructions géométriques qui possèdent de très jolies propriétés. Nous allons développer ces constructions dans le plan cartésien usuel, puis elles seront traduites dans le contexte très calculatoire des entiers modulo q, un grand nombre premier fixé.

Une courbe elliptique, dans le plan cartésien, est l'ensemble des points (x, y) du plan qui satisfont une équation de la forme

$$y^2 = x^3 + ax + b, \tag{1}$$

avec a et b certains nombres réels. Pour des raisons techniques, nous allons supposer que a et b ont été choisi de façon à ce que

$$4\,a^3 + 27\,b^2 \neq 0. \tag{2}$$

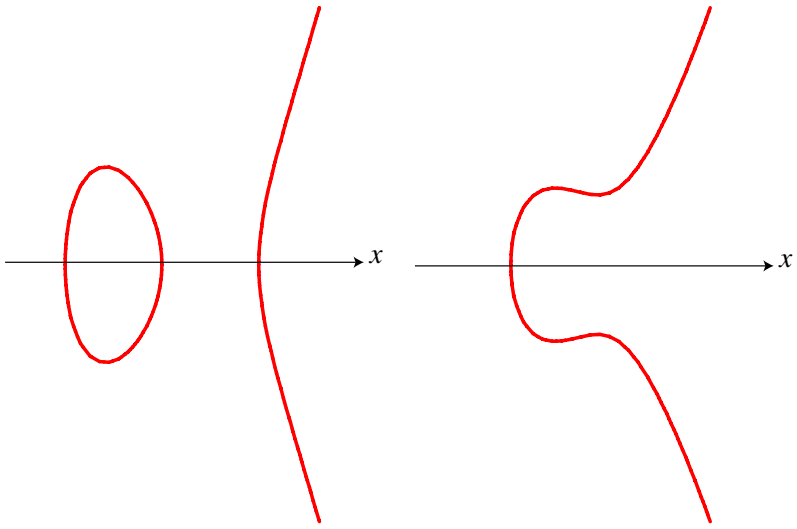


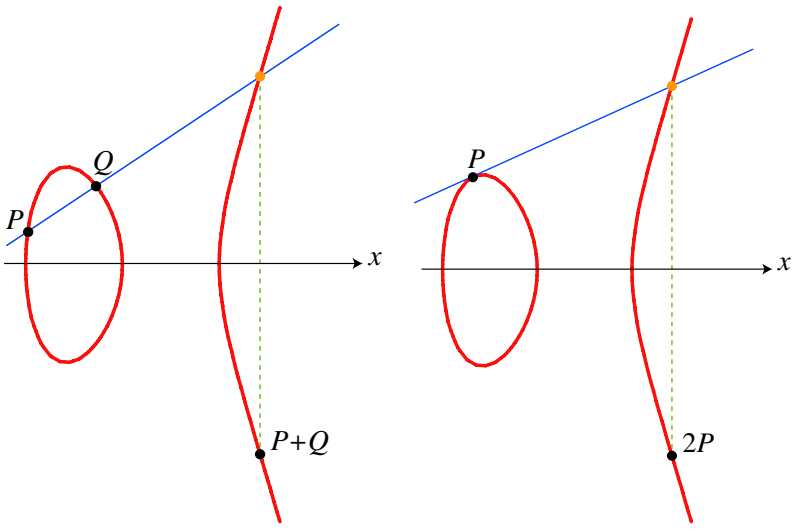
FIGURE 7.1 – Deux courbes elliptiques typiques

Typiquement, ces courbes prennent l'une des deux formes [1] illustrées à la figure 7.1. On peut observer que ces courbes sont toujours symétriques par rapport à l'axe des x. Autrement dit, la partie sous l'axe des x est l'image miroir de la partie au-dessus de l'axe des x. Cela résulte essentiellement du fait qu'on a y^2 dans le membre de gauche de l'équation (1), qui définit la courbe. Cependant, pour nos besoins, la propriété cruciale de ces courbes est la suivante (voir figure 7.2).

