Leçon élémentaire de cryptographie *

Yaogan Mensah †

24 novembre 2020

1 Introduction

Le mot *cryptographie* vient du grec "kruptos" qui signifie "caché" et de "graphein" qui veut dire "écrire".
La cryptographie est l'ensemble des techniques permettant de protéger une communication au moyen de codes secrets.

La cryptanalyse est l'ensemble des techniques permettant de trouver le code secret d'une communication.
La cryptologie est l'ensemble formé par la cryptographie et la cryptanalyse.

Un message étant donné en langage naturel (texte clair), le *chiffrement* (ou encryptage ou codage) consiste à le traduire en langage codé (texte chiffré ou cryptogramme). Le *déchiffrement* (ou décodage) consiste à transformer le message codé en langage naturel grâce à la connaissance du code.

Tout système de cryptage (cryptosystème) est composé d'un algorithme de codage plus ou moins compliqué utilisant une ou plusieurs clés de sécurité. Si l'on désigne par $\mathcal{P}$ l'ensemble des messages clairs possibles sur un alphabet $\mathcal{A}$ et par $\mathcal{C}$ l'ensemble des messages codés alors un algorithme de codage est une application injective de $\mathcal{P}$ dans $\mathcal{C}$ en ce sens que deux textes clairs différents ne doivent pas correspondre au même texte chiffré.

Il existe deux types de cryptographie :
- La cryptographie symétrique à clé secrète : la même clé est utilisée pour coder et décoder l'information. Le problème est de transmettre de manière sécurisée la clé à son correspondant.
- La cryptographie asymétrique à clé publique. Deux clés différentes servent à chiffrer et déchiffrer les messages ; une clé publique publiée dans des annuaires et qui sert à coder le message et une clé privée gardée secrète qui permet de déchiffrer le message.

*Cours dispensé à l'Institut Africain d'Informatique (IAI) Version 2.0

†Département de Maths, Université de Lomé, Togo. email: mensahyaogan2@gmail.com

2 Qualité d'un cryptosystème

Tout cryptosystème doit possèder les qualités suivantes :

1. Confidentialité : seules les personnes habilitées ont accès au contenu du message.
2. Intégrité des données : le message ne peut être manipulé/falsifié (insertion, substitution, suppression) sans qu'on s'en aperçoive.
3. Authentification : l'émetteur est sûr de l'identité du récepteur et vice versa.
4. Non-répudiation :

3 Cryptographie symétrique

3.1 Le code de César

3.1.1 Le code de César

J. César [1] pour envoyer des messages à ses généraux déplaçait les lettres de trois rangs vers l'avant. Ainsi, A→D, B→E, ..., X→A, Y→B, Z→C.
On associe à chaque lettre un équivalent numérique : $A \leftrightarrow 0, B \leftrightarrow 1, \cdots, Z \leftrightarrow 25$.
Si m est l'équivalent numérique de la lettre claire et c celui de la lettre codée alors La formule de codage est

$$c = m + 3 \mod 26 \tag{1}$$

et la formule du décodage est

$$m = c - 3 \mod 26. \tag{2}$$

Exemple 3.1 1. Coder le message ATTAQUER

2. Décoder le message FH PHVVDJH HVW WRS VHFUHW

3.2 Le code affine

3.2.1 Codage et décodage

Il généralise le code de J. César. La formule de codage est

$$c = am + b \mod 26 \tag{3}$$

Ici la clé est le couple (a, b). Le code de César est un cas particulier du code affine. Il correspond à la clé $(a, b) = (1, 3)$. La formule de décodage est

$$m = a^{-1}(c - b) \mod 26 \tag{4}$$

où a^{-1} est l'inverse module 26 de a ; ceci sous-entend donc que a et 26 doivent être premiers entre eux.

1. Empereur romain

Le nombre de bonnes clés est $\varphi(12) \times 26 - 1 = 311$, où φ est la fonction d'Euler donnée par

$$\varphi(n) = n \prod_{i=1}^{k} (1 - \frac{1}{p_i}) \tag{5}$$

lorsque la décomposition en produit de facteurs premiers de n est $n = \prod_{i=1}^{k} p_i^{\alpha_i}$.

Exemple 3.2 La formule de codage est $c = 3p + 5 \mod 26$.

