La cryptologie, ou cryptographie et cryptanalyse, de l'ère pré-informatique

Razafimahefa F. O. [1], Randriamitantsoa P.A. [2], Randriamitantsoa A. A[3]

Laboratoire de Recherche en Télécommunication, Automatique, Signal et Images (LR-TASI)

École Doctorale en Sciences et Techniques de l'Ingénierie et de l'Innovation (ED-STII)

Équipe d'Accueil Doctorale Télécommunication, Automatique, Signal et Images (EAD-TASI)

Université d'Antananarivo

BP 1500, Ankatso – Antananarivo 101 – Madagascar

[1]ombana.razafimahefa@gmail.com , [2]rpauguste@gmail.com, [3]andriau23@gmail.com

Résumé

L'être humain était et est fasciné par l'utilité de cacher un message et de ne divulguer celui-ci qu'à ceux ou celles à qui il est destiné. Deux disciplines distinctes existent pour ce faire. D'un côté, la stéganographie, consiste à cacher le message dans un contenu. De l'autre, la cryptographie, consiste à présenter le message sous une autre forme pour que seul celui à qui il est destiné, soit en mesure de le lire. Le présent ouvrage tournera autour de la cryptographie.

D'une manière générale la cryptographie se définit comme étant l'étude des techniques permettant la communication sécurisée en présence de partie tierce que l'on nomme adversaires. Dans la littérature, l'envoyeur est souvent appelé sous le nom Alice ou personne A. Le récepteur Bob ou personne B. L'adversaire est quant à lui dénommé Eve de l'anglais eavesdropper, l'eavesdropping se définissant comme étant l'écoute secrète de conversation privée d'autrui, sans son consentement.

Avant d'aller plus loin dans notre recherche doctorale dans le domaine de la cryptographie, il nous semble impératif de connaitre dans un premier temps ce que fut l'historique dudit domaine, soit en d'autres termes en quoi consistait la cryptographie au tout début, durant l'ère pré-informatique.

Mots clés : *Cryptologie, cryptographie, cryptanalyse, pré-informatique, formulation mathématique.*

Abstract

Humans were and are fascinated by the usefulness of hiding a message and only disclosing it to those to whom it is intended. Two separate disciplines exist to do this. On the one hand, steganography consists in hiding the message in a content. On the other hand, cryptography, consists in presenting the message in another form so that only the one for whom it is intended, is able to read it. This work will revolve around cryptography.

Generally speaking, cryptography is defined as the study of techniques allowing secure communication in the presence of a third party known as adversaries. In the literature, the sender is often called under the name Alice or person A. The receiver Bob or person B. The adversary is called Eve (from the English eavesdropper), the eavesdropping being defined as listening the private conversation of others without their consent.

Before going further in our doctoral research in the field of cryptography, it seems imperative to us to know at first what was the history of the said field, that is to say in other words what consisted of cryptography at the very beginning, during the pre-Information Technology era.

Keywords: *Cryptology, cryptography, cryptanalysis, pre-computer science, mathematical formulation.*

1 Chiffrement par substitution

Le terme *substitution* se réfère au fait qu'une lettre en claire est remplacée par une lettre chiffrée suivant une règle bien définie.

1.1 Chiffrement par addition

Désignons par C le texte chiffré et par P le texte en clair. Posons k la clé de chiffrement, qui sera dans le cas du présent chiffrement par addition (et pour le cas des quelques méthodes de chiffrement qui vont suivre d'ailleurs) un nombre entier. [1], [2]

$$C \equiv P + k \ modulo\ 26 \qquad (1.01)$$

Notons que le nombre 26 n'est autre que le nombre de lettres dans l'alphabet.

Exemple : En prenant $k = 3$, la lettre *a* après chiffrement devient *D* (*a* est équivalent à 1, $1 + 3$ donne 4, ce qui revient à *D*).

Le déchiffrement se fait par la formule qui suit : [1], [3]

$$P \equiv C - k \ modulo\ 26 \qquad (1.02)$$

Le nombre de caractères dans l'alphabet étant de 26, le nombre de valeurs que peut prendre notre clé de chiffrement k est limité à 26. Cela rend la méthode de chiffrement par addition sensible à l'*attaque par force brute* (*brute force attack*). Cette attaque consiste à essayer un à un les valeurs de clés possibles jusqu'à obtenir un résultat satisfaisant. [3]

1.2 Chiffrement par multiplication

La formule utilisée revient à ce qui va suivre : [1]

$$C \equiv kP \ modulo\ 26 \qquad (1.03)$$

Notons quand même que la valeur que peut prendre k ici devra suivre certaines conditions. Etant donné que 26 (le nombre de lettres de l'alphabet pour rappel) si décomposé en nombre

premiers donne 2×13. Pour que chacune des lettres en claire de l'alphabet ait une et une seule correspondance, k devrait être ni multiple de 2 ni de 13. k peut-être donc parmi $1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23$ et 25. Soit 13 valeurs possibles au total.

Exemple : Pour $k = 3$. a devient C après chiffrement $(3 \times 1 = 3)$. x devient T (soit $3 \times 24\ modulo\ 26 = 20$).

Le déchiffrement se fait logiquement selon la formule ci-dessous : [1]

$$P \equiv \frac{1}{k} C\ modulo\ 26 \qquad (1.04)$$

Il y a cependant un léger problème à utiliser cette formule. En effet, prenons le cas où C n'est pas divisible par k (la lettre chiffrée est A par exemple, ce qui équivaut à 1 et $\frac{1}{3}$ n'est pas un entier donc difficilement transformable en lettre).