Une droite coupe une courbe elliptique en au plus trois points.

En effet, on peut exploiter cette propriété remarquable pour introduire abstraitement une *opération d'addition* sur l'ensemble des points de la courbe. De prime abord, une telle démarche peut sembler mystérieuse et surprenante. En fait, elle est le résultat de travaux de Leonhard Euler (au dix-huitième siècle) visant à développer des généralisations importantes des lois de la trigonométrie. Pour définir l'addition de deux points P et Q sur une courbe elliptique, on considère la droite qui passe par les points P et Q, voir figure 7.2. Sauf pour quelques cas exceptionnels (discutés ci-dessous), cette droite coupe la courbe en exactement un autre point. On définit alors la *somme* $P + Q$, comme étant la réflexion de ce point par rapport à l'axe des x. Si on cherche à additionner un point P avec lui-même, cette dernière

1. On doit souligner ici que la courbe de gauche est constituée de deux morceaux.



Cas 1) Addition de points Cas 2) Doubler un point

FIGURE 7.2 – Opération d'addition sur une courbe elliptique.

construction n'a plus de sens. La façon toute naturelle de faire est alors de considérer la droite qui est tangente à la courbe au point P. Dans tous les cas, cette droite tangente coupe la courbe en au plus un autre point. Quand c'est le cas, on pose encore que $2\,P = P + P$ est la réflexion de cet autre point par rapport à l'axe des x. On dit alors qu'on a *doublé* P.

Quelques cas exceptionnels restent à discuter pour combler les lacunes laissées par les situations générales considérées plus haut. Nous allons tout d'abord convenir que l'image miroir de P, par réflexion par rapport à l'axe de x, est $-P$. On veut en effet pouvoir additionner ces deux points, avec comme résultat O, un nouveau point spécial qui joue un rôle analogue au nombre 0 usuel. Si on cherche à retrouver ce point sur la droite (parallèle à l'axe de y) qui joint P et $-P$, le choix qui s'impose (après réflexion) est d'imaginer que ce point est à «l'infini». Il faut inclure dans ce cas la situation limite correspondante à $P = -P$, c'est-à-dire que P se trouve sur l'axe des x. On trouve alors que $2P = O$. On peut montrer qu'on a ainsi couvert toutes les possibilités, en posant $P + O = P$ pour clarifier comment additionner n'importe quel point au point spécial à l'infini.

Calcul des coordonnées de $P + Q$

Rappelons que notre but véritable est de transposer toute cette discussion au contexte des entiers modulo un nombre premier q. Pour ce faire, nous allons décrire comment calculer les coordonnées de $P + Q$, à partir de celles de P et de Q. On aura alors une recette de calcul de la somme, facile à traduire dans le contexte de calculs modulaires.

Pour la courbe elliptique donnée par la formule (1), supposons que les coordonnées de P, Q et $P + Q$ soient respectivement (x_1, y_1), (x_2, y_2) et (x_3, y_3). On considère deux situations possibles, illustrées à la figure 7.2 :

1. Les points P et Q sont distincts et ne sont pas sur la même droite verticale. Alors on pose :
$$m := \frac{y_2 - y_1}{x_2 - x_1} \tag{3}$$
2. Les points P et Q sont égaux et ne sont pas sur l'axe des x, et on pose
$$m := \frac{3x_1^2 + a}{y_1 + y_2} \tag{4}$$

Dans les deux cas, les coordonnées (x_3, y_3) de $P + Q$ s'obtiennent par les formules :

$$x_3 = m^2 - (x_1 + x_2), \tag{5}$$
$$y_3 = -y_1 + m\,(x_1 - x_3) \tag{6}$$

Ainsi, sur la courbe elliptique d'équation $y^2 = x^3 - 25\,x$, pour les points $P = (-4, 6)$ et $Q = (-3, 4\sqrt{3})$, on trouve

$$P + Q = (91 - 48\sqrt{3}, 1140 - 668\sqrt{3}), \qquad \text{et} \qquad 2P = (1681/144, 62279/1728).$$

Courbes elliptiques sur les entiers modulaires.