1. Coder le message BONJOUR.
2. Décoder le message ERQRSRC.

Exercices 3.3

1. On reçoit le message suivant : JWPNWMRCFWMY
   On sait que le chiffrement est affine et que la lettre E est codée par la lettre E et que la lettre J est codée par la lettre N. Décoder le message reçu.
2. On utilisera le codage informatique des lettres avec le code ASCII, le code ASCII consiste à associer à chaque caractère (lettre de l'alphabet, chiffre, signe de ponctuation, $\cdots$) un code numérique qu'on appelle son code ASCII. Par exemple le code de A est 65, celui de B est 66, celui de a est 97, celui de l'espace est 32. Le code ASCII est un entier x entier tel que $0 \leq x < 256$. Le code ASCII ne constituant pas un codage bien secret, la ligne 3 du tableau ci-dessous consiste à chiffrer le code ASCII en utilisant la fonction du chiffrement suivante : $e(x) = 7x \mod 256$, c-à-d cette fonction associe, à tout x entier appartenant à $[0, 255]$ le reste de la division de $7x$ par 256. Soit $e(x)$ ce reste.
   (a) Complèter le tableau suivant

Message	e	x	a	m	e	n		A	M	D
Code ASCII	101	120	97	109	101	110	32	65	77	68
Message codé										

   (b) Montrer que $183 \times 7 = 1 \mod 256$. En déduire la formule de déchiffrement $d(y)$.
   (c) On généralise l'algorithme précédent en utilisant la fonction de chiffrement $e(x) = ax \mod 256$ avec $a, x \in \mathbb{Z}_{256}$. Quelle est la propriété que doit vérifier a ? Donner la fonction de déchiffrement $d(y)$. Quel est l'espace de clés de cet algorithme ? Combien y a-t-il de bonnes clés ?

3.2.2 Cryptanalyse du code affine

Lorsque la langue de départ et la technique de chiffrement sont connues, on peut exploiter les régularités du langage par le principe d'analyse de la fréquence d'une lettre. Cette technique ne fonctionne bien que si le texte codé est suffisamment long pour avoir des moyennes significatives. Par exemple en français la lettre la plus fréquente est E suivi de S suivi de A. En anglais les lettres les plus fréquentes sont dans l'ordre E, T, A.

Exemples 3.4 Décoder le texte suivant sachant qu'il provient d'un chiffrement affine d'un texte écrit en français : YM QMGGKAM MGN NEL GMYZMN

3.3 Le code de Vigenère

C'est aussi une amélioration du chiffre de César. Son but est de contrer le décodage par analyse statistique des codes par substitution monoalphabétique. Ici on regroupe les lettres par blocs de longueur k. On choisit ensuite une clé constituée de k nombres de 0 à 25 soit $(n_1, \ldots, n_k)$. Pour coder un message on effectue une translation dont le décalage dépend du rang de la lettre dans le bloc : un décalage de n_1 pour la première lettre du bloc, de n_2 pour la deuxième lettre du bloc, ..., de n_k pour la dernière lettre du bloc. Pour décoder il suffit de faire des décalages vers l'arrière. Le grand intérêt du code de Vigenère est que la même lettre sera chiffrée de différentes manière, ce qui sécurise le code du point de vue de l'analyse fréquentielle des lettres.

Exercices 3.5

1. Chiffrer le texte CETTEPHRASENEVEUTRIENDIRE avec la clé (3,1,5, 2).
2. Le texte suivant a été chiffré avec un code Vigenère dont la clef est RIVA. Déchiffrez-le.
   CMN MVBCOUMN SFVO LVA CASQOUUMN DV T'ZSGZDT VB GEJ MXOEWHIVA YE CI HEDWDRV.

3.4 Le code de Hill

3.4.1 Rappel mathématique : Les matrices

Soient la matrice carrée $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ et la matrice colonne $X = \begin{pmatrix} x \\ y \end{pmatrix}$. Le produit AX est donné par

$$AX = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}. \tag{6}$$

Le déterminant de A est

$$\det(A) := ad - bc. \tag{7}$$

Lorsque l'on travaille dans $\mathbb{Z}_n$, la matrice A est inversible si et seulement si son déterminant est premier avec n. Dans ce cas on a

$$A^{-1} = (det(A))^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{8}$$

Et si B est une autre matrice colonne alors

$$AX = B \Longleftrightarrow X = A^{-1}B. \tag{9}$$

3.4.2 Codage et décodage

C'est l'exemple le plus simple de chiffrements polygraphiques. On chiffre deux lettres par deux autres. Chaque digramme clair (m_1, m_2) sera chiffré (c_1, c_2) selon le système de codage (formule de codage)

$$\begin{cases} c_1 & \equiv & am_1 + bm_2 & \mod n \\ c_2 & \equiv & cm_1 + dm_2 & \mod n. \end{cases} \tag{10}$$

où n est le nombre de caractères de l'alphabet utilisé. Le système admet la représentation matricielle