Nous allons faire entrer la notion d'inverse multiplicative modulaire $\bar{k}$ telle que $\frac{1}{k} C\ modulo\ 26 = \bar{k} C\ modulo\ 26$. Dans ce cas nous avons notre formule de déchiffrement finale :

$$P \equiv \bar{k} C\ modulo\ 26 \qquad (1.05)$$

L'*algorithme d'Euclide étendu*, nous permet d'arriver à l'égalité qui suit : $9 \times 3 - 1 \times 26 = 1$, d'où notre $a = 9$ et $b = 1$. Soit $a = \bar{k} = 9$.

Pour en revenir au déchiffrement de notre lettre encodée A. A équivaut à 1 la lettre en claire correspondant est donc $P = 9 \times 1\ modulo\ 26 = 9$, ce qui revient à la lettre i.

Une autre forme d'attaque peut en venir à bout de la présente méthode de chiffrement, celle connue sous le nom de l'*analyse de la fréquence des lettres*. Cette technique revient à l'arabe *Abu Yusuf Yaqub ibn Ishaq al-Sabbah al-Kindi*, développé aux alentours du IX^{ème} siècle.

Dans toutes les langues, certaines lettres se répètent plus souvent que d'autres. Dans la langue anglaise par exemple, les lettres *e*, *t* et *a* sont celles qui apparaissent les plus fréquemment, respectivement de l'ordre de 12%, 9% et 8%. Pour un message encodé relativement long, nous pourrions ainsi étudier la fréquence d'apparition des lettres, en sortir les 3 premières qui reviennent le plus. Elles correspondront probablement aux lettres en claire *e*, *t* et *a*. Pour les messages courts par contre, cette méthode perd en précision. [1]

1.3 Chiffrement affine

Le chiffrement se fait par la formule qui suit : [1], [3]

$$C \equiv kP + m\ modulo\ 26 \qquad (1.06)$$

Exemple : pour $k = 3$ et $m = 13$, notre lettre en claire x deviendra T $(3 \times 24\ modulo\ 26 = 20)$ après application du chiffrement par

multiplication. *T* deviendra par la suite *G* ($20 + 13\ modulo\ 26 = 7$) après application de la seconde méthode de chiffrement, le chiffrement par addition.

Le déchiffrement se fait en ordre inversé. Effectuer le déchiffrement par addition, puis faire suivre par le déchiffrement par multiplication : [1], [3]

$$P \equiv \bar{k}(C - m)\ modulo\ 26 \qquad (1.07)$$

Exemple : G après le déchiffrement par addition devient *T* ($7 - 13\ modulo\ 26 = 20$). Nous avons vu dans notre paragraphe 1.2 que $\bar{3} = 9$. Après déchiffrement par multiplication *T* revient à notre lettre en claire *x* ($9 \times 20\ modulo\ 26 = 24$).

1.4 Chiffrement homophonique

Le *chiffrement homophonique*, consiste à attribuer plus d'une correspondance en lettre chiffrée à une lettre en claire ayant une fréquence d'apparition élevée. A la lettre en claire *e* par exemple, à laquelle la fréquence d'apparition 12% est attribuée, on en fait correspondre 4 (quatre) lettres chiffrées (*V, @, &* et – par exemple). De sorte que la fréquence de la lettre en claire *e* retombe à 3% ($3\% = 12\%/4$) lorsqu'elle est chiffrée, sous condition que le choix entre les lettres chiffrées *V, @, &* et – se fasse d'une manière totalement aléatoire. Et comme 3% étant approximativement la fréquence d'apparition de toutes les lettres de l'alphabet, si chacune d'entre elles avaient la même probabilité d'apparition ($3\% \simeq 1/26$), nous sommes ici donc en train d'outrepasser le problème lié à la fréquence d'apparition des lettres. [1], [2], [3]

1.5 Phi test ou calcul de l'indice de coïncidence

Le *phi test* a été introduit par *William Friedman* dans les années 1920. Il consiste à calculer l'*indice de coïncidence (IC)* correspondant à un message chiffré. L'indice de coïncidence est la probabilité pour que 2 mêmes lettres apparaissent simultanément dans le message. La formule est comme suit :

$$IC = \sum_{i=1}^{z} (\mathbb{P}(c_i))^2 \qquad (1.08)$$

Où z étant le nombre de lettres de l'alphabet ;

c_i le i-ième lettre de l'alphabet du message chiffré ;

$\mathbb{P}(c_i)$ dénote la probabilité d'apparition de la lettre c_i dans le message chiffré.

$$IC_{CUD} = \sum_{i=1}^{26} \left(\frac{1}{26}\right)^2 = 26 \times \frac{1}{26^2} = \frac{1}{26} \simeq 0.038 \qquad (1.09)$$

IC_{CUD}, pour Indice de Coïncidence pour Chiffre Uniformément Distribué.

Notre formule sera équivalente à $IC_{anglais} = (\mathbb{P}(A))^2 + (\mathbb{P}(B))^2 + \cdots + (\mathbb{P}(Z))^2$, qui revient à la valeur ci-dessous selon la bibliographie [1] :

$$IC_{anglais} = 0.066 \qquad (1.10)$$

Ce que l'on peut conclure ici est que, pour tout message chiffré, l'on devrait avoir $0.038 \leq IC \leq 0.066$. Si IC est plutôt proche de 0.038, c'est que le chiffrement est probable d'avoir utilisé la méthode homophonique. Si plutôt proche de 0.066, le chiffrement est sûrement un simple chiffrement mono-alphabétique où la fréquence d'apparition des lettres a été répercutée dans le message chiffré.

