LES FICHES DU CLUB ALKINDI juillet 2021
FICHE N° 3

LES CHIFFREMENTS PAR SUBSTITUTION

Quelques définitions pour bien se comprendre :

Clair : message d'origine à transmettre. On le désigne par le « clair »

Chiffrer, crypter, coder : transformer le message d'origine en message secret et incompréhensible pour l'ennemi ou pour n'importe qui d'autre que son destinataire.

Cryptogramme : message chiffré. On le désigne aussi par le « crypto ».

Algorithme de chiffrement : Méthode utilisée pour chiffrer. C'est une suite d'opérations à effectuer pour parvenir à chiffrer un texte. On effectue les opérations inverses pour déchiffrer.

Clef : un algorithme de chiffrement nécessite généralement une clef, qui peut être un mot, une phrase, une suite de chiffres connus seulement par l'expéditeur et le destinataire du message.

Déchiffrer : reconstituer le message clair en connaissant le mode de chiffrement et le clef.

Décrypter : trouver le message clair sans connaître le mode de chiffrement ni la clé.

1. Le chiffrement par substitution simple

C'est le système le plus courant : chaque lettre du message clair est remplacée par une autre lettre dans le message chiffré. Commençons par un peu d'Histoire :

1. 1 Le chiffre de Jules César

C'est l'un des plus simples : chaque lettre de l'alphabet est décalée de 3 lettres dans le sens opposé, ainsi :

Clair : A B C D E F G H ...
Crypto : X Y Z A B C D E ...

Le message : VENI VIDI VICI

sera chiffré par : SBKF SFAF SFZF

On peut également décaler les lettres de 3, 4, 5… positions, dans un sens ou dans un autre. Dans ce cas général où la longueur du décalage est inconnue, le déchiffrement s'effectue en testant des positions de lettres jusqu'à obtenir un message qui ait un sens.
En fait ce code était utilisé par Jules César pour sa correspondance privée. Avec ses généraux, il employait un autre procédé : il écrivait en grec.

1.2 Inversion de l'alphabet

On peut chiffrer en inversant l'ordre des lettres dans l'alphabet, comme suit :

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

Dans la Bible hébraïque, il existe un mode de cryptage appelé l'*Atbash.* Dans ce système, chaque lettre de l'alphabet est remplacée par la lettre qui occupe la même place en partant de la fin de l'alphabet, comme ci-dessus. A= Z, B=Y, etc. Le mot « atbash » est formé par la première lettre de l'alphabet hébreu, aleph, puis tav, la dernière, puis la seconde, beth, et l'avant dernière, shin.
Dans Jérémie, chapitre 25, verset 26, la ville de Babel (Babylone) est appelée Shéshakh,

2. Substitution par bigrammes : le carré de Polybe

Pour rester dans l'Antiquité, citons *le carré de Polybe* (général grec, 200-155 avant J-C,). Les lettres de l'alphabet sont placées dans un carré de 5 x5, comme suit :

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

Chaque lettre est repérée par un couple de nombres. Par exemple, S sera chiffré 43, puisqu'il se trouve ligne 4 et colonne 3. Le J est supprimé pour avoir 25 lettres.

Pour le rendre plus difficile à déchiffrer, on peut convenir d'un mot que l'on commence à écrire dans la grille, en supprimant les doublons, puis on écrit les lettre manquantes.

Exemple avec le mot Polybe :

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

L'énigme n°24 repose sur le principe d'un carré de Polybe, avec une légère variante.

3. Les méthodes classiques de substitution simple

Outre ces méthodes très anciennes et historiquement intéressantes, la méthode de chiffrement par substitution la plus classique consiste tout simplement à remplacer une lettre de l'alphabet par n'importe quelle autre lettre. Mais une même lettre du cryptogramme correspond toujours à la même lettre du message clair et réciproquement.

L'énigme n° 3 est un parfait exemple de ce mode de chiffrement. Voici la première phrase :

PT F'LWOL HAL R'WFT ZTAWMT LWHKTDT DAOL YAWM HAL L'YOTK A ET JW'GF XGOM PT XTWV ZOTF DT DAFUTK DGO-DTDT LO XGWL MKGWXTN HSWL DASOF JW'DGO

L'auteur de l'énigme a conservé les séparations entre les mots ainsi que la ponctuation, sans doute du fait du style très original, ce qui rend cette énigme un peu moins difficile à déchiffrer.

