

Codage, cryptologie et applications

Bruno Martin

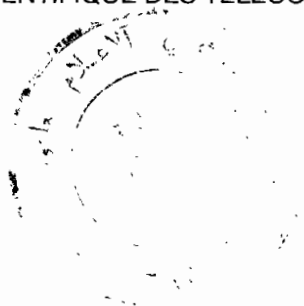


IST 2841

**Codage, cryptologie
et applications**

BIBLIOTHEQUE DU CERIST

COLLECTION TECHNIQUE ET SCIENTIFIQUE DES TÉLÉCOMMUNICATIONS



Codage, cryptologie et applications

Bruno Martin

PRESSES POLYTECHNIQUES ET UNIVERSITAIRES ROMANDES

BIBLIOTHEQUE DU CERIST

Dans la même collection:

VHDL

Langage, modélisation et synthèse

R. AIRIAU, J.-M. BERGE, V. OLIVE & J. ROUILLARD

Objets réactifs en Java

FRÉDÉRIC BOUSSINOT

Egalement disponibles aux Presses polytechniques et universitaires romandes

Programmation séquentielle avec ADA 95

PIERRE BREGUET, LUIGI ZAFFALON

Programmation concurrente et temps réel avec Ada 95

LUIGI ZAFFALON, PIERRE BREGUET

Algorithmes et structures de données avec Ada, C++ et Java

ABDELALI GUERID, PIERRE BREGUET, HENRI ROTHLSBERGER

Programmation orientée objets en C++

Une approche évolutive

MARYLÈNE MICHELOU

MÉDARD RIEDER

Illustration de couverture réalisée par Atelier Isatis, Dijon.



Le catalogue des publications des Presses polytechniques et universitaires romandes peut être obtenu par courrier à l'adresse suivante, EPFL – Centre Midi, CH-1015 Lausanne, par E-Mail à ppur@epfl.ch, par téléphone au (0)21 693 41 40, ou par fax au (0)21 693 40 27.

<http://www.ppur.org>

© 2004, Première édition

Presses polytechniques et universitaires romandes.

CH – 1015 Lausanne

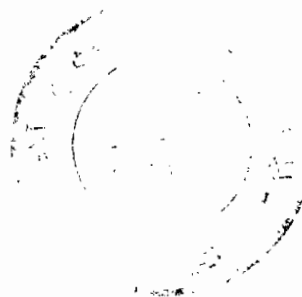
Tous droits réservés

ISBN 2-88074-569-1

Imprimé en Italie

Reproduction, même partielle, sous quelque forme ou sur quelque support que ce soit, interdite sans l'accord écrit de l'éditeur.

Préface



Au cours de nos premiers pas dans le domaine de la cryptologie, nous sommes tous confrontés à la subtile nuance entre “codage” et “chiffrement”. Il est vrai que l’usage courant de la langue française nous induit souvent en erreur. D’un point de vue formel, le “codage” se réfère à une représentation, quelle qu’elle soit, de l’information. D’un point de vue plus technique, on cherche à y représenter l’information de manière aussi efficace que possible pour communiquer, archiver, calculer. Le “chiffrement” peut être vu comme un cas particulier du codage : son objectif est également la représentation de l’information. Il est aussi la protection contre l’accès de personnes non habilitées et mal intentionnées. “Chiffrer” est par conséquent synonyme de “code secret”. Il est donc naturel en apparence d’englober la cryptographie dans les sciences des techniques de codage.

Ce serait pourtant ignorer leurs différences fondamentales. La cryptographie doit faire face à un adversaire malicieux. On cherche à savoir s’il parviendra sans effort à percer le système. La théorie de la complexité permet de quantifier la notion d’effort et joue donc un rôle central. En revanche, l’adversaire du codage (non cryptographique) est un phénomène aléatoire : un “bruit”. On cherche donc à coder ou à comprimer mieux sur un support donné. C’est ici la théorie de l’information qui permet de quantifier le bruit et la capacité à le surmonter. Codage et cryptographie font donc face à des adversaires différents au moyen de théories différentes.