1.6 Chiffrement affine de Hill, un chiffrement type polygraphique

Notons le fait que polygraphique signifie que le processus de chiffrement traite plus d'une lettre en claire à la fois.

La présente méthode de chiffrement fut développée par *Lester Hill* dans les années 1929. Elle consiste à subdiviser le texte en clair en plusieurs blocs de *n* lettre chacun. Et d'appliquer le processus de chiffrement tour à tour sur ces blocs de *n* lettres. [1], [2], [3]

La formule pour le chiffrement est comme ce qui suit : [1]

$$C \equiv KP + M \ modulo\ 26 \qquad (1.11)$$

Où :

C et P sont des vecteurs de taille n, dont les éléments sont respectivement les n lettres chiffrées et celles en claires.

$C = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ et $P = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}$;

K est une matrice carrée inversible de dimension $n \times n$. Elle représente un genre de facteur multiplicatif comme celui que nous avons vu dans notre paragraphe 1.2. Cette fois ci, c'est le déterminant de la matrice K, noté $det(K)$ et le nombre de lettre de l'alphabet 26 qui doivent être premiers entre eux :

$K = \begin{pmatrix} k_{1,1} & \cdots & k_{1,n} \\ \vdots & \ddots & \vdots \\ k_{n,1} & \cdots & k_{n,n} \end{pmatrix}$;

M est également un vecteur de taille n et représente le facteur additionnel qui assure le chiffrement par addition, se référer au paragraphe 1.1 :

$M = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}$;

Exemple : Prenons $n = 2$, $K = \begin{pmatrix} 3 & 5 \\ 6 & 1 \end{pmatrix}$, et $M = \begin{pmatrix} 13 \\ 9 \end{pmatrix}$. Et essayons de chiffrer le digramme en clair *et*, soit $P = \begin{pmatrix} 5 \\ 20 \end{pmatrix}$:

$C \equiv \begin{pmatrix}3 & 5\\ 6 & 1\end{pmatrix}\begin{pmatrix}5\\ 20\end{pmatrix} + \begin{pmatrix}13\\ 9\end{pmatrix}\ modulo\ 26 = \begin{pmatrix}128\\ 59\end{pmatrix}\ modulo\ 26 = \begin{pmatrix}24\\ 7\end{pmatrix}$, soit le digramme chiffré *XG*.

Le déchiffrement s'effectue par la formule qui suit : [1]

$$P \equiv \overline{\det(K)}\ {}^{t}com(K)(C - M)\ modulo\ 26 \qquad (1.12)$$

Où :

$\det(K)$ se réfère au déterminant de la matrice K. $\overline{\det(K)}$ donc représente la multiplicative inverse de $\det(K)$ modulo 26 ;

$com(K)$ représente la comatrice de la matrice K et ${}^{t}com(K)$ se réfère à la transposée de $com(K)$;

Exemple : Déchiffrons le premier digramme *XG* que nous avons obtenu dans notre paragraphe précédent 1.5.2.

$det(K) = (3)(-1)^{1+1}(1) + (6)(-1)^{1+2}(5) = 3 - 30 = -27$;

L'inverse multiplicatif de -27 modulo 26 en suivant la méthode détaillée dans le paragraphe 1.3.3 donne $\overline{det(K)} = -1$.

La matrice comatrice de *K* est :

$com(K) = \begin{pmatrix}(-1)^{1+1}(1) & (-1)^{1+2}(6)\\ (-1)^{1+2}(5) & (-1)^{2+2}(3)\end{pmatrix} = \begin{pmatrix}1 & -6\\ -5 & 3\end{pmatrix}$;

${}^{t}com(K) = \begin{pmatrix}1 & -5\\ -6 & 3\end{pmatrix}$;

Notre digramme en clair s'obtient donc comme suit :

$P \equiv (-1)\begin{pmatrix}1 & -5\\ -6 & 3\end{pmatrix}\left(\begin{pmatrix}24\\ 7\end{pmatrix} - \begin{pmatrix}13\\ 9\end{pmatrix}\right)\ modulo\ 26 = \begin{pmatrix}-27\\ 72\end{pmatrix}\ modulo\ 26 = \begin{pmatrix}5\\ 20\end{pmatrix}$, soit *ET*.

Nous avons connu jusqu'ici des attaques qui se basent uniquement sur la connaissance du texte chiffré. Un autre cas qui pourrait se présenter est que l'attaquant pourrait être en possession d'une portion de texte en clair également. L'attaquant peut venir à bout du présent chiffrement s'il est en possession d'autant de couple de (lettre en claire, lettre chiffrée) que d'inconnus dans notre équation de chiffrement (1.11). [1]

Prenons le cas par exemple des 2 digrammes que nous avons chiffrés précédemment (*et*, *XG*) et (*es*, *SF*) auxquels nous allons rajouter le couple de (lettre en claire, lettre chiffrée), également obtenu par le même processus de chiffrement, (*bi*, *LD*). Nous aurons :

(*et*, *XG*) donne $\begin{pmatrix}24\\ 7\end{pmatrix} = \begin{pmatrix}k_{1,1} & k_{1,2}\\ k_{2,1} & k_{2,2}\end{pmatrix}\begin{pmatrix}5\\ 20\end{pmatrix} + \begin{pmatrix}m_1\\ m_2\end{pmatrix}$;