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \mod n. \tag{11}$$

Ici la clé de codage est la matrice $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Lorsque la matrice $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ est inversible, la formule de décodage est donnée par

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \mod n. \tag{12}$$

Exemple 3.6

1. On utilise l'alphabet de 26 caractères A à Z d'équivalents numériques respectifs 0 à 25. Chiffrer le message REVENIR avec la clé $A = \begin{pmatrix} 9 & 4 \\ 5 & 7 \end{pmatrix}$. Déchiffrer le message UBZDTR.
2. Vous interceptez le cryptogramme GFPYJP X ?UYXSTLADPLW
   Vous savez que le chiffrement a été réalisé avec une matrice 2×2 et un alphabet de 29 caractères A à Z d'équivalents numériques 0 à 25, espace=26, ?=27, !=28. Vous savez aussi que les cinq dernières lettres correspondent à la signature de l'expéditeur KARLA. Déterminer le texte clair.
3. Vous savez que votre adversaire utlise un alphabet de 27 caractères A à Z d'équivalent numériques 0 à 25 et espace ␣=26. Si deux lettres ont pour équivalents numériques respectifs x et y alors le digraphe a l'équivalent numérique $27x + y$. La formule de codage est $c = ap + b \mod 27^2$. On suppose que l'étude d'un grand nombre de cryptogrammes révèle que les digraphes les plus fréquents sont dans l'ordre ZA, IA, IW. On admet que les digraphes les plus fréquents en anglais (pour un texte écrit dans notre alphabet de 27 lettres) sont E␣, S␣ et ␣T. Déterminer la formule de décodage et décoder le message NDXBHO.

3.5 Le code de Vernam

Le code de Vernam est aussi appelé Masque jetable/One-time-pad. Le code de Vernam est comme un chiffre de Vigenère avec la caractéristique que la clé de chiffrement a la même longueur que le message clair. Le message clair m et le message chiffré c sont des suites de bits de longueur n. La clé k est une suite binaire de la

même longueur que le clair. Le chiffrement consiste en l'ajout modulo 2, bit à bit, du clair et de la clé.

$$c = m \oplus k = (m_1 \oplus k_1, \cdots, m_n \oplus k_n). \tag{13}$$

Le Masque jetable est le seul système de codage connu comme étant indécryptable. Ce système est parfait dans le cas d'une attaque à chiffré seul, en ce sens que si l'attaquant peut décrypter un (seul) message, c'est-à-dire trouver m connaissant c alors c'est qu'il détient la clé. La clé se déduit aisément de m et de c par

$$k = m \oplus c. \tag{14}$$

C'est pourquoi le code de Vernam est vulnérable aux attaques à clairs connus. Aussi la clé ne doit-elle être utilisée qu'une seule fois. De plus ce code est impraticable à moins que le message soit de petite longueur. Cependant ce système est à la base de la cryptographie par flots qui concerne tous les chiffrements en ligne (très rapides). Il est aussi couramment utilisé de nos jours par les États. En effet, ceux-ci peuvent communiquer les clefs à leurs ambassades de manière sûre via la valise diplomatique.

Exemple 3.7

$m =$	1	0	0	1	0	1	0	0	0	1	0	1	1
$k =$	0	0	1	1	1	0	0	1	0	1	1	0	1
$c =$	1	0	1	0	1	1	0	1	0	0	1	1	0

4 Cryptographie asymétrique

4.1 Fonctions à sens unique

Une fonction (algorithme) f est dite à sens unique si connaissant x il est facile de déterminer $f(x)$ mais connaissant $y = f(x)$ il est calculatoirement impossible de trouver x. Une application possible est *la sécurisation du mot de passe*. En effet la plupart des logiciels demande *un code d'accès* à l'utilisateur. Si ce code d'accès est inscrit directement sur le disque dur un utilisateur qui a de l'expertise pourra le lire. On utilise alors une fonction à sens unique f. On inscrit $f(x)$ sur le disque dur plutôt que le mot de passe x. Voici quelques exemples de foncions à sens unique

1. Dans $\mathbb{Z}_n$, pour x donné, il est facile de calculer x^2 mais pour y donné il est difficile de trouver x tel que $x^2 = y$.
2. La décomposition en facteurs premiers d'un entier n est un problème difficile dès que n est suffisamment grand (par exemple factoriser le nombre $2^{67} - 1$). Étant donnés deux nombres premiers p et q il est aisé de calculer $n = pq$ mais étant donné n il est difficile de le factoriser c'est-à-dire de retrouver les facteurs premiers p et q.
3. Le problème du sac à dos Soit une suite d'entiers positifs $a_1, \cdots, a_n$ et soit un nombre t. Déterminer un sous-ensemble I de $\{1, \cdots, n\}$ tel que $\sum_{i \in I} a_i = t$ est un problème difficile. Par exemple soit la suite de nombres 14, 28, 56, 82, 90, 132, 197, 284, 341, 455. Le nombre 516 peut-il s'écrire comme somme de certains de ces nombres ? Même question avec 515.