On suppose maintenant qu'on travaille avec l'ensemble des nombres modulo un nombre premier q. La courbe elliptique d'équation

$$y^2 \equiv x^3 + ax + b \pmod{q},$$

est l'ensemble des couples (x, y) avec x, et y des entiers modulo q. C'est donc un certain sous-ensemble des q^2 couples possibles. À titre d'exemple, pour $q = 23$, les points de la courbe elliptique

$$y^2 = x^3 + x,$$

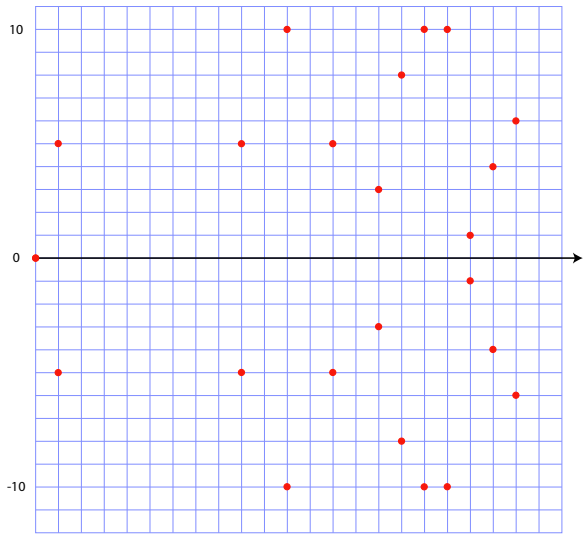


FIGURE 7.3 – Un courbe elliptique modulo q.

sont les 23 points

$$\begin{array}{lllllll} (0,0), & (1,5), & (1,-5), & (9,5), & (9,-5), & (11,10), & (11,-10), \\ & (13,5), & (13,-5), & (15,3), & (15,-3), & (16,8), & (16,-8), \\ & (17,10), & (17,-10), & (18,10), & (18,-10), & (19,1), & (19,-1), \\ & (20,4), & (20,-4), & (21,6), & (21,-6) & & \end{array}$$

illustrés à la figure 7.4. L'addition de points, pour les courbes elliptiques modulo q, se fait avec les mêmes formules que dans le cas précédent, sauf que tous les calculs se font maintenant modulo q.

7.2 Cryptosystèmes elliptiques

Nous allons construire un cryptosystème basé sur l'addition de points sur des courbes elliptiques modulo q. À la base des calculs propres à ce système se retrouve la notion de *multiples d'un point*. Plus spécialement, nous allons exploiter une forte analogie entre le calcul de la puissance n-ième d'un entier, et celui du multiple $n\,P$ d'un point sur une courbe elliptique. Par exemple, on a les multiples successifs suivants, pour le point $P = (11, 115)$ sur la courbe

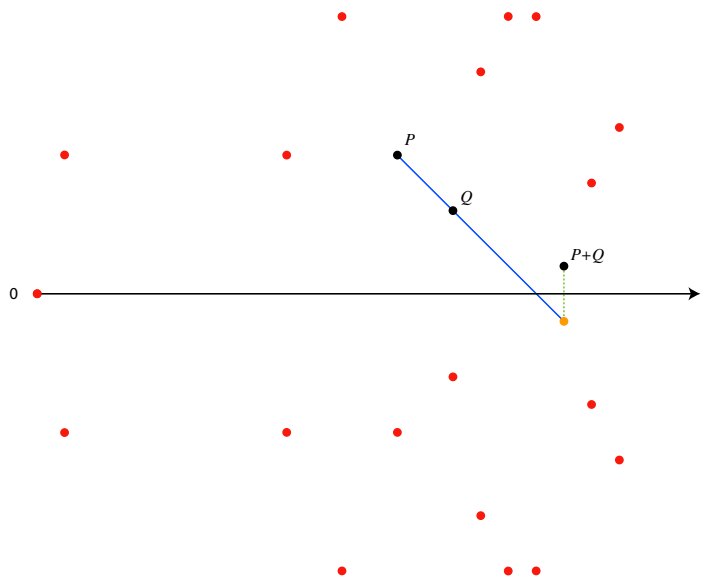


FIGURE 7.4 – Addition sur une courbe elliptique modulo q.