(*es*, *SF*) donne $\binom{19}{6} = \begin{pmatrix} k_{1,1} & k_{1,2} \\ k_{2,1} & k_{2,2} \end{pmatrix} \binom{5}{19} + \binom{m_1}{m_2}$;

Et (*bi*, *LD*) donne $\binom{12}{4} = \begin{pmatrix} k_{1,1} & k_{1,2} \\ k_{2,1} & k_{2,2} \end{pmatrix} \binom{2}{9} + \binom{m_1}{m_2}$;

Notre attaquant serait alors en possession d'une sorte de 6 équations à 6 inconnues, à l'issue desquelles il est en mesure de reconstituer les paramètres *K* et *M* du chiffrement affine de Hill utilisé.

$$\begin{cases} 24 = 5k_{1,1} + 20k_{1,2} + m_1 \ modulo\ 26\ (1) \\ 7 = 5k_{2,1} + 20k_{2,2} + m_2 \ modulo\ 26\ (2) \\ 19 = 5k_{1,1} + 19k_{1,2} + m_1 \ modulo\ 26\ (3) \\ 6 = 5k_{2,1} + 19k_{2,2} + m_2 \ modulo\ 26\ (4) \\ 12 = 2k_{1,1} + 9k_{1,2} + m_1 \ modulo\ 26\ (5) \\ 4 = 2k_{2,1} + 9k_{2,2} + m_2 \ modulo\ 26\ (6) \end{cases} ;$$

1.7 Chiffre d'Alberti

Le chiffre d'Alberti est notre premier exemple de chiffre *poly-alphabétique*. Un chiffre est dit comme tel (poly-alphabétique) lorsqu'il fait correspondre plus d'une lettre chiffrée à chacune des lettres en claires existantes.

Pour chiffrer un texte en clair, Alberti introduisit un outil nommé *Disque d'Alberti*. Il consistait en 2 cercles concentriques. [1]

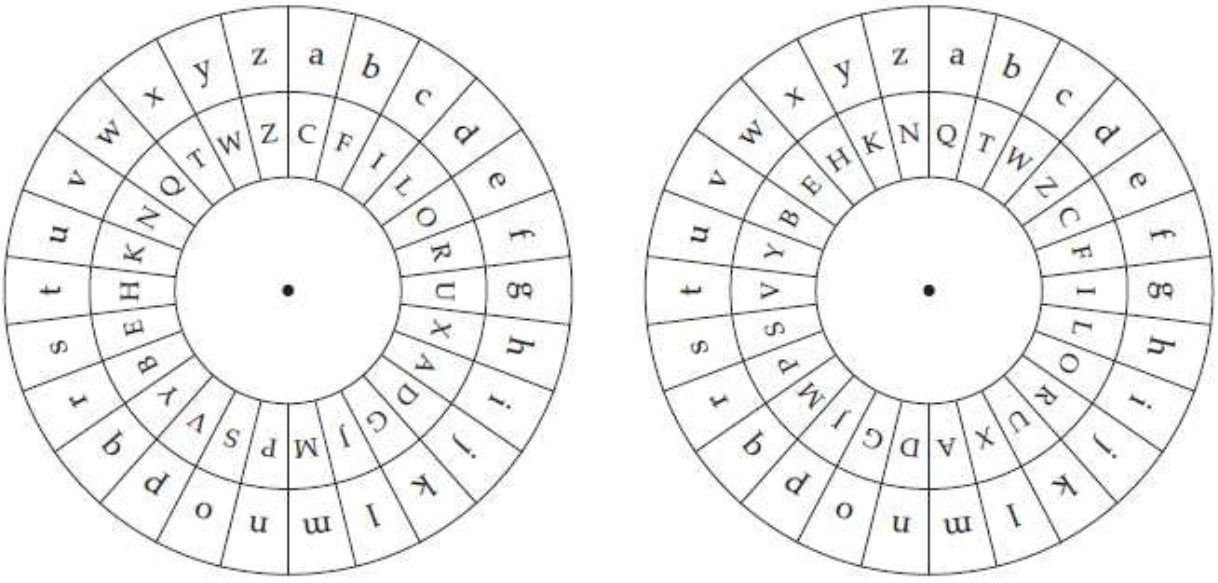

Figure 1.01 : *Disque d'Alberti avec dans un premier temps* $\boldsymbol{m_1 = 0}$ *à gauche, puis* $\boldsymbol{m_2 = -4\ modulo\ 26 = 22}$ *s à droite.*

En termes de modélisation mathématique, le chiffre d'Alberti est une combinaison d'un chiffre de permutation de clé k, que nous avons noté $f(k, P)$, et d'un chiffre d'addition de clé m_i.

La clé du chiffre d'addition, comme nous avons vu plus haut, a été proposé à être changer tous les 3 à 4 mots. D'où l'indice i dans m_i.

$$C \equiv f(k,(P+m_i))\ modulo\ 26 \qquad (1.13)$$

Une simplification que nous pouvons proposer est de prendre $f(k,P) = kP\ modulo\ 26$ Cela nous amène à un genre de chiffre affine comme suit :

$$C \equiv k(P+m_i)\ modulo\ 26 \qquad (1.14)$$

Exemple : Nous allons chiffrer la lettre *a* de notre premier syllabe *as* : $a \equiv 1, C \equiv 3 \times (1+0)\ modulo\ 26 \equiv 3 \equiv C$.