L'algorithme suivant, appelé *algorithme glouton*, permet de résoudre le problème du sac à dos.

4.2 Le code de Merkle-Hellman ou méthode des empilements

4.3 Un exercice introductif

4.4 Cours

On considère une suite super-croissante de k entiers (généralement appelée le sac). Une telle suite est telle que chacun de ses éléments est plus grand que la somme des entiers qui le précèdent. Par exemple

$$S = (1, 2, 4, 9). \tag{15}$$

On choisit ensuite un multiplicateur m et un module n tels que n soit premier et supérieur à la somme des éléments dans S et m et n soient premiers entre eux. Par exemple

$$m = 15 \text{ et } n = 17. \tag{16}$$

On obtient la clé publique en multipliant modulo n le nombre m par les éléments de S. Par exemple ici la clé publique est

$$H = (15, 13, 9, 16). \tag{17}$$

tandis que la clé secrète est (S, m, n).
Le message clair est composé de blocs de h bits (h=longueur du sac). Par exemple on prend

$$P = 0100101110100101$$

On a les blocs 0100 1011 1010 0101. Le codage se réalise comme suit.

$[0, 1, 0, 0]\cdot[15, 13, 9, 16] = 13$, $[1, 0, 1, 1]\cdot[15, 13, 9, 16] = 40$, $[1, 0, 1, 0]\cdot[15, 13, 9, 16] = 24$, $[0, 1, 0, 1]\cdot[15, 13, 9, 16] = 29$. Le message codé est alors $(13, 40, 24, 29)$. Pour obtenir le déchiffrement le destinataire qui connaît S, m et n, détermine l'inverse modulo n de m. Dans notre exemple $m^{-1} = 8$. Ensuite il fait

$$13 \times 8 \mod 17 = 104 = 2 = [1, 2, 4, 9] \cdot [0, 1, 0, 0]$$

Et ainsi de suite.....

4.5 Exercices

1. Alice choisit le sac à dos $S = (s_i)_{1\leq i\leq 4} = (4, 5, 10, 21)$, $n = 53$ et $m = 13$.
   (a) Calculer le sac à dos déguisé $H = mS = (ms_i \mod n)_{1\leq i\leq 4}$. La clé publique de Abalo est H.
   (b) Binéta veut envoyer à Abalo le message $M = (1, 0, 0, 1)$. Calculer le message crypté envoyé par Binéta.
   (c) Abalo a reçu de Mazalo le message crypté $C = 18$. XXXXX à revoir Quel est le message clair ?

2. On considère l'alphabet de 32 caractères A à Z d'équivalent numériques 0 à 25, -=26, ,=27, .=28, :=29, !=30 et ?=31.

   (a) Soit $\omega = (1, 8, 15, 50, 105)$ une suite super-croissante, $n = 211$, $m = 5$ et la permutation $\sigma = (3, 2, 5, 1, 4)$. Calculer la clé publique $\omega' = (\omega'_1, \omega'_2, \omega'_3, \omega'_4, \omega'_5)$ avec $\omega'_i = m\omega_{\sigma(i)} \mod n$.

   (b) Utiliser ce système pour chiffrer le message ALERT (chaque lettre est codée sur 5 bits et chaque suite de 5 bits est ensuite codée avec le système de Merkle-Hellman)

   (c) Vous recevez le message chiffré suivant : (187, 40, 45, 0, 5, 147, 178). Le déchiffer.

4.6 RSA

4.6.1 L'exponentiation modulaire

La méthode d'exponentiation modulaire connue sous le nom de *square-and-multiply* (mettre au carré et multiplier) réduit le nombre de multiplications modulaires nécessaires pour calculer $z = x^c \mod n$. On décompose d'abord c en binaire :

$$c = \sum_{i=1}^{l} c_i 2^i \tag{18}$$

1. $$z \longleftarrow 1$$

2. $$\text{pour } i = l \text{ jusqu'à } 0 \text{ fais}$$

3. $$\text{Si } c_i = 0 \text{ alors } z \longleftarrow z^2 \mod n$$

4. $$\text{Si } c_i = 1 \text{ alors } z \longleftarrow z^2 \times x \mod n$$

Exemples 4.1 Calculer $41^{37} \mod 527$.
Réponse : 113

4.6.2 Le chiffrement RSA

Il fut publié en 1978 par Rivest, Shamir et Adleman. Ce chiffrement est basé sur la difficulté de factorisation des grands nombres. Pour la mise en oeuvre du RSA,

1. Bob engendre deux grands nombres premiers p et q.
2. Bob calcule $n = pq$ et $\varphi(n) = (p-1)(q-1)$.
3. Bob choisit un nombre e aléatoire entre 1 et $\varphi(n)$ et premier avec $\varphi(n)$.
4. Bob calcule d l'inverse modulo $\varphi(n)$ de e par l'algorithme d'Euclide étendu.