elliptique

$$y^2 \equiv x^3 + x \pmod{233},$$

sont :

$P = (11, 115)$,	$2\,P = (202, 148)$,	$3\,P = (86, 127)$,	$4\,P = (98, 160)$,
$5\,P = (78, 3)$,	$6\,P = (196, 156)$,	$7\,P = (42, 227)$,	$8\,P = (18, 228)$,
$9\,P = (146, 69)$,	$10\,P = (218, 173)$,	$11\,P = (111, 126)$,	$12\,P = (215, 21)$,
$13\,P = (20, 129)$,	$14\,P = (121, 154)$,	$15\,P = (137, 175)$,	$16\,P = (81, 7)$,
$17\,P = (143, 82)$,	$18\,P = (181, 44)$,	$19\,P = (176, 43)$,	...

La première étape du processus de codage consiste à transformer le texte clair m, en un point P sur la courbe elliptique, de telle sorte qu'on puisse récupérer le message m lorsqu'on connaît P. Bien qu'on ne sache pas encore trouver rapidement, systématiquement, et sans erreur, un grand nombre de points sur une courbe elliptique ; on connaît de bonnes méthodes probabilistes. En gros, elles consistent à piger x au hasard, puis à tester s'il est possible de trouver y tel que $y^2 - (x^3 + a\,x + b) \equiv 0 \pmod{q}$. Cela peut se faire très rapidement. Nous allons donc supposer qu'on dispose d'outils qui permettent de trouver une courbe elliptique, et un point sur cette courbe, pour q grand. Nous allons aussi voir plus loin qu'on peut très efficacement calculer un grand multiple du point sélectionné. Par contre le problème inverse semble très difficile. En effet, on a là une situation tout à fait analogue à celle du logarithme discret. Le problème consiste à chercher, pour des points P et Q donné, s'il existe, un entier

n tel que $n\,P = Q$. Pour l'instant, les spécialistes pensent que ce problème est plus difficile à résoudre que celui sur du logarithme discret. Cela suggère donc d'exploiter cette situation à des fins cryptographiques.

De façon très semblable au système de Elgamal, le système se met en place de la façon suivante. Une courbe elliptique est fixée pour l'ensemble des participants, ainsi qu'un point P choisit sur cette courbe. Chaque utilisateur X choisit aléatoirement un grand nombre entier n_X, qui constitue sa clé secrète, et il rend public le point $N_X := n_X\,P$, sa clé publique. Celle-ci peut être calculée très rapidement de la façon qui est expliquée ci-dessous. Ainsi, pour envoyer un message $m\,P$ à Laurie, Justine choisit aléatoirement un nombre k, et utilise la clé publique $N_L := n_L\,P$ de Laurie pour lui envoyer le couple (K, M), où

$$K := k\,P, \qquad \text{et} \qquad M := m\,P + k\,N_L.$$

Pour décoder le message, Laurie multiplie le premier nombre par sa clé secrète n_L et soustrait le résultat du second nombre. Elle obtient alors :

$$\begin{aligned} M - n_L\,K &= (m\,P + k \cdot (n_L\,P)) - n_L \cdot (k\,P) && (7)\\ &= m\,P && (8) \end{aligned}$$

Ainsi Justine envoie le message $m\,P$ masqué par $k \cdot N_L$. On ne peut enlever le masque que lorsqu'on connaît la clé secrète n_L.

Aspects techniques

Pour mettre en place un tel système, nous devons savoir trouver une courbe elliptique et un point sur cette courbe. On procède de façon probabiliste, en choisissant d'abord aléatoirement trois grands nombres x, y, et a modulo q. On pose alors

$$b := y^2 - (x^3 + ax).$$

On vérifie ensuite que

$$4a^3 + 27b^2 not \equiv 0 \pmod{q}.$$

Si cette condition n'est pas satisfaite, on recommence tout le processus. La théorie générale assure qu'un petit nombre d'essais conduit à une situation acceptable.