Chiffrons maintenant la première lettre *o* de notre seconde syllabe plus haut *ont*. $o \equiv 15, C \equiv 3 \times (15+22)\ modulo\ 26 \equiv 111\ modulo\ 26 \equiv G$.

Mentionnons quand même notre message chiffré final qui est *aCEeGDV* où *a* étant notre première lettre clé, et *e* en est la seconde.

1.8 La Tabula Recta

Le chiffrement utilisant le *Tabula Recta* ou *tableau dans le sens normal* (par opposition à sens inverse) fut introduit par *Johannes Trithemius* dans les années 1508. [1], [3]

En d'autres termes la modélisation mathématique est comme suit :

$$c_i \equiv p_i + i\ modulo\ 26 \qquad (1.15)$$

Où c_i étant la i-ième lettre chiffrée, p_i la i-ième lettre en claire et i étant le numéro d'itération en commençant par $i = 1$.

Exemple : Si nous voulons chiffrer la syllabe *as*, cela deviendra *BU*.

Il va de soi qu'une certaine Eve qui connait que la méthode de chiffrement utilisée par Alice et Bob se trouve être un simple Tabula Recta sans clé, déchiffrera sans problème le texte chiffré. Ce qui nous amène à l'amélioration proposée par *Vigenère*, celle d'introduire une notion de clé dans le présent mode de chiffrement.

1.9 Chiffre de Vigenère

Vigenère a suggéré qu'Alice et Bob devrait s'entendre sur un mot ou une phrase à définir comme clé à utiliser dans leur chiffrement faisant introduire le Tabula Recta.

La modélisation mathématique est donc comme suit : [1], [2], [3]

$$c_i \equiv p_i + m_i\ modulo\ 26 \qquad (1.16)$$

m_i étant la i-ième lettre du mot de passe, revenir à la première lettre à chaque fois que toutes les lettres aient été utilisées.

Exemple : Chiffrons la syllabe *ont* par le chiffre de Vigenère, en utilisant comme clé *as*.

- La lettre en claire *o* avec la lettre clé *a* donnent

P si directement lu sur le Tabula Recta. Si calculé $c \equiv 15\ (soit\ o) + 1(soit\ a)\ modulo\ 26\ \equiv 16\ modulo\ 26 \equiv 16 \equiv P$;

- *n* et *s* donnent *G* ;

- *t* et *a* (nous étions arrivés à la dernière lettre *s* de la clé et revenons à la première, d'où *a*) donnent *U*.

	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	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
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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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

Figure 1.02 : *Tabula Recta par Johannes Trithemius et amélioré par Vigenère.*

1.10 Indice de coïncidence pour un chiffre à clé périodique

L'indice de Coïncidence pour un Chiffre à Clé Périodique ou IC_{CCP}, est donné par la formule (1.17). [1]

Où n représente la longueur du message ;

l la longueur de la clé utilisée ;

IC_{CUD} et $IC_{anglais}$ respectivement l'indice de coïncidence d'un chiffre uniformément distribué et celui de la langue anglaise ;

$$IC_{CCP} = \frac{n-\frac{n}{l}}{n-1} \times IC_{CUD} + \frac{\frac{n}{l}-1}{n-1} \times IC_{anglais} = \frac{n-\frac{n}{l}}{n-1} \times 0.038 + \frac{\frac{n}{l}-1}{n-1} \times 0.066 \quad (1.17)$$

Remarquons que $IC_{CCP} = IC_{CUD} = 0.038$ pour $l = n$, en d'autres termes toutes les lettres du message chiffré ont été chiffrées différemment. Et $IC_{CCP} = IC_{anglais} = 0.066$ pour $l = 1$, en d'autres termes pour un chiffrement mono-alphabétique.

La connaissance de IC_{CCP} (peut être calculée vu que le message chiffré est connu par Eve) et de n (connu également donc) nous emmène à un genre d'équation à une inconnue l. D'où Eve qui peut suggérer une valeur de la longueur de la clé utilisée.

Exemple : Pour $n = 235$ et $IC_{CCP} = 0.044$, nous avons :

$$0.044 = \frac{235-\frac{235}{l}}{235-1} \times 0.038 + \frac{\frac{235}{l}-1}{235-1} \times 0.066;$$

Soit après résolution, $l = 4.59$. Une bonne suggestion en est $l = 5$.

1.11 Superposition et réduction

Eve a maintenant à grouper chacune des lettres du message chiffré selon qu'elles aient été chiffrées par la 1ère lettre du mot clé, par la 2nde, … jusqu'à la 5ème pour le cas de notre exemple de méthode de chiffrement qui est de période 5. Cette première action est connue sous le terme *superposition*. [1]

L'on obtient donc une portion de message chiffré par la même lettre clé, en d'autre termes utilisant une méthode de chiffrement mono-alphabétique – voire un chiffre additif – dans le cas du chiffre de Vigenère. Cette passage du chiffre de Vigenère, chiffre plus ardu à déchiffrer, à un plus simple chiffre qu'est le chiffre additif, se nomme *réduction*.

L'analyse de la fréquence des lettres de la première colonne nous indique que la lettre chiffrée L est celle qui revient le plus, elle pourrait correspondre à la lettre en claire e donc. Ce qui entraine que la première lettre du mot clé utilisé est G. $7\ (G) \equiv 12(L) - 5\ (e)$.