4. Comment déchiffrer ce type de texte ? L'analyse de fréquence

Au IXe siècle, un philosophe et mathématicien arabe, *AlKindi*, a mis au point une technique pour déchiffrer ce type de message : *l'analyse de fréquence.*

Dans une langue donnée, il y a des lettres qui reviennent plus fréquemment que d'autres. En français, par exemple, le E est la lettre la plus répandue, suivie par le A, puis le S, le I etc.
Sur des milliers de pages de texte, la fréquence des lettres est la suivante : 17 ,3 % de E, 8,4 % de A, 8,1 % de S , 7,3 % de I, 7,1 % de N et de T, etc. Vous trouverez partout sur Internet des tables de fréquence des lettres, et elles ne sont pas toutes d'accord entre elles sur les statistiques. Mais peu importe.

L'idée d'AlKindi est donc la suivante (en adaptant de l'arabe au français) : dans le cryptogramme, on compte le nombre de fois où l'on trouve chaque lettre. Si la lettre la plus fréquente est par exemple un M, on supposera qu'il est mis pour le E. Si la 2ème lettre la plus fréquente est le P, elle représentera sans doute le A ou le S, et ainsi de suite pour les 10 premières lettres (E, S, A, I, N, T, R, U, L, O). Bien sûr, on ne tombe pas juste au premier coup, il faut chercher un peu.

Pour continuer le décryptage, on travaille sur les *bigrammes*, c'est à dire sur les couples de lettres. Les bigrammes les plus fréquents en français sont ES, LE, DE, RE… Donc on essaye d'identifier, dans le cryptogramme, des couples de lettres qui pourraient représenter ces bigrammes en clair. On observe également les *répétitions de lettres* : SS, LL, TT … Même travail sur les trigramme : LES, le verbe être à la 3ème personne du singulier EST…

Tout ceci permet, petit à petit, de reconstituer des fragments du texte, et d'attribuer à chaque lettre du crypto une lettre du clair. C'est parfois assez long et fastidieux. Actuellement, certains sites Internet permettent de faire automatiquement ce travail.

Un très bon exemple d'analyse de fréquence est donné dans *Le Scarabée d'or*, la célèbre nouvelle d'Edgar Poe publiée dans le recueil des *Histoires extraordinaires.* L'histoire est écrite à l'origine en anglais, et la traduction de Charles Baudelaire explicite parfaitement tous les raisonnements du narrateur pour parvenir à décrypter un mystérieux parchemin.

Après l'étude de lettres, des bigrammes et des trigrammes, au fur et à mesure que l'on trouve des fragments de mots, on peut faire des hypothèses sur des *mots courants* ou sur des *mots probables* à partir d'indices que l'on possède sur le message : d'où vient-il, quel est son *environnement* ?

Un point important : pour réaliser une analyse de fréquence valable, il faut que le message soit assez long, ou bien il faut avoir plusieurs messages du même auteur. C'est par exemple le cas de l'énigme n° 3, déjà citée. Mais certaines énigmes, qui ne comportent que quelques mots, voire un seul, sont totalement réfractaires à cette méthode.

Face aux tentatives et aux réussites en matière de déchiffrement par l'analyse de fréquence, une parade a été mise au point par un cryptographe, Blaise de Vigenère, au cours du XVIe siècle :

5. Neutraliser la fréquence des lettres : le Chiffre de Vigenère (XVIe siècle)

La méthode de chiffrement inventée par Blaise de Vigenère vers les années 1580 présente quelque complications inhérentes à toute nouveauté. Il utilisait des substitutions mono-alphabétiques, confondait (volontairement) le I, le J et le Y ainsi que le V et le W. On ne s'attardera pas sur la création de cette méthode. Elle a été rapidement simplifié, et est devenue un mode de chiffrement très utilisé jusqu'au XXe siècle. Elle fait partie des bases de la cryptographie.

Le chiffre de Vigenère repose sur une idée simple : l'utilisation d'une clef. La clef est un mot ou une phrase, connue seulement de l'expéditeur et du destinataire du message. Le principe repose sur un décalage des lettres, comme dans le Chiffre de César, mais ce décalage varie à chaque lettre en fonction de la clef.

Pour chiffrer, on utilise une *table de Vigenère*, sorte de table de Pythagore, comme celle utilisée pour l'addition. *Voir le grand tableau page suivante*.

On appelle le message d'origine le « clair » , le message crypté le « crypto ». On procède comme suit : chaque lettre du message clair est repérée sur la 1ère ligne horizontale. Puis on va descendre sur la colonne verticale de cette lettre, jusqu'à la ligne horizontale de la lettre de la clé. La lettre à l' intersection de la lettre « claire » en colonne et de la lettre « clef » en ligne est la lettre cryptée.

Exemple sur le tableau de Vigenère page suivante (suivre les lettres en jaune) : la lettre du clair M avec la lettre de la clef J donnera ==> lettre cryptée V.

Exemple sur un texte : Message clair : **Les sanglots longs des violons de l'automne**
Clef : **Verlaine**
On aura le chiffrement suivant :

Clair	L	**E**	**S**		**S**	**A**	**N**	**G**	**L**	**O**	**T**	**S**		**L**	**O**	**N**	**G**	**S**		**D**	**E**	**S**
Clef	**V**	**E**	**R**		**L**	**A**	**I**	**N**	**E**	**V**	**E**	**R**		**L**	**A**	**I**	**N**	**E**		**V**	**E**	**R**
Crypto	**G**	**I**	**J**		**D**	**A**	**V**	**T**	**P**	**J**	**X**	**J**		**W**	**O**	**V**	**T**	**W**		**Y**	**I**	**J**

Clair	**V**	**I**	**O**	**L**	**O**	**N**	**S**		**D**	**E**		**L**	**A**	**U**	**T**	**O**	**M**	**N**	**E**
Clef	**L**	**A**	**I**	**N**	**E**	**V**	**E**		**R**	**L**		**A**	**I**	**N**	**E**	**V**	**E**	**R**	**L**
Crypto	**G**	**I**	**W**	**Y**	**S**	**I**	**W**		**U**	**P**		**L**	**I**	**H**	**X**	**J**	**Q**	**E**	**P**

Table de Vigenère

Lettres de la clef	Lettres du message clair																									
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

L'importance de ce type de chiffrement est capitale : chaque lettre claire est combinée avec une lettre différente de la clef. Une même lettre est donc chiffrée de différentes manières. L'analyse de la fréquence des lettres dans le message chiffré n'est plus possible.

On vient de voir que les lettres du message clair sont généralement remplacées par d'autres lettres. Mais on peut les remplacer par des chiffres (Énigme n° 1), par d'autres caractères, voire par des symboles. La fiche n° 3, qui traitera des *Codes et des dictionnaires*, en donnera quelques exemples.

Pour introduire la notion *d'algorithme*, on peut écrire que l'algorithme du chiffre de Vigenère est

Clair + Clef = Crypto

Pour déchiffrer ce type de messages, on effectue l'opération inverse à l'aide du tableau de Vigenère :

Crypto - Clef = Clair

On présente souvent le crypto par blocs de 5 lettres, ce qui complique l'analyse des mots probables dans le déchiffrement.

Ce chiffrement avec une clef est-il indéchiffrable ?

A vrai dire, non. En effet, Charles Babbage, vers 1854, a mis au point une méthode qui permet, malgré la clé, de casser ce type de chiffrement. Voici le principe de cette méthode :

Il faut d'abord essayer de trouver la longueur de la clef. Pour se faire, Babbage recherche dans le message chiffré les répétitions de 2 lettres ou 3 lettres. Il suppose que dans ces cas, la même partie de texte clair a été chiffrée avec la même partie de la clef. A partir de là, et en comptant le nombre de lettres entre chaque occurrence de répétitions de lettres identiques, il en déduit la longueur de la clef (avec plusieurs possibilités). Ayant fait une hypothèse sur la longueur de la clef, par exemple une clef de 7 lettres, il divise le message en blocs de 7 lettres et écrit ces blocs les uns en dessous des autres, formant ainsi 7 colonnes. Ainsi, dans une même colonne, les lettres ont toutes été chiffrées par la même lettre de la clef. Enfin, il effectue une analyse de fréquence sur les lettres de ces colonnes en appliquant la méthode expliquée au paragraphe 4. Toutes ces analyses supposent que le crypto soit assez long pour pouvoir être effectuées.

La méthode est assez complexe et dépasse un peu le cadre de ce « cours » qui se veut être une initiation. Si l'on veut bien la comprendre et la travailler, on trouvera aisément des livres ou des sites sur Internet consacrés à la cryptographie qui développent le sujet. (Voir la fiche n° 10).

Pour éviter cette tentative de déchiffrement en utilisant la longueur de la clef, l'idée est venue d'utiliser des clef aussi longues que le message clair, par exemple chiffrer un texte en clair par une clef qui est un texte aussi long que le message clair. Cela complique les choses pour le décrypteur, qui ne peux s'appuyer sur aucune fréquence de lettres. Dans ce cas, il faut travailler à partir de mots probables et supposés dans le texte : mots fréquents, vocabulaire usuel ou vocabulaire technique relatif au sujet du message (militaire, diplomatique…).

En combinant les lettres d'un mot clair supposé avec les lettre du cypto, on peut en déduire un fragment de clef qui a un sens (Crypto moins clair = clef).

Le carré de Vigenère est constitué de lettres, mais on peut travailler sur des chiffres. En posant A=1, B=2, C=3…, le chiffrement peut s'effectuer mathématiquement. On remplace chaque lettre du clair et de la clé par son rang dans l'alphabet, on additionne et on retire 26 si le nombre est supérieur à 26. Le résultat donne le rang de la lettre cryptée.. On peut donc travailler directement avec des nombres, ce qui est plus facile. Faites-le avec un papier et un crayon, vous verrez.

6. Le chiffre inviolable : la clé aléatoire utilisée une fois

Avec l'utilisation d'un algorithme de chiffrement (ici une simple addition) et d'une clef, une idée s'est rapidement imposée : si le clef est composée de lettres au hasard , *aléatoires* comme on dit, une lettre du clair est chiffrée avec n'importe quelle lettre de la clé. Exemple :

Clair : **B L E S S E N T M O N C O E U R D' U N E L A N G U E U R M O N O T O N E**

Clé : **H N T F S M P K D C B Z M etc. n'importe quelle lettre...**

Le résultat sera indéchiffrable, car on n'a aucune idée de la lettre employée pour chiffrer. De plus, il faut que la clef ne soit utilisée *qu'une seule fois*. Si elle est réutilisée, il y a le risque que l'analyse de deux cryptos permette le déchiffrement. On parle donc de « *clef aléatoire une fois* » (elle n'est pas belge.)

Le système est parfait, ce type de message chiffré est indéchiffrable.

Dans les années 1970 / 1980, les transmissions entre les bâtiments de la Marine nationale et entre les bâtiments et le continent étaient fondées sur ce principe. Concrètement, il y avait un téléscripteur (une grosse machine) sur laquelle défilaient deux bandes de papier, larges d'environ 2 cm. Une bande avec le clair, une bande avec la clef. Les deux bandes étaient calées au départ pour être synchronisées puis défilaient côte à côte dans le téléscripteur. La machine faisait l'addition des lettres, obtenait le crypto et le transmettait par radio au destinataire.

Les navires français étaient souvent accompagnés par des « chalutiers» soviétiques, bardés d'antennes et de radars, qui ne pêchaient pas beaucoup et suivaient tous les exercices en mer. Ils interceptaient les communications, mais ne parvenaient pas à les déchiffrer.

Cette méthode présente un inconvénient. L'échange de clefs entre l'émetteur et le destinataire est long et délicat. Le système est un peu lourd et suppose des conditions de communications des clefs en toute sécurité. Elle est encore utilisée de nos jours par certaines ambassades qui peuvent transmettre leurs clefs par la valise diplomatique.

7. Le principe de Kerckhoffs

Nous avons vu en détail le mécanisme de l'utilisation d'un algorithme de chiffrement et d'une clef, jusqu'à la méthode de la *clé aléatoire une fois.*

L'importance de la clef par rapport à l'algorithme est un principe de base de la cryptographie. Ce principe a été fixé en 1883 par le linguiste hollandais Auguste Kerckhoff Van Nieuwenhoff dans son traité « *La cryptographie militaire* ». Principe de Kerckhoffs :

« La sécurité d'un système de chiffrement ne doit reposer que sur le secret de la clef