Nous devons maintenant montrer comment calculer rapidement un grand multiple $n\,P$ d'un point P. La démarche est très similaire à celle qui est utilisée pour calculer de grandes puissances. Une des approches possibles consiste à calculer d'abord, par doublement successif, les puissances 2^j de 2 qui sont plus petites que n :

$$P,\ 2\,P,\ 4\,P,\ \ldots\ , 2^k\,P,$$

avec

$$2^k < n < 2^{k+1}.$$

On peut alors obtenir facilement $n\,P$ par la succession d'additions

$$a_i\,P + 2^{j_i}\,P,$$

où $a_1 = 2^k$, et 2^{j_i} est la plus petite puissance de 2 telle que

$$2^{j_i} < n - a_i.$$

En d'autres termes, on calque les calculs sur le développement de n en base 2. Pour illustrer, considérons le calcul de $210\,P$. Les étapes de calcul sont les suivantes. On calcule d'abord

$$P, \quad 2\,P, \quad 4\,P \quad 8\,P, \quad 16\,P, \quad 32\,P, \quad 64\,P, \quad 128\,P,$$

puis on calcule :

$$192\,P = 128\,P + 64\,P, \qquad \text{et enfin} \qquad 210\,P = 192\,P + 18\,P.$$

7.3 Chaînes d'additions

Dans les calculs du paragraphe précédent, l'important est, qu'à chaque étape le point calculé s'obtienne soit par doublement d'un point déjà calculé, ou par addition de deux points déjà calculés. Ce type de calcul fait apparaître la notion de *chaîne d'additions*. Nous allons conclure notre discussion par un bref survol de cette notion, d'abord parce que l'élaboration d'algorithmes de calculs nécessite une bonne compréhension de ces chaînes, mais aussi parce que plusieurs problèmes restent à résoudre les concernant. Il serait dommage de ne pas profiter de cette occasion pour illustrer comment des problématiques simples nécessitent encore des recherches, même si la question a été soulevée il y a plusieurs décennies.

Techniquement, une *chaîne d'additions* est une suite d'entiers :

$$a_0,\ a_1,\ a_2,\ \ldots\ ,\ a_\ell,$$

avec $a_0 = 1$, et la propriété que chaque terme est la somme de deux termes (peut-être égaux) que le précède. Autrement dit,

$$a_i = a_j + a_k,$$

avec $j \leq k$, et tous deux plus petits que i. Si n est la valeur du dernier terme a_ℓ de la chaîne, on dit avoir une chaîne d'addition de *longueur* ℓ pour n. Par exemple, on a

$$1,\ 2,\ 4,\ 6,\ 12,\ 18,\ 22,\ 44,\ 88,\ 90$$

qui est une chaîne de longueur 9 pour 90. Une chaîne d'additions pour n permet de trouver les étapes d'un calcul du multiple $n\,P$ d'un point P, et sa longueur correspond au nombre de ces étapes. Ainsi, plus la longueur d'une chaîne est courte, plus nous aurons une façon efficace d'obtenir le multiple cherché. Or, la méthode dite *binaire* décrite à la section précédente n'est pas la plus efficace. Par exemple, on a la chaîne

$$1,\ 2,\ 4,\ 8,\ 9,\ 17,\ 34,\ 43,$$

pour 43, qui est clairement plus courte que la chaîne binaire :

$$1,\ 2,\ 4,\ 8,\ 16,\ 32,\ 40,\ 42,\ 43.$$

Dans certaines situations, la chaîne binaire est environ deux fois plus longue que nécessaire. Ainsi, la chaîne binaire pour l'entier

179769313486231590772930519078902473361797697894230657273430081157732
675805500963132708477322407536021120113879871393357658789768814416622
492847430639474124377767893424865485276302219601246094119453082952085
005768838150682342462881473913110540827237163350510684586298239947245
938479716304835356329624224137215

est de longueur 2046, mais on connaît une chaîne de longueur 1033.