1.12 Amélioration de la robustesse du chiffre de Vigenère

Les ouvrages que nous avons lus parlent d'amélioration conséquente de la robustesse du chiffre si utilisé à deux reprises par exemple. Sous conditions que les deux clés utilisées aient une longueur dont les valeurs sont premiers entre eux, soit $PGCD(l_1, l_2) = 1$. [1]

Le double chiffre est équivalent à un chiffre unique dont le mot clé est de longueur l_3 avec :

$$l_3 = l_1 \times l_2 \qquad (1.18)$$

Exemple : Considérons les deux mots clés *OK* et puis *CURIOUSER*, de longueurs respectivement 2 et 9. Comme ces deux derniers sont premiers entre eux, la formule (1.18) dit qu'elles (les deux clés, et par conséquent le chiffre qui en résulte) équivaut à une clé de longueur $18 = 2 \times 9$. Cette clé résultat est *RFGTDFHPGNJCXZJDTC*, le calcul étant explicité dans le **Tableau 1.01**.

Clé 1	*O*	*K*	*O*	*K*	*O*	*K*	*O*	*K*	*O*	*K*	*O*	*K*	*O*	*K*	*O*	*K*	*O*	*K*
Equivalent en nombre	15	11	15	11	15	11	15	11	15	11	15	11	15	11	15	11	15	11

Clé 2	*C*	*U*	*R*	*I*	*O*	*U*	*S*	*E*	*R*	*C*	*U*	*R*	*I*	*O*	*U*	*S*	*E*	*R*
Equivalent en nombre	3	21	18	9	15	21	19	5	18	3	21	18	9	15	21	19	5	18

Equivalent en nombre	18	6	7	20	4	6	8	16	7	14	10	3	24	26	10	4	20	3
Clé 3	*R*	*F*	*G*	*T*	*D*	*F*	*H*	*P*	*G*	*N*	*J*	*C*	*X*	*Z*	*J*	*D*	*T*	*C*

Tableau 1.01 : *Clé résultant de OK et CURIOUSER.*

1.13 Enigma, un type de machine cryptographique à rotor

La machine *Enigma* est une machine cryptographique à rotors dont la première version fut inventée vers la fin de la première guerre mondiale par un ingénieur allemand connu sous le nom d'*Arthur Scherbius*. Toujours dans cette première version, la machine possédait : [1], [2]

- 3 *rotors* qui vont être dans un premier temps traversés, de droite vers la gauche, par le courant électrique représentant la première étape de chiffrement ;

- Un *réflecteur*, qui effectue encore une fois, un chiffre par substitution. Le réflecteur subdivise l'alphabet de 26 lettres en 2 groupes de 13 lettres chacun. Fait correspondre par la suite chacune des lettres du premier groupe, à une autre dans le second groupe. Le réflecteur assure dans un premier temps que chacune des lettres en claires ne corresponde à une lettre chiffrée égale à elle (la lettre en claire) même. Et dans un second temps, de rendre réciproque le chiffre effectué par la machine Enigma ;

- Le courant électrique après avoir passé dans le réflecteur, ré-entre dans le labyrinthe des 3 rotors cités dans notre premier point, mais cette fois ci d'un ordre inversé de gauche vers la droite ;

- Dans les versions améliorées de la machine, l'on trouvait un *tableau de connexion*. Qui une n-ième fois encore, effectue une substitution. Le

tableau de connexion permet de définir manuellement le comportement du chiffre par substitution qui est appliqué. Le chiffre effectué par le tableau de connexion enveloppe (Effectue un chiffrement de la lettre en claire en premier. Passe la lettre ainsi chiffré à l'ensemble rotors, réflecteur puis rotors. Effectue le déchiffrement) tous les chiffre précédemment effectués dans nos précédents points ;

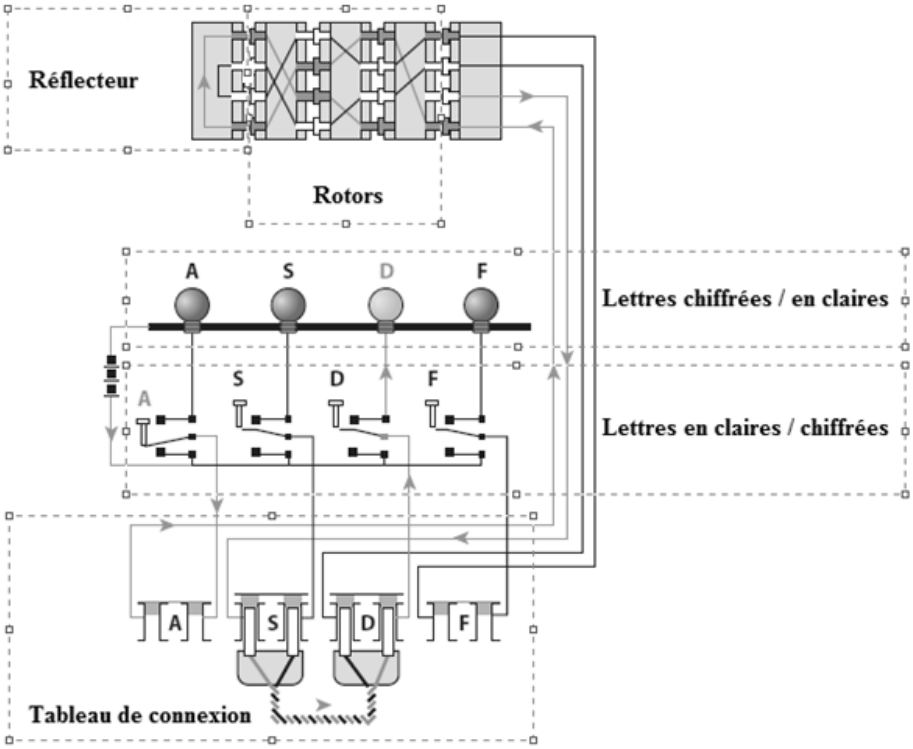

Figure 1.03 : *Le couple de (lettre en claire, lettre chiffré) ou (lettre chiffrée, lettre en claire) (A, D) traité par la machine Enigma.*