5. Bob publie (n, e) (clé publique) et garde secret d (clé privée)

A la fin de cette opération Bob peut détruire p, q et $\varphi(n)$ car il n'en aura plus besoin. Pour envoyer un message m à Bob, Alice utilise la formule de codage

$$c = m^e \mod n \tag{19}$$

Bob décode le message reçu par la formule

$$m = c^d \mod n \tag{20}$$

Exercices 4.2

1. On considère la clé publique RSA $(n, e) = (319, 11)$.
   (a) Quel est le message codé C avec cette clé correspondant au message $M = 100$?
   (b) Calculer d la clé privée correspondant à la clé publique e.
   (c) Déchiffrer le message $C = 1033$.
2. On donne la clé publique RSA $(n, e) = (21, 5)$.
   (a) Donner la clé privée associée.
   (b) On se place sur l'alphabet $\mathscr{A}$ formé des 21 premières lettres de l'alphabet français ; A se code 0, ..., U se code 20. Coder le message ILFAITBEAU. Décoder le message NONICPCQUK
3. Montrer que dans le système RSA si l'on connaît $\varphi(n)$ alors on peut factoriser n. Application : $n = 7663$ et $\varphi(n) = 7488$.
4. Un professeur envoie ses notes au Secrétariat de l'Ecole par email. La clé publique RSA du professeur est $(n_1, e_1) = (55, 3)$ et celle du Secrétariat est $(n_2, e_2) = (33, 3)$.
   (a) Déterminer la clé privée du professeur et du Secrétariat.
   (b) Pour assurer la confidentialité de ses messages, le professeur chiffre les notes avec la clé RSA du Secrétariat. Quel message chiffré correspond à la note 13 ?
   (c) Pour assurer l'authenticité de ses messages, le professeur signe chaque note avec sa clé privée et chiffre le résultat avec la clé publique RSA du Secrétariat. Le Secrétariat reçoit ainsi le message 26. Quelle est la note correspondante ?

4.7 Le chiffrement ElGamal

Ce code est inventé par Taher ElGamal en 1984. C'est un chiffrement à clé publique qui est à la base de la norme U.S. de signature électronique. Sa solidité est basée sur la difficulté de calculer des logarithmes discrets. Étant donné x, y et p, le problème du logarithme discret consiste à trouver λ tel que $y = x^\lambda \mod p$.

Le principe du chiffrement ElGamal est le suivant :

1. Bob choisit un nombre premier p très grand tel que $p-1$ ait un grand facteur premier.
2. Bob produit une clé privée a telle que $a \in \{1, \cdots, p-2\}$.
3. Bob produit une clé publique (p, α, β) telle que $\beta = \alpha^a \mod p$, le nombre α est pris tel que $\alpha \in \{0, \cdots, p-1\}$ et pour tout $k \in \{1, \cdots, p-2\}$, $\alpha^k \neq 1 \mod p$.

Alice souhaite envoyer un message $m < p$ à Bob. Pour ce faire elle choisit un nombre aléatoire $k \in \{1, \cdots, p-2\}$. Elle calcule alors

$$c_1 = \alpha^k \mod p \tag{21}$$

$$c_2 = m\beta^k \mod p \tag{22}$$

Le message chiffré est alors $c = (c_1, c_2)$. C'est un message deux fois plus long que le message clair.

Le principe de déchiffrement est le suivant : à la réception, Bob calcule

$$r_1 = c_1^a \mod p. \tag{23}$$

Ensuite il divise c_2 par r_1 modulo p. Il retrouve alors m.

Pour décrypter le message il faut trouver la clé privée a solution de l'équation $\beta = \alpha^a \mod p$ (problème du logarithme discret.)

Exercices 4.3 1. Bob choisit la clé privée $a = 765$ et la clé publique $(p = 2579, \alpha = 2, \beta)$

   (a) Préciser β.

   (b) Pour coder le message $m = 1299$, Alice choisit $k = 853$. Préciser le cryptogramme reçu par Bob.

   (c) Décoder le message reçu par Bob.