Bien que ce problème ait été étudié depuis le début du XX$^{\text{e}}$-siècle, on ne sait toujours pas trouver efficacement une plus courte chaîne d'additions pour un entier n. La question est toujours activement étudiée de nos jours. On peut trouver un survol intéressant des travaux la concernant, à la section 4.6.3 du second volume du fameux livre de Donald Knuth : *The Art of Computer Programming*.

Bibliographie

[1] H. Anton, *Elementary linear algebra, Application version*, 7e éd., John Wiley and sons,1994.

[2] H. Beker, F. Piper, *Cipher systems : the protection of communications*, Wiley-Interscience, 1982.

[3] J.A. Buchmann, *Introduction to cryptography*, Springer, 2001.

[4] H. Delfs, H. Knebl, *Introduction to Cryptography*, Springer, 2002.

[5] G. Dubertret, *Initiation à la cryptographie*, Vuibert, 1998.

[6] S. W. Hawking, I. Naddeo-Souriau, *une brève histoire du temps, du Big-bang aux trous noirs*, J'ai lu éd., juillet 2000 .

[7] D. Kahn, *La Guerre des codes secrets*, (traduction de *The codebreakers*), Inter éditions, 1980.

[8] N. Koblitz, *A course in Number theory and Cryptography*, Springer, 1994.

[9] A. G. Konheim, *Cryptography, a Primer*, New York, Wiley-interscience, 1981.

[10] A.J. Menezes, P.C. Van OOrschot, S.A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 2001.
http ://www.cacr.math.uwaterloo.ca/hac/.

[11] K.H. Rosen, *Mathématiques discrètes*, Chenelière McGraw-Hill, 2001.

[12] A. Sinkov, *Elementary Cryptanalysis*, a mathematical approach, MAA math. library, 1996.

[13] B. Schneier, *Cryptographie appliquée*, 2e édition, John Wiley and sons, 1996.

[14] D. Stinson, *Cryptographie théorie et pratique*, 2e édition, Vuibert, 2003.

[15] A.M. Yaglom, I.M. Yaglom, *Probabilité et information, théorie et application*, 2e édition, Dunod, Paris, 1969.

[16] G. Zémor, *Cours de Cryptographie*, Cassini, 2000.

Articles

[17] W. Diffie, M.E. Hellman, *New directions in cryptography*, IEEE Transactions on Information theory, 22, 1976, p. 644-654.

[18] L. S. Hill, *Cryptography in an algebraic alphabet*, american Mathematical monthly, 36, 1929.

[19] A. Kerckhoffs, *La cryptographie militaire*, Journal des sciences militaires, vol. IX, pp. 5–38, Janvier 1883, pp. 161–191, Février 1883.

[20] R. E. Lewand, *Cryptological Mathematics*, the mathematical association of america, 2000, p. 124-140.

[21] J. Silverman, *The arithmetic of Elliptic curves*, Springer-Verlag, 1986.

[22] C. Shannon, *A mathematical theory of communication*, Bell telephone systems technical publication, 1948.

Sites web

[23] Ars Cryptographica, Didier Müller, Lycée cantonal de Porrentruy,
http ://www.apprendre-en-ligne.net/crypto/menu/index.html

[24] La cryptographie expliquée, Frédéric Bayart,
http ://www.bibmath.net/crypto/index.php3

[25] Le site du cours : La cryptographie de l'Antiquité à l'Internet,
http ://www.lacim.uqam.ca/crypto/index.html

[26] La folle course informatique 2000,
http ://www.gel.usherb.ca/fci/fci_2k3/français/éditions_precedentes.html

[27] L'encyclopédie en ligne Wikipedia,
http ://en.wikipedia.org/wiki/Topics_in_cryptography

[28] Le site d'Adriano Garsia,
http ://math.ucsd.edu/ garsia/