La formule du chiffrement effectué par une machine Enigma est représentée comme suit :

$$E \equiv TRMLUL^{-1}M^{-1}R^{-1}T^{-1} \qquad (1.19)$$

Où E représente le chiffrement (Encryption) de la machine Enigma dans sa globalité ;

T et T^{-1} représentent le chiffre du tableau de connexion (Table of connection ou plugboard) et son inverse ;

R et R^{-1} représentent le chiffre du rotor de droite (R pour Right) et son inverse ;

M et M^{-1} représentent le chiffre du rotor du milieu (M pour Middle) et son inverse ;

L et L^{-1} représentent le chiffre du rotor de gauche (L pour Left) et son inverse ;

U représente le chiffre symétrique du réflecteur.

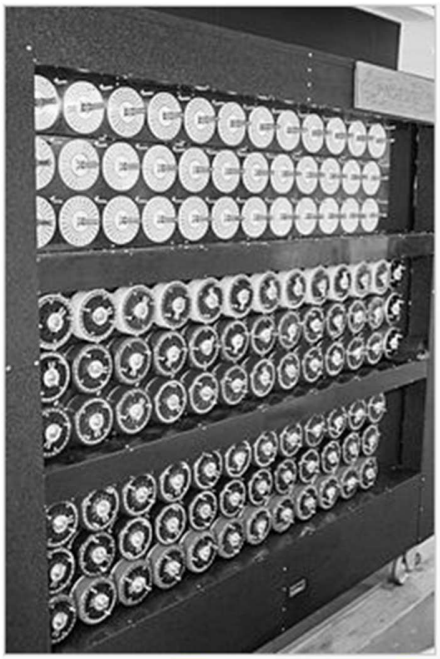

Figure 1.04 : *Une version reconstruite de la bombe qui marche réellement, exposée au musée du parc de Bletchley. La bombe possède* 3 *groupes de* 36 *rotors. Chaque rotor de la bombe simulant l'action d'un rotor de la machine Enigma.*

En termes de cryptanalyse de la machine Enigma, l'on peut avancer la machine connue sous le nom de *La bombe* et illustrée dans la **Figure 1.04**. Cette dernière est une machine électromécanique utilisée par les cryptologues britanniques afin de casser les codes allemands d'Enigma pendant la Seconde Guerre mondiale. La version la plus connue fut celle conçue par *Alan Turing* en 1939. Celle-ci (la version d'Alan Turing) tire toute fois son idée d'un même genre de machine créé par le cryptologue et mathématicien polonais *Marian Rejewski* un an plus tôt (en 1938). Pour être un an plus tard (en 1940) améliorée avec la participation cette fois ci de *Gordon Welchman*, un mathématicien anglais. [1], [4]

2 Chiffrement par transposition

Ce que nous avons vu jusqu'ici est que pour une lettre en claire donnée, ou un groupe de lettre en claires données, l'on associe une lettre chiffrée, ou groupe de lettres chiffrées. C'est ce qui a été connu sous le terme chiffrement par substitution.

Une autre manière de procéder est de laisser les lettres en claires telles qu'elles sont, mais d'uniquement chambouler leur position, pour donner vers la fin le message chiffré. Une telle méthode est connue sous le nom de chiffrement par transposition.

2.1 Méthodes de chiffrement dérivées du remplissage d'une forme géométrique

Il existe différentes manières de remplir une forme géométrique. Le colonel américain *Parker Hitt* en a fait l'objet de son ouvrage sorti durant la première guerre mondiale. [1], [3]

(a) Horizontal simple			
ABCDEF	FEDCBA	STUVWX	XWVUTS
GHIJKL	LKJIHG	MNOPQR	RQPONM
MNOPQR	RQPONM	GHIJKL	LKJIHG
STUVWX	XWVUTS	ABCDEF	FEDCBA

(c) Horizontal alterné			
ABCDEF	FEDCBA	XWVUTS	STUVWX
LKJIHG	GHIJKL	MNOPQR	RQPONM
MNOPQR	RQPONM	LKJIHG	GHIJKL
XWVUTS	STUVWX	ABCDEF	FEDCBA

(b) Vertical simple			
AEIMQU	DHLPTX	UQMIEA	XTPLHD
BFJNRV	CGKOSW	VRNJFB	WSOKGC
CGKOSW	BFJNRV	WSOKGC	VRNJFB
DHLPTX	AEIMQU	XTPLHD	UQMIEA

(d) Vertical alterné			
AHIPQX	DELMTU	XQPIHA	UTMLED
BGJORW	CFKNSV	WROJGB	VSNKFC
CFKNSV	BGJORW	VSNKFC	WROJGB
DELMTU	AHIPQX	UTMLED	XQPIHA

Tableau 2.01 : *Quelques manières de remplir une forme rectangulaire.*

2.2 Le chiffre par permutation

Le chiffre par permutation consiste à assigner une autre position à toutes les lettres en claires données du message. Pour praticité, le message est subdivisé en un nombre donné de lettres, puis la permutation est effectuée sur chaque groupe de lettres. Encore pour praticité, la longueur de subdivision du message en claire ainsi que l'ordre de permutation proviennent d'un mot que l'on pourrait désigner comme clé du chiffre. [1]

Le chiffrement est représenté sous la forme d'une matrice à 2 lignes et l colonnes, l étant la longueur du mot clé utilisé. La première ligne représente tout simplement un nombre croissant allant de 1 au nombre de lettres du mot clé (ici 4). La seconde, la succession de chiffre qui représente la permutation à effectuer : [1], [2]

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix} \qquad (2.01)$$

La première colonne de la matrice dans la formule (2.01) dicte donc que la première lettre du groupe de lettre du message chiffré est en fait la 4ème lettre du groupe de lettre correspondant dans le message en clair.