Pourquoi alors traiter ces deux disciplines dans un même ouvrage ? Comme les chapitres suivants le démontrent, ce sont en fait les applications et leurs outils qui les rassemblent. Tous les systèmes de communication (téléphone, courrier électronique, Internet ...) et les systèmes de stockage de données (disque compact, DVD, carte à puce ...) utilisent le codage. La plupart utilisent aussi des moyens cryptographiques pour sécuriser les données ou protéger la propriété intellectuelle. Les outils fondamentaux proviennent de l’algèbre discrète et combinatoire et de l’algorithmique. On utilisera donc corps finis, polynômes et espaces vectoriels, mais aussi arbres, graphes et automates.

Codage et cryptographie sont aujourd'hui omniprésents. Il est donc important de maîtriser ces technologies. Peu de livres les traitent de manière équilibrée. Le présent ouvrage présente brillamment les deux techniques. Il occulte volontairement les aspects trop techniques et offre une présentation homogène sous une forme mathématique. Ce livre est donc recommandé aux étudiants en systèmes de communication et en informatique qui cherchent à comprendre leurs mécanismes de base. Nous leur souhaitons bonne lecture.

Serge Vaudenay
Professeur à l'EPFL

Préambule

Ce livre présente de façon accessible à tous deux disciplines : la *cryptologie* (ou science du chiffre) dont l'origine remonte aux débuts de l'histoire de l'homme et celle des *codes correcteurs* qui est née pendant la seconde moitié du 20^e siècle, après l'invention de la théorie de l'information. Toutes deux traitent de la transmission sûre de l'information. La cryptographie (ou l'art de concevoir de bons chiffres) assure la confidentialité des communications en présence d'ennemis. Les codes correcteurs permettent de retirer d'éventuelles erreurs de transmission.

Nous avons choisi d'orienter cet ouvrage plutôt vers les applications en télécommunications numériques que vers l'abstraction théorique. Nous avons volontairement simplifié le contenu mathématique de ces disciplines dans le but de mieux comprendre certains aspects du fonctionnement d'appareils que nous utilisons tous les jours comme les téléphones cellulaires, les disques compacts et les ordinateurs. Par exemple, sans codes correcteurs et sans cryptographie, le développement des réseaux comme Internet ou bientôt des réseaux dits «ad hoc» (il s'agit des réseaux sans fil qui se construisent à la demande et qui commencent à être utilisés par les ordinateurs portables, les téléphones cellulaires, les appareils photos numériques ou les imprimantes) ne pourrait avoir lieu.

Il est d'ailleurs intéressant de noter que les différentes notions que nous présentons sont à la croisée d'un grand nombre de disciplines comme l'informatique (tant théorique que pratique), les mathématiques, mais aussi l'économie et même le droit. L'économie à cause du développement quasi continu d'Internet et de son impact social et économique et le droit qui régit l'utilisation des systèmes de chiffrement. Ce n'est que depuis quelques années que l'utilisation de chiffres est autorisée en France. La taille des clés est toujours limitée, sauf pour l'utilisation de signatures électroniques.

Présentation

Cet ouvrage est découpé en cinq parties. La première présente la *théorie de l'information* de Weaver et Shannon dont le but était de quantifier le contenu en information. A ce titre, on lui doit la notion de bit (pour *binary digit*). Il s'agit de la théorie qui permet de relier sous un cadre commun la compression des données, la théorie des codes et la cryptologie ; ces différentes disciplines sont toutes prises en compte dans les problématiques de la théorie de l'information.

La deuxième partie traite de la *compression des données*. Celle-ci devient de plus en plus présente dans notre vie quotidienne. C'est grâce à la compression

des données que nous pouvons communiquer au moyen de téléphones cellulaires, regarder des DVD, transmettre des quantités d'informations toujours plus grandes au moyen d'un modem ou même d'Internet. Nous nous limitons ici aux seuls algorithmes de compression travaillant conjointement avec les mécanismes de chiffrement dans les logiciels qui permettent d'assurer la sécurité des communications au travers d'Internet.

Pour transmettre un message au travers d'un canal de communication qui peut être perturbé (ou bruité), on lui ajoute de la redondance pour qu'il soit possible de détecter, voire de corriger des erreurs de transmission. C'est le but de la *théorie des codes* qui fait l'objet de la troisième partie de ce livre. Pour présenter cette théorie qui peut devenir assez mathématique, nous avons privilégié le point de vue de l'ingénieur en informatique. Notre but est de faire comprendre les algorithmes de codage et de décodage nécessaires au bon fonctionnement du mécanisme de détection ou de correction des erreurs de différents media de communication. Nous y présentons le principe de fonctionnement du mécanisme de correction des erreurs des disques compacts, celui des téléphones cellulaires de type GSM et ceux qui régissent certaines communications par satellite. Nous nous sommes efforcés de présenter la théorie des codes de la manière la plus intuitive possible, sans trop entrer dans les détails mathématiques.

Dans la quatrième partie, nous rappelons brièvement quelques faits de *théorie de la complexité* initiée par Rabin, McNaughton, Yamada, Hartmanis et Stearns dans les années 1960. Cette théorie est constamment présente dans la problématique de la cryptographie. En effet, l'usage d'un chiffre est une course entre les utilisateurs légitimes du système de chiffrement et des adversaires qui cherchent à retrouver leurs messages. Il faut donc que le temps de calcul d'une cryptanalyse excède l'espérance de vie de la confidentialité du cryptogramme. La théorie de la complexité permet de ranger des problèmes algorithmiques dans différentes classes selon le temps de fonctionnement de leurs algorithmes de résolution. Nous y traitons la complexité de certains problèmes liés à la théorie des codes (notamment celui du décodage) et nous présentons quelques problèmes algorithmiques qui nous serviront dans la partie sur la cryptographie à clé publique, en particulier pour définir la notion de fonction à sens unique.

Enfin, nous introduisons la *cryptologie* ou science du chiffre qui permet de transmettre des messages en préservant leur confidentialité en présence d'adversaires. Cette discipline regroupe à la fois l'art de concevoir de bons chiffres (cryptographie) et celle de les cryptanalyser. Dans cette partie nous nous sommes efforcés de respecter cette interaction en proposant pour la plupart des chiffres présentés, soit leur cryptanalyse lorsque celle-ci est connue, soit des attaques qui présentent leurs principales faiblesses, tant pour les chiffres à clé secrète que pour les chiffres à clé publique. Le choix des méthodes de chiffrement repose d'une part sur leur intérêt pédagogique et sur leur intérêt «pratique» en extrayant un sous-ensemble consistant des méthodes adoptées par les différents standards. En effet, l'objectif principal de cette dernière partie est d'expliquer le fonctionnement de quelques protocoles sécurisés qui sont utilisés par Internet. Nous présentons notamment le fonctionnement d'un des

protocoles qui permettent d'assurer la sécurité des échanges du Web en chiffrant les communications. Son bon fonctionnement est généralement symbolisé par un petit cadenas verrouillé sur la fenêtre de nos navigateurs.

Utilisation

Le contenu de ce livre provient de différents enseignements dispensés à l'École Supérieure en Sciences Informatiques de l'Université de Nice Sophia-Antipolis et à l'Institut Supérieur d'Informatique et d'Automatique de l'École des Mines de Paris. Il s'inspire également de mon cours du DEA d'Informatique et de celui du DEA de Mathématiques, en collaboration avec S. Julia. C'est la raison pour laquelle il est divisé en plusieurs parties qui peuvent être lues de façon presque indépendante.

Chaque partie est découpée en plusieurs chapitres courts qui présentent une notion précise. De nombreux exemples et quelques exercices permettent d'illustrer les notions introduites et de vérifier leur bonne compréhension.

Cet ouvrage peut servir de support de cours pour des enseignements de deuxième ou de troisième cycle universitaire ou d'introduction pour des ingénieurs ou des chercheurs qui souhaiteraient se familiariser avec les différentes disciplines qui y sont présentées.

Remerciements

Tout d'abord, je dois beaucoup à la Maîtrise de Mathématiques Discrètes de l'Université de Lyon 1 et à ses enseignants-chercheurs. Ce sont eux qui m'ont fait connaître, entre autres, la théorie des codes et la cryptographie.

Je voudrais remercier J. Gruska pour m'avoir suggéré d'écrire ce livre à partir de mes notes de cours, C. Peyrat pour m'en avoir relu des versions préliminaires et H. Ben Azza pour ses encouragements.

Plusieurs personnes ont contribué à cet ouvrage : J. Leroux m'a fourni des explications sur la théorie de l'information ; P. Solé et Ling San ont répondu aux nombreuses questions que je leur ai posées sur les codes correcteurs, J. Bond, M. Bronstein et M. Santha ont éclairci des points d'algorithmique et de complexité. Je remercie également mes étudiants ; c'est grâce à leurs questions et à leurs réactions que le contenu de mes notes de cours s'est peu à peu enrichi.

Enfin, un grand merci à S. Julia. Son soutien, ses relectures et nos discussions m'ont permis de mener à bien l'écriture de ce livre.

BIBLIOTHEQUE DU CERIST

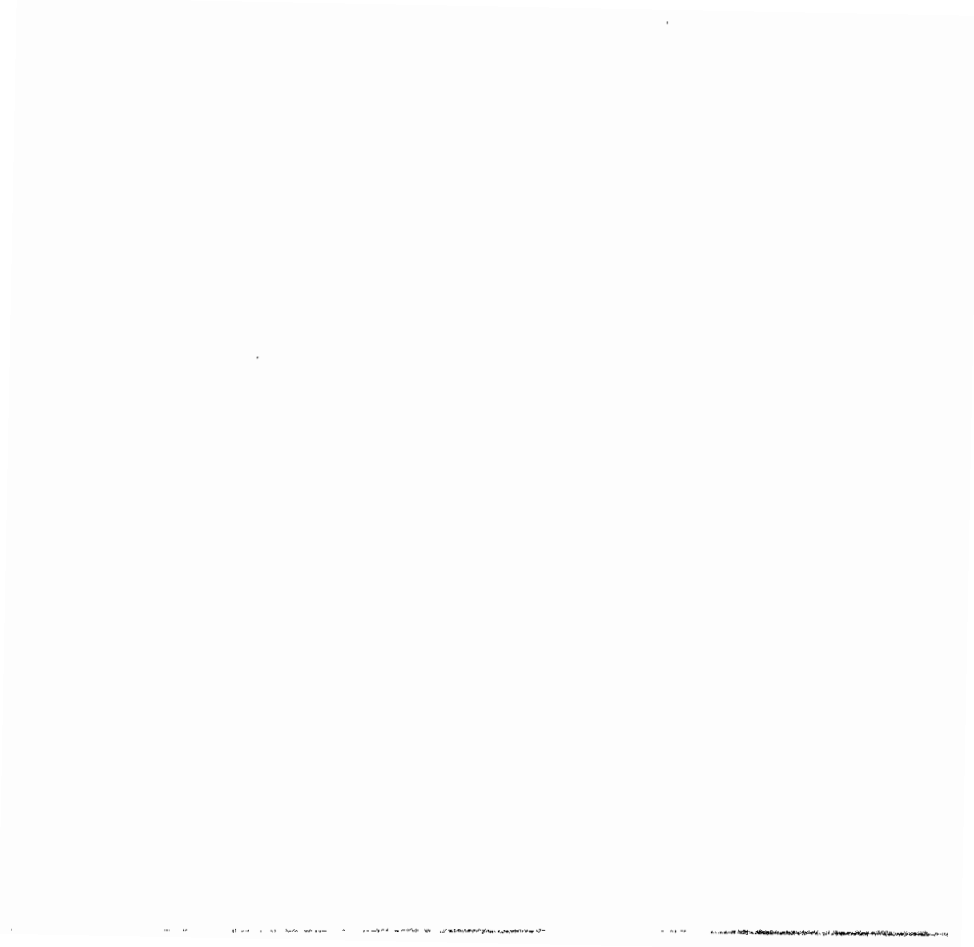


Table des matières

Préface	v
Préambule	vii
Théorie de l'information	1
1 Théorie de l'information	3
1.1 Introduction	3
1.2 Information et mesure de l'information	4
1.3 Codage pour un canal non bruité	7
1.4 Codage en présence de bruit	10
1.5 Théorèmes de Shannon	14
1.6 Observations	14
Compression de données	15
2 Compression de données	17
2.1 Introduction	17
2.2 Types d'algorithmes de compression	18
2.3 Définitions	19
2.4 Techniques de base	19
2.5 Algorithmes statistiques	21
2.6 Algorithmes dynamiques	24
2.7 Limites de la compression	32
2.8 Observations	33
Codes correcteurs d'erreurs	35
3 Généralités sur la théorie des codes	37
3.1 Distance de Hamming et boules	37
3.2 Problème du codage à longueur fixe	39
3.3 Problème du décodage	42

4	Codes linéaires	45
4.1	Préambule mathématique	45
4.2	Définitions	46
4.3	Décodage par les classes latérales	49
4.4	Exemple de décodage par le tableau standard	51
4.5	Codes duaux	53
4.6	Décodage par le syndrome	55
4.7	Exemple de décodage par la table des syndromes	56
4.8	Quelques propriétés des codes linéaires	56
4.9	Observations	59
5	Codes de Hamming	61
5.1	Définition	61
5.2	Propriétés des codes de Hamming	62
5.3	Décodage	63
5.4	Codes de Hamming étendus	63
5.5	Observation	64
6	Code de Golay étendu	65
6.1	Matrice génératrice	65
6.2	Propriétés du code	66
6.3	Décodage	67
6.4	Exemples	68
6.5	Observations	69
7	Codes de Reed-Muller	71
7.1	Définition inductive	71
7.2	Matrices génératrices	72
7.3	Propriétés du code	73
7.4	Décodage	75
7.5	Observations	75
8	Codes cycliques	77
8.1	Description	77
8.2	Représentation polynomiale	77
8.3	Les facteurs de $x^n - 1$ sur $\mathbb{F}_2$	80
8.4	Implantation du codage par les codes cycliques	82
8.5	Implantation du décodage des codes cycliques	83
8.6	Exemples de codes cycliques	85

9 Codes correcteurs de paquets d'erreurs	87
9.1 Définitions	87
9.2 Vers un nouveau problème de décodage	88
9.3 Décodage	89
9.4 Technique d'entrelacement	90
9.5 Technique d'entrelacement avec retard	91
9.6 Technique d'entrelacement croisé	92
10 Introduction aux codes convolutifs	95
10.1 Éléments de base	95
10.2 Codage	96
10.3 Capacité de correction	101
10.4 Décodage	104
11 Application des codes correcteurs dans l'industrie	107
11.1 Code du disque compact	107
11.2 Code des CD-ROM	109
11.3 Code du minitel	110
11.4 Codes des réseaux informatiques	110
11.5 Codage de la parole du GSM	112
11.6 Codes de transmissions satellitaires	116
Complexité	119
12 Théorie de la complexité	121
12.1 Problèmes de décision, de calcul et leur codage	121
12.2 Machines de Turing déterministes	123
12.3 Classe du temps polynomial pour les machines déterministes	125
12.4 Machines de Turing non déterministes	126
12.5 Classe du temps polynomial pour les modèles non déterministes	127
12.6 Les langages NP-complets	130

13 Complexité des problèmes de théorie des codes	133
13.1 Problème du décodage des codes linéaires	133
13.2 La NP-complétude du décodage linéaire	134
13.3 La NP-complétude du problème de l'existence d'un vecteur de poids donné	135
13.4 La NP-complétude du problème de la distance minimale sur $\mathbb{F}_{2^m}$	136
13.5 La NP-complétude du problème de la distance minimale	138
14 Complexité des problèmes de cryptographie	139
14.1 Problème de la somme de sous-ensembles	139
14.2 Problème de la primalité	141
14.3 Une amélioration récente, Primes $\in P$	143
14.4 Problème du logarithme discret	145
Cryptologie	147
15 Introduction à la cryptologie historique	149
15.1 Introduction	149
15.2 Chiffres monoalphabétiques	150
15.3 Cryptanalyse des chiffres monoalphabétiques	153
15.4 Chiffres polyalphabétiques	156
15.5 Cryptanalyse du chiffre de Vigenère	158
15.6 Chiffres à transposition	161
16 Cryptologie technique à clé secrète	163
16.1 Chiffre de Vernam	163
16.2 Machines à rotors	164
16.3 Chiffres produits et itérés	168
16.4 DES	169
17 Cryptanalyses différentielle et linéaire des chiffres itérés	177
17.1 Présentation du chiffre utilisé	177
17.2 Cryptanalyse différentielle	179
17.3 Cryptanalyse linéaire	181

18 Deux chiffres robustes : IDEA et AES	193
18.1 IDEA	193
18.2 La sélection de AES	197
18.3 Rijndael	198
19 Différents modes de fonctionnement	205
19.1 Mode ECB	205
19.2 Mode CBC	206
19.3 Modes OFB et CFB	207
19.4 <i>Modification Detection Code</i> et <i>Message Authentication Code</i> .	208
20 Cryptographie à clé publique	209
20.1 Merkle Hellman	210
20.2 RSA	213
20.3 Le problème du logarithme discret	221
20.4 El Gamal	221
21 Signatures numériques	223
21.1 Mécanisme général de signature	224
21.2 Signature par RSA	224
21.3 Signature par El Gamal	225
21.4 Digital Signature Standard (DSS)	227
22 Fonctions de hachage	231
22.1 Définition d'une fonction de hachage	231
22.2 Fonctions de hachage à collisions difficiles	232
22.3 Dimensionnement du nombre de bits d'une fonction de hachage	233
22.4 Hachage par le logarithme discret	234
22.5 Hachage compressif	235
22.6 Fonction MD5	237
22.7 Fonction SHA-1	241
22.8 Digital Signature Algorithm (DSA)	243
22.9 Autres applications des fonctions de hachage	244
23 Sûreté des chiffres à clé publique	247
23.1 Cryptanalyse du chiffre de Merkle-Hellman	247
23.2 Attaques par factorisation de RSA	255
23.3 Calcul du logarithme discret	261

24 Génération de suites pseudo-aléatoires	267
24.1 Motivation	267
24.2 Génération de suites aléatoires	268
24.3 Génération de suites pseudo-aléatoires	268
24.4 Générateur RSA	271
25 Certification	273
25.1 Introduction	273
25.2 Forme générale d'un certificat	274
25.3 Utilisation et réalisation asymétrique	275
25.4 Utilisation et réalisation symétrique	278
26 Gestion des clés	281
26.1 Exemple introductif	281
26.2 Gestion des clés	281
26.3 Distribution des clés publiques	283
26.4 Distribution des clés secrètes	284
26.5 Mise en accord par Diffie Hellman	288
26.6 Mise à jour des clés	290
27 Applications de la cryptographie à la sécurité des réseaux	291
27.1 Motivation	291
27.2 Introduction à la sécurité	291
27.3 Introduction aux protocoles réseaux	293
27.4 Mots de passe jetables	298
27.5 Kerberos	301
27.6 Sécuriser les courriers électroniques	304
27.7 Sécuriser les échanges TCP/IP	307
27.8 Sécuriser la couche de transport	314
A Utilisation des registres linéaires à décalage	319
A.1 Arithmétique sur les polynômes	320
B Table de polynômes irréductibles primitifs sur $\mathbb{F}_2$	325
C Tables des fréquences relatives des lettres	327
Index	329
Bibliographie	339