Exercices et problèmes de cryptographie

Damien Vergnaud

Préface de Jacques Stern

Exercices et problèmes de cryptographie

2e édition

DUNOD

Illustration de couverture :
Energy of fractal realms © agsandrew-Fotolia.com

5 rue Laromiguière, 75005 Paris
www.dunod.com

ISBN 978-2-10-072145-0

PRÉFACE

« *Pour devenir habile en quelque profession que ce soit, il faut le concours de la nature, de l'étude et de l'exercice* ». Cette maxime d'Aristote semble bien mal s'appliquer à la cryptologie tant l'exercice y est absent. Il existe de multiples ouvrages de référence de qualité mais, pour la plupart, ils sollicitent très peu l'initiative des étudiants. Et même ceux – rares – qui sont accompagnés d'un véritable choix de problèmes à résoudre, par exemple sous forme d'un livre compagnon, ne couvrent pas totalement une discipline qui connaît une évolution rapide. C'est donc un réel manque que vient combler le recueil que propose Damien Vergnaud.

Le livre que j'ai le plaisir de présenter est issu d'un vrai travail de terrain puisqu'il est le résultat de plusieurs années d'enseignement de la cryptologie à l'École normale supérieure. À l'évidence, l'auteur a beaucoup de talent pour éveiller l'intérêt des étudiants et les conduire, pas à pas, à s'approprier les concepts et les méthodes de la science du secret. Beaucoup de culture également, puisque les sujets choisis sont extrêmement variés à l'image d'une science qui emprunte à l'algèbre, à la théorie des probabilités, à l'algorithmique, à la théorie de l'information. D'ailleurs, ils débordent largement le cadre strict de la cryptographie. Ce talent et cette culture conduisent à un choix d'exercices qui ne demandent pas simplement à l'étudiant de faire des gammes mais lui proposent de s'attaquer à de véritables compositions : ici un effort raisonnable de programmation illustre des cryptanalyses célèbres comme celle de l'Enigma ou celle du programme Venona qui a permis l'interception de communications où les services russes mettaient incorrectement en œuvre le chiffrement jetable ; là une invitation à « mettre la main à la pâte » permet d'entrer de plain-pied dans les méthodes modernes de cryptanalyse – différentielle et linéaire – des algorithmes conventionnels tels que le DES ou l'AES ; là encore, une initiation progressive aux méthodes de factorisation d'entiers, intimement liées à la sécurité du RSA est proposée.

Présenter un tel ouvrage comme un simple livre d'exercices est le reflet de la modestie de son auteur. Certes, il permet la pratique nécessaire à l'acquisition des éléments essentiels de la cryptologie. Mais il va au-delà de cet objectif : chaque chapitre inclut une présentation qui est un véritable cours d'introduction et l'ensemble constitue de fait une forme d'ouvrage d'enseignement avancé fondé sur la pratique. En d'autres termes, le lecteur qui va au terme de tous les exercices proposés est déjà un

véritable spécialiste, capable de se confronter aux multiples concepts que la cryptologie moderne a développés ces trente dernières années. À un moment où la cryptologie est au cœur de la société de l'information, de l'internet aux moyens de paiement en passant par les téléphones portables, une telle expertise est indispensable et il faut souhaiter au livre de Damien Vergnaud des lecteurs à la fois nombreux et actifs.

Jacques Stern
Professeur à l'École normale supérieure

TABLE DES MATIÈRES

AVANT-PROPOS

La cryptologie est un ensemble de techniques permettant d'assurer la sécurité des systèmes d'information. Cette discipline permet notamment de conserver aux données leur caractère de confidentialité, de contrôler leur accès ou d'authentifier des documents. L'utilisation de la cryptographie est de plus en plus répandue et les utilisateurs des systèmes cryptographiques doivent être en mesure non seulement de comprendre leur fonctionnement mais aussi d'en estimer la sécurité.

Cet ouvrage s'adresse aux étudiants de second cycle d'informatique ou de mathématiques. Il s'est développé à partir de textes de travaux dirigés et de travaux pratiques proposés à des étudiants du Master parisien de recherche en informatique (MPRI) et aux élèves de première année de l'École normale supérieure. Il a été conçu pour aider à assimiler les connaissances d'un cours d'introduction à la cryptologie et à se préparer aux examens. Il présente les outils mathématiques et algorithmiques utiles en cryptographie et les fonctionnalités cryptographiques de base dans le cadre de la cryptographie symétrique et asymétrique.

Les exercices destinés aux étudiants de « master 1 » pourront cependant être abordés par un étudiant motivé de licence ayant un goût pour l'algorithmique dans ses aspects mathématiques et pratiques. À l'intention des étudiants plus avancés, nous avons inclus des énoncés plus difficiles qui sont alors signalés par un astérisque. Enfin, l'ouvrage sera utile aux enseignants de cryptologie qui y trouveront un support pour leurs travaux dirigés.

La cryptologie est liée à d'autres disciplines mathématiques et informatiques comme l'arithmétique, l'algèbre, l'algorithmique, ou la théorie de la complexité. Le bagage informatique et mathématique requis pour aborder ce livre est celui que l'on acquiert lors des deux premières années de licence ou en classes préparatoires scientifiques augmenté de quelques notions de théorie des nombres de niveau troisième année. Ces notions plus avancées font l'objet de brefs rappels qui n'ont cependant pas pour ambition de remplacer un livre de cours.

Le but de cet ouvrage est de permettre à ceux qui le souhaitent de s'initier à la cryptographie par l'exemple. Il propose plus d'une centaine d'exercices et problèmes entièrement utilisés dans le cadre de travaux dirigés, de travaux pratiques ou d'examens. Ces exercices sont entièrement corrigés mais le lecteur ne tirera profit de ce livre que s'il cherche des solutions personnelles avant d'en étudier les corrections. L'étude de la cryptologie moderne ne peut se concevoir sans un ordinateur à portée

de main et le livre propose de nombreux exercices de programmation qui ont pour but notamment d'acquérir une pratique de la cryptanalyse. Les données numériques de ces exercices sont disponibles en ligne sur :

www.dunod.com/contenus-complementaires/9782100721450

Le lecteur pourra recopier les énoncés des exercices avant de les traiter. Il trouvera également une vingtaine d'exercices supplémentaires et des références complémentaires.

Cette **deuxième édition** a été inspirée par les nombreuses demandes et remarques que m'ont envoyées des étudiants et collègues, utilisateurs de la première édition. Cette nouvelle édition m'a donné l'occasion de modifier et réécrire des parties importantes du texte en suivant ces remarques sur le contenu, le style et l'organisation de l'ouvrage. En plus d'épurer le texte de ses inévitables erreurs typographiques et coquilles, j'ai simplifié et clarifié une grande partie des énoncés des exercices et de leurs solutions. J'ai notamment ajouté des questions intermédiaires pour simplifier la résolution de certains exercices et détaillé certains points techniques dans des solutions d'exercices complexes. J'ai supprimé certains exercices jugés trop difficiles et j'en ai également ajouté de nouveaux. Enfin, les compléments en ligne qui accompagnent l'ouvrage ont été enrichis d'une vingtaine d'exercices supplémentaires (avec leurs solutions complètes) et d'autres exercices et compléments de cours seront ajoutés progressivement.

Remerciements

J'adresse un chaleureux merci à David Naccache et Jacques Stern avec qui j'ai eu le plaisir d'enseigner le cours d'*Introduction à la cryptologie*. Ma gratitude va également aux étudiants du MPRI et aux élèves normaliens de ces dernières années qui ont testé, malgré eux, la majorité des exercices présentés dans ce livre. Ce texte doit beaucoup à des conversations de couloirs et je tiens également à remercier les doctorants, post-doctorants et membres permanents de l'équipe Cryptographie de l'ENS – et particulièrement Pierre-Alain Fouque – pour toutes les discussions que nous avons pu avoir. Pour terminer, je voudrais remercier Aurélie Bauer, Guilhem Castagnos, Céline Chevalier, Aurore Guillevic, Pierre-Alain Fouque, Fabien Laguillaumie, Roch Lescuyer et Juliette Vergnaud-Gauduchon pour la rigueur et la pertinence de leurs nombreux commentaires.

Notations

Les conventions et notations suivantes sont utilisées dans cet ouvrage :

a) Ensembles

Nous utilisons les notations ensemblistes classiques : $\emptyset$ désigne l'ensemble vide ; $x \in A$ signifie que x est un élément de l'ensemble A. Pour deux ensembles A et B, $A \subseteq B$ indique que A est un sous-ensemble de B (alors que $A \subset B$ indique que A est un sous-ensemble strict de B). De plus, $A \cup B$ désigne la réunion de A et B, $A \cap B$ désigne l'intersection de A et B, $A \setminus B$ l'ensemble des éléments de A qui ne sont pas dans B et $A \times B$ le produit cartésien des ensembles A et B. Le cardinal d'un ensemble A est noté $\#A$. Nous utilisons les notations classiques suivantes pour désigner certains ensembles :

$\mathbb{N}$	ensemble des entiers naturels
$\mathbb{P}$	ensemble des nombres premiers
$\mathbb{Z}$	anneau des entiers relatifs
$\mathbb{Q}$	corps des nombres rationnels
$\mathbb{R}$	corps des nombres réels
$\mathbb{C}$	corps des nombres complexes
$(\mathbb{Z}/N\mathbb{Z})$	anneau des résidus modulo un entier $N \geqslant 1$
$\mathbb{F}_q$	corps fini à q éléments
$\mathfrak{S}_A$	groupe de permutations de l'ensemble A
A^*	groupe des éléments inversibles d'un anneau A
$\mathscr{M}_\ell(A)$	anneau des matrices carrées $\ell \times \ell$ à coefficients dans un anneau A
$A[X]$	anneau des polynômes à une indéterminée X à coefficients dans un anneau A

La lettre p désigne le plus souvent un nombre premier $p \in \mathbb{P}$ et nous notons $(p_n)_{n\geqslant 1}$ la suite croissante des nombres premiers (avec $p_1 = 2$, $p_2 = 3, \ldots$). Pour un polynôme $P \in A[X]$, nous notons $\deg P$ le degré de P. La notation $\mathbb{G}$ désigne un groupe dont la loi est notée multiplicativement. L'élément neutre pour la multiplication dans $\mathbb{G}$ est noté $1_{\mathbb{G}}$. L'ordre d'un groupe $\mathbb{G}$ est noté $|\mathbb{G}| = \#\mathbb{G}$ et $\langle g \rangle$ désigne le sous-groupe de $\mathbb{G}$ engendré par $g \in \mathbb{G}$.

b) Fonctions

Nous notons $f : A \longrightarrow B$ pour indiquer que f est une fonction d'un ensemble A dans un ensemble B. Pour un sous-ensemble $A' \subseteq A$, nous notons $f(A') = \{f(a), a \in A'\} \subseteq$

B. Pour un sous-ensemble $B' \subseteq B$, nous notons $f^{-1}(B') = \{a \in A, f(a) \in B'\} \subseteq A$. La composition de fonctions est notée $\circ$. Nous utilisons les notations classiques suivantes pour désigner certaines fonctions :

$\lfloor x \rfloor$	partie entière par défaut de $x \in \mathbb{R}$ ($\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$)
$\lceil x \rceil$	partie entière par excès de $x \in \mathbb{R}$ ($\lceil x \rceil - 1 < x \leq \lceil x \rceil$)
$\ln(x)$	logarithme népérien de $x \in \mathbb{R}$ ($x > 0$)
$\log(x)$	logarithme en base 2 de $x \in \mathbb{R}$ ($x > 0$)
$\log_g(h)$	logarithme discret de $h \in \langle g \rangle$ en base $g \in \mathbb{G}$
$\pi(x)$	nombre de nombres premiers inférieurs ou égaux à x ($\#\{p \in \mathbb{P}, p \leqslant x\}$)
$\Psi(x, y)$	fonction de Dickman-De Bruijn ($\#\{n \in \mathbb{N}, n \leqslant x$ et n est y-friable$\}$)
$\binom{n}{m}$	coefficient binomial $\binom{n}{p} = \frac{n!}{m!(n-m)!}$ pour $0 \leqslant m \leqslant n$
$\left(\frac{x}{m}\right)$	symbole de Jacobi
$\text{pgcd}(a, b)$	plus grand commun diviseur de $a, b \in \mathbb{Z}$
$\text{ppcm}(a, b)$	plus petit commun multiple de $a, b \in \mathbb{Z}$
$\Pr(E)$	probabilité d'un événement E

c) Chaînes binaires

Nous utilisons les notations classiques suivantes sur les chaînes binaires :

$\{0, 1\}^n$	ensemble des chaînes binaires de longueur n
$\{0, 1\}^*$	ensemble des chaînes binaires de longueur finie
$\wedge$	ET logique (bit à bit pour deux chaînes de même longueur)
$\vee$	OU logique (bit à bit pour deux chaînes de même longueur)
$\neg$	NON logique (bit à bit pour deux chaînes de même longueur)
$\oplus$	« OU exclusif » (bit à bit pour deux chaînes de même longueur)
$\lvert x \rvert$	longueur binaire d'une chaîne $x \in \{0, 1\}^*$
$\overline{x}$	chaîne binaire complémentaire de x ($\overline{x} = \neg x = x \oplus 1^n$ avec $n = \lvert x \rvert$)
$x \Vert y$	concaténation des chaînes x et y
x^n	concaténation de la chaîne x n fois ($\underbrace{x\Vert \ldots \Vert x}_{n\text{fois}}$)
$\lll i$	rotation à gauche d'une chaîne de bits de i positions

Dans les chapitres 2, 3 et 4, nous utilisons une fonte de type « machine à écrire » pour représenter la valeur d'un octet avec deux chiffres hexadécimaux : `00` = 0, `01` = 1, …, `0A` = 10, …, `10` = 16, …, `FF` = 255.

d) Notations algorithmiques

Les algorithmes sont présentés sous forme de pseudo-code simple (notamment en s'affranchissant des problèmes de mémoire). Les entrées et les sorties sont toujours précisées. Les structures de contrôle classiques sont notées en gras (**tant que** condition **faire** instructions **fin tant que**, **si** condition **alors** instructions **sinon** instructions **fin si**, …). Les commentaires dans les algorithmes sont signalés par le symbole ▷. Le symbole $a \leftarrow b$ indique l'assignation algorithmique (*i.e.* a prend la valeur de b) et le symbole $a \xleftarrow{u.a.} A$ l'assignation d'un élément tiré uniformément aléatoirement (*i.e.* un élément est tiré uniformément aléatoirement dans l'ensemble A et la valeur obtenue est enregistrée dans a).

CRYPTOGRAPHIE CLASSIQUE

1

La cryptologie est une science très ancienne : les hommes ont toujours eu besoin de dissimuler des informations et de transmettre des messages en toute confidentialité. Le terme cryptologie vient du grec *kruptos* signifiant *secret, caché* et de *logos* signifiant *discours*. La cryptologie est donc la *science du secret*. Elle regroupe la cryptographie et la cryptanalyse : la première a pour but de concevoir des systèmes visant à assurer la sécurité des communications sur un canal public et la seconde vise à trouver des failles dans ces systèmes.

La cryptographie est traditionnellement utilisée pour dissimuler des messages aux yeux de certains utilisateurs et le chiffrement des communications militaires a depuis l'Antiquité été une préoccupation majeure des diverses forces armées. Le *chiffrement* regroupe les techniques mises en œuvre pour brouiller la signification d'un message qui est matériellement visible. Le contenu du message ne doit alors être retrouvé que par les personnes auxquelles le message est adressé. Le chiffrement fait appel à deux processus élémentaires impliquant la transformation des lettres d'un message pour satisfaire ces propriétés : la *substitution* qui consiste à remplacer, sans en bouleverser l'ordre, les symboles d'un texte clair par d'autres symboles et la *transposition* qui repose sur le bouleversement de l'ordre des symboles (mais pas leur identité).

Dans ce chapitre, nous allons étudier des systèmes de chiffrement relativement simples qui ont été utilisés de l'Antiquité (*chiffrement de César* ou *scytale*) jusqu'au début du XX^{e} siècle (*chiffrement de Vernam*, *chiffrement de Hill*, *machine Enigma*).

1.1 CHIFFREMENT PAR SUBSTITUTION MONO-ALPHABÉTIQUE

Le *chiffrement par substitution* consiste à remplacer dans un message un ou plusieurs symboles par un ou plusieurs symboles (généralement du même alphabet) tout en conservant l'ordre de succession des symboles du message. Dans cette section, nous considérons le chiffrement par substitution *mono-alphabétique* qui consiste à remplacer chaque symbole individuel du message par un autre symbole de l'alphabet. Nous allons étudier les techniques de cryptanalyse permettant d'attaquer un tel système.

Elles reposent sur l'analyse des fréquences des symboles utilisés dans le texte chiffré et utilisent le fait que, dans chaque langue, certains symboles ou combinaisons de symboles apparaissent plus fréquemment que d'autres. Les systèmes de chiffrement par substitution mono-alphabétique conservent la répartition des fréquences et

si le message chiffré est suffisamment long, la recherche d'un symbole ayant une fréquence élevée permettra parfois de retrouver tout ou partie du message clair associé.

La fréquence d'apparition des lettres varie bien évidemment en fonction de la langue et du type de texte considérés. Pour un texte rédigé en français, nous obtenons généralement les fréquences d'apparition (en pourcentage) proches des valeurs suivantes[1] (*cf.* Figure 1.1) :

a	b	c	d	e	f	g	h	i	j	k	l	m
8,46	1,02	3,21	3,78	17,60	1,11	1,12	1,07	7,40	0,48	0	6,05	2,70
n	o	p	q	r	s	t	u	v	w	x	y	z
6,38	5,19	2,68	1,21	6,56	7,56	7,26	6,63	1,65	0	0,03	0,03	0,01

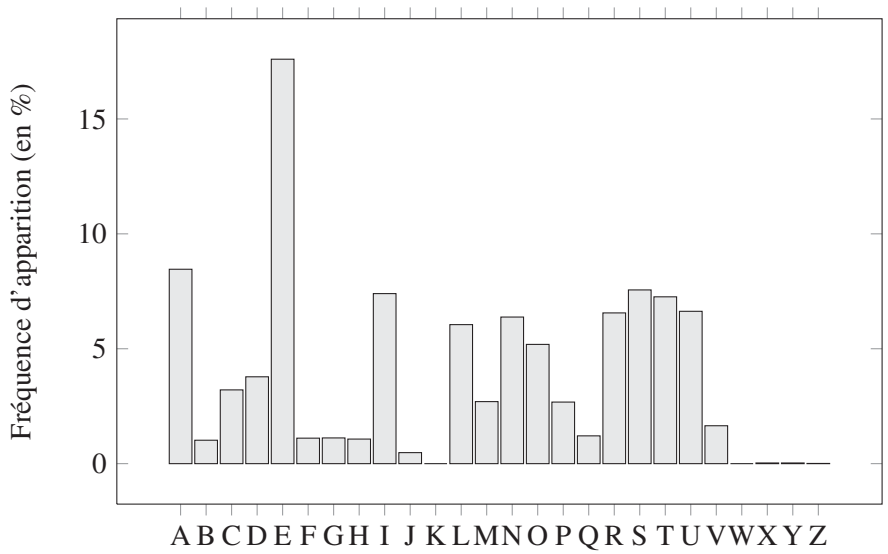

Figure 1.1– Fréquence d'apparition des lettres en français

De même, certains couples de lettres (ou *bigrammes*) apparaissent plus souvent que d'autres dans une langue donnée. Les vingt bigrammes les plus fréquents de la langue française sont (du plus fréquent au moins fréquent) : *es*, *de*, *le*, *en*, *re*, *nt*, *on*, *er*, *te*, *el*, *an*, *se*, *et*, *la*, *ai*, *it*, *me*, *ou*, *em* et *ie*.

Le chiffrement par substitution mono-alphabétique le plus simple est le *chiffrement par décalage*, aussi connu sous le nom de *chiffrement de César*. Il consiste simplement à décaler les lettres de l'alphabet d'un nombre de positions constant vers la

1. Ces valeurs correspondent aux fréquences d'apparition des lettres dans le roman *Notre-Dame de Paris* de V. Hugo (en identifiant les lettres accentuées et les lettres non accentuées).

droite ou la gauche. Par exemple, en décalant les lettres de trois rangs vers la gauche (comme le faisait J. César), le texte clair *veni vidi vici* devient *yhql ylgl ylfl*.

Exercice 1.1 (avec programmation). *Chiffrement de César*

Décrypter le texte suivant qui a été obtenu en appliquant le chiffrement de César sur un texte en langue française dans lequel les espaces ont été supprimées :

```
vcfgrwqwfsbhfsntowbsobgfsbhfsnqvsnjcigsghqsoixcif
rviwtshseicwbsgojsnjcigdogeisjcigoihfsgofhwgobgjc
igbsrsjsnqwfqizsfrobgzsgfisgzsgxcifgcijfopzsgeioj
sqzsggwubsgrsjchfsdfctsggwcbdofzseiszsghhcbashwsf
```

Solution

Le chiffrement de César est un mode de chiffrement par substitution, il ne modifie donc pas la fréquence d'apparition des lettres. La lettre la plus fréquente dans un texte français étant le « e », le décalage entre la lettre la plus fréquente dans ce texte chiffré et la lettre « e » doit donc nous révéler la clé utilisée pour le chiffrement. La lettre qui apparaît le plus souvent dans le texte chiffré est le « s » avec 33 occurrences (puis vient la lettre « g » avec seulement 24 occurrences). Le décalage utilisé est donc vraisemblablement de 14 lettres vers la gauche et l'on obtient le message clair suivant :

```
horsdicirentrezfaineansrentrezchezvousestceaujour
dhuifetequoinesavezvouspasquevousautresartisansvo
usnedevezcirculerdanslesrueslesjoursouvrablesquav
eclessignesdevotreprofessionparlequelesttonmetier
```

En ajoutant les espaces et la ponctuation, nous reconnaissons la première réplique de la pièce *Jules César* écrit par W. Shakespeare en 1599 (dans la traduction de M. Guizot) :

« *Hors d'ici, rentrez, fainéans ; rentrez chez vous. Est-ce aujourd'hui fête ? Quoi ! Ne savez-vous pas que vous autres artisans vous ne devez circuler dans les rues les jours ouvrables qu'avec les signes de votre profession ? Parle, quel est ton métier ?* »

Le chiffrement affine est un système de chiffrement par substitution mono-alphabétique. La clé consiste en un couple d'entiers $(a, b) \in (\mathbb{Z}/26\mathbb{Z})^* \times (\mathbb{Z}/26\mathbb{Z})$. En remplaçant chaque lettre de l'alphabet par son rang (la lettre « a » est remplacée par 0, la lettre « b » est remplacée par 1, … et la lettre « z » est remplacée par 25), une lettre du texte clair de rang $i \in \{0, \ldots, 25\}$ est remplacée dans le chiffré par la lettre de rang $a \cdot i + b \bmod 26$. Puisque a est inversible dans $(\mathbb{Z}/26\mathbb{Z})$, cette transformation est bien une permutation de $(\mathbb{Z}/26\mathbb{Z})$.

Exercice 1.2 (avec programmation). *Chiffrement affine*

Décrypter le texte suivant qui a été obtenu en appliquant le chiffrement affine sur un texte en langue française dans lequel les espaces ont été supprimées :

```
ntjmpumgxpqtstgqpgtxpnchumtputgfsftgthnngxnchumwx
ootrtumhpyctgktjqtjchfooxujqhgztumxpotjxotfoqtohr
 xumhzutwftgtopfmntjmpuatmfmshodpfrxpjjtqtghbxuj
```

Solution

Comme le chiffrement de César, le chiffrement affine ne modifie pas la fréquence d'apparition des lettres. La lettre la plus fréquente dans un texte français est le « e » et les lettres suivantes sont par ordre de fréquence le « a », le « i », le « n », le « s » et le « t » (avec des fréquences très variables d'un texte à l'autre). La lettre qui apparaît le plus souvent dans le texte chiffré est le « t » avec 19 occurrences puis vient la lettre « m » avec 12 occurrences.

En supposons que la lettre « t » correspond à la lettre « e » et que la lettre « m » correspond à l'une des lettres « a », « i », « n », « s » ou « t », nous devons résoudre pour chaque choix un système linéaire à deux équations et deux inconnues dans $(\mathbb{Z}/26\mathbb{Z})$ de la forme

$$\begin{cases} a \cdot 19 + b = 4 \\ a \cdot 12 + b = \ell \end{cases}$$

où ℓ est l'entier 0, 8, 13, 18 ou 19 selon que la lettre testée pour « m » est « a », « i », « n », « s » ou « t » (respectivement). Le couple (a, b) obtenu est testé en déchiffrant les premiers caractères du chiffré et nous obtenons les résultats suivants :

Lettre testée	« a »	« i »	« n »	« s »	« t »
ℓ	0	8	13	18	19
(a, b)	$(8, 8)$	$(18, 0)$	$(21, 21)$	$(24, 16)$	$(9, 15)$
Début du « clair » associé	`iecaym`	`aegikw`	`iecnyz`	`qeysmc`	`cestun`

La clé à utiliser pour le déchiffrement est donc vraisemblablement le couple $(9, 15)$ et nous obtenons le message clair suivant :

```
cestuntroudeverdureouchanteunerivieреaccrochantfol
lementauxherbesdeshaillonsdargentoulesoleildelamon
  tagnefiereluitcestunpetitvalquimoussederayons
```

Il s'agit bien sûr du premier quatrain du sonnet *Le Dormeur du val* écrit par A. Rimbaud en 1870.

Les systèmes de chiffrement par substitution mono-alphabétique les plus généraux utilisent une permutation aléatoire des symboles de l'alphabet utilisé. Pour l'alphabet latin formé de 26 lettres, le nombre de clés possibles est donc égal à $26! \simeq 4 \cdot 10^{26} \simeq 2^{88}$. Cependant même si le nombre de clés rend toute recherche

exhaustive impossible, les techniques d'analyse fréquentielle permettent de décrypter facilement un chiffré suffisamment long.

Exercice 1.3 (avec programmation).
Chiffrement par substitution mono-alphabétique

Le texte suivant résulte du chiffrement d'un texte français par une substitution mono-alphabétique.

```
v ubcfb osu ymoqsuu n cxqfj dqmfnu ub vjcfqu juz amqjmruz zmsscfusb bquflu
auoquz hfszbms zwfba ju wusbms qusbqu ncsz ju vmo z uddmqvcfb n uxfbuq ju
xusb wcoxcfz fj eczzc qcefnuwusb jc emqbu xfbquu no ijmv nuz wcfzmsz nu jc
xfvbmfqu ecz czzul qcefnuwusb vueusncsb emoq uweuvauq kou z usrmoddqu us
wuwu buwez kou jof os bmoqifjjms nu emozzfuqu ub nu zciju
ju acjj zusbcfb ju vamo vofb ub ju xfuog bcefz c j osu nu zuz ugbquwfbuz
osu cddfvau nu vmojuoq bqme xczbu emoq vu nuejmfuwusb fsbuqfuoq ubcfb
vjmouu co woq ujju quequzusbcfb zfwejuwusb os usmqwu xfzcru jcqru nu ejoz
n os wubqu ju xfzcru n os amwwu n usxfqms kocqcsbu vfsk csz c j uecfzzu
wmozbcvau smfqu cog bqcfbz cvvusbouz ub iucog

hfszbms zu nfqfruc xuqz j uzvcjfuq fj ubcfb fsobfju n uzzcpuq nu
equsnqu j czvuszuoq wuwu cox wufjjuoquz uemkouz fj dmsvbfmsscfb qcquwusb
cvboujjuwusb n cfjjuoqz ju vmoqcsb ujuvbqfkou ubcfb vmoeu ncsz jc ymoqsuu
v ubcfb osu nuz wuzoquz n uvmsmwfu eqfzuz us xou nu jc zuwcfsu nu jc acfsu
zms ceecqbuwusb ubcfb co zuebfuwu hfszbms kof cxcfb bqusbu suod csz ub
zmoddqcfb n os ojvuqu xcqfkouog co nuzzoz nu jc vauxfjju nqmfbu wmsbcfb
jusbuwusb fj z cqqubc ejozfuoqz dmfz us vauwfs emoq zu quemzuq c vackou
ecjfuq zoq osu cddfvau vmjjuu co woq dcvu c jc vcru nu j czvuszuoq j
usmqwu xfzcru xmoz dfgcfb no qurcqn v ubcfb os nu vuz emqbqcfbz cqqcsruz
nu bujju zmqbu kou juz puog zuwijusb zofxqu vujof kof eczzu osu jurusnu
zmoz ju emqbqcfb nfzcfb ifr iqmbauq xmoz qurcqnu

c j fsbuqfuoq nu j ceecqbuwusb nu hfszbms osu xmfg zovquu dcfzcfb usbusnqu
osu zuqfu nu smwiquz kof cxcfusb bqcfb c jc eqmnovbfms nu jc dmsbu jc xmfg
eqmxuscfb n osu ejckou nu wubcj mijmsrou wfqmfq buqsu usvczbqu ncsz ju woq
nu nqmfbu hfszbms bmoqsc os imobms ub jc xmfg nfwfsoc nu xmjowu wcfz juz
wmbz ubcfusb usvmqu nfzbfsvbz ju zms nu j ceecqufj no bujuvqcs vmwwu ms
nfzcfb emoxcfb ubqu czzmoqnf wcfz fj s p cxcfb covos wmpus nu j ubufsnqu
vmwejubuwusb hfszbms zu nfqfruc xuqz jc dusubqu fj ubcfb nu zbcboqu dquju
ejobmb eubfbu ub zc wcfrquoq ubcfb zmojfrsuu ecq jc vmwifscfzms ijuou
osfdmqwu no ecqbf fj cxcfb juz vauxuog bquz ijmsnz ju xfzcru scboqujjuwusb
zcsrofs jc euco noqvfu ecq ju zcxms rqmzzfuq juz jcwuz nu qczmfq uwmozzuuz
ub ju dqmfn nu j afxuq kof xuscfb nu equsnqu dfs
```

Décrypter ce texte.

Solution

Le nombre d'occurrences de chaque caractère du texte chiffré est donné dans le tableau suivant :

a	b	c	d	e	f	g	h	i	j	k	l	m
17	141	150	23	48	146	11	6	13	106	14	2	95
n	o	p	q	r	s	t	u	v	w	x	y	z
71	113	4	129	20	129	0	329	51	57	35	2	128

Il est vraisemblable que le caractère « u » représente le caractère « **e** » dans le texte clair (nous utiliserons une fonte grasse pour indiquer les lettres appartenant au texte clair). Les autres lettres les plus fréquentes sont le « c » et le « f » mais leurs fréquences sont trop proches pour décider quels caractères elles représentent.

Les bigrammes les plus fréquents de la langue française sont « es », « de », « le », « en », « re »« nt », et « on ». Les bigrammes commençant par « u » dans le texte chiffré ne sont pas assez fréquents pour décider quel caractère correspond à la lettre « **s** » dans le texte clair. Les bigrammes les plus fréquents du texte chiffré finissant par « u » sont « ju » (39 occurrences) et « nu » (35 occurrences) ce qui suggère que « j » représente « **l** » et « n » représente « **d** ». Ces trois substitutions donnent, pour le premier paragraphe, le texte suivant :

```
v ebcfb ose ymoqsee d cxqfl dqmfde eb vlcfqe lez amqlmrez zmsscfesb
bqefle aeoqez hfszbms zwfba le wesbms qesbqe dcsz le vmo z eddmqvcfb
d exfbeq le xesb wcoxcfz fl eczzc qcefdewesb lc emqbe xfbqee do ilmv
dez wcfzmsz de lc xfvbmfqe ecz czzel qcefdewesb veeesdcsb emoq eweevaeq
koe z esrmoddqe es wewe bewez koe lof os bmoqifllms de emozzfeqe eb de
zcile
```

Le mot « **do** » suggère que le caractère « o » représente « **u** ». Avec cette déduction, le mot « ose » devient « **use** » qui suggère que le caractère « s » représente « **n** ». Le premier paragraphe du texte devient alors :

```
v ebcfb une ymuqnee d cxqfl dqmfde eb vlcfqe lez amqlmrez zmnncfenb
bqefle aeuqez hfnzbmn zwfba le wenbmn qenbqe dcnz le vmu z eddmqvcfb
d exfbeq le xenb wcuxcfz fl eczzc qcefdewenb lc emqbe xfbqee du ilmv
dez wcfzmnz de lc xfvbmfqe ecz czzel qcefdewenb veeendcnb emuq eweevaeq
kue z enrmuddqe en wewe bewez kue luf un bmuqifllmn de emuzzfeqe eb de
zcile
```

Le mot « **kue** » qui apparaît deux fois suggère que la lettre « k » représente « **q** ». Les mots « qcefdewenb » et « eb » suggèrent que la lettre « b » représente « **t** ». Les mots « **luf** » et « fl » suggèrent que la lettre « f » représente « **i** » et nous obtenons :

```
v etcit une ymuqnee d cxqil dqmide et vlciqe lez amqlmrez zmnncient
tqeile aeuqez hinztmn zwita le wentmn qentqe dcnz le vmu z eddmqvcit
```