Pour obtenir la matrice de déchiffrement, échanger la première et la seconde ligne de la matrice. Ce qui donne $\begin{pmatrix} 4 & 1 & 3 & 2 \\ 1 & 2 & 3 & 4 \end{pmatrix}$. Puis trier suivant la première ligne, ce qui revient à la formule (2.02).

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix} \qquad (2.02)$$

La plage de clés (le nombre de clé possible) pour un chiffre par permutation est donnée par la valeur $l!$ (l factorielle) pour une clé de longueur l. Cette valeur est égale à $4! = 24$ pour un chiffre par permutation dont la clé est de longueur 4.

$$l! = \sum_{1 \le n \le l} n = l \times (l-1) \times \ldots \times 2 \times 1 \qquad (2.03)$$

2.3 Cryptanalyse de chiffre de type colonnaire ou à permutation. Méthode de contact de William Friedman

Nous allons accepter que l'on ait déjà pu savoir la longueur de la clé k utilisée. Il nous reste donc

à détecter, quelle colonne va en premier, suivie par laquelle.

Pour ce faire, William Friedman a proposé ce que l'on connait par méthode de contact. Cette méthode consiste à calculer la probabilité pour que les lettres de 2 colonnes données se suivent. En effet, l'on connait déjà pour la langue anglaise la probabilité d'apparition de tout digraphe donné.

Le produit de chacune de ces valeurs, ou la somme de leur logarithme donne la probabilité pour que chacune des lettres de 2 colonnes données se suivent. Une probabilité proche de 0 (soit 1 en valeur pour laquelle le logarithme n'a pas été appliqué) équivaut à deux colonnes qui a une forte probabilité de se suivre.

I	VII	Probabilité d'apparition	Logarithme - Probabilité d'apparition
O	H	0.001	-3.301
H	A	0.013	-1.886
I	T	0.010	-2.000
V	E	0.008	-2.097
R	H	0.001	-3.000
S	M	0.001	-3.301
V	E	0.008	-2.097
A	I	0.001	-3.000
H	E	0.017	-1.783
T	L	0.002	-2.824
B	B	-	
L	O	0.002	-2.699
R	H	0.001	-3.000
H	I	0.006	-2.222
L	U	0.002	-2.824
H	S	-	
L	I	0.005	-2.347
B	E	0.006	-2.260
Probabilité sur la totalité du message :			**-40.640**

I	VIII	Probabilité d'apparition	Logarithme - Probabilité d'apparition
O	K	-	
H	I	0.006	-2.222
I	U	-	
V	E	0.008	-2.097
R	U	0.002	-2.824
S	H	0.005	-2.301
V	E	0.008	-2.097
A	S	0.008	-2.097
H	L	0.001	-3.301
T	M	0.001	-3.301
B	T	0.001	-3.301
L	K	-	
R	S	0.005	-2.347
H	E	0.017	-1.783
L	C	0.002	-2.699
H	R	0.001	-3.000
L	E	0.009	-2.046
B	P	-	
Probabilité sur la totalité du message :			**-35.415**

Tableau 2.02 : *Méthode de contact appliqué aux couples de colonnes (I, VII) et (I, VIII). (I, VIII) a une plus forte probabilité de se suivre.*

3 Conclusion

Cet article nous a permis de se former une solide base sur ce qu'était la cryptographie depuis son apparition jusqu'à la fin de l'ère pré-informatique. Et en parallèle de s'initier à la cryptanalyse des différentes méthodes citées.

Rappelons que les méthodes citées dans le présent article constituent le fondement de ceux qui vont se faire en termes d'améliorations dans la discipline qu'est la cryptologie, cette fois-ci, pour l'ère de l'arrivée de l'informatique. Notre ère donc.

Mais le domaine de l'informatique étant surtout caractérisé par une capacité de faire faire des calculs complexes en un temps réduit, il est primordial pour nous de passer par des fondements mathématiques indispensable à la cryptologie. Pour ne citer que la théorie de l'information, celle des nombres ainsi que les mathématiques discrètes. Ceci fera l'objet de notre 2nd article qui s'intitulera *Quelques fondements mathématiques indispensables à la cryptologie*.

4 Bibliographie

[1]. J. Holden, *« The mathematics of secrets: cryptography from Caesar ciphers to digital encryption »*, Princeton University Press, New Jersey, Feb. 2017 ;

[2]. W. Stallings, *« Cryptography and network security, principles and practice, seventh edition »*, Pearson Education Limited, England, 2017 ;

[3]. C. Easttom, *« Modern cryptography, applied mathematics for encryption and information security »*, McGraw-Hill Education, New York, 2016 ;

[4]. F. Carter, *« The Turing bombe »*, http://www.rutherfordjournal.org/article030108.html, Jun. 2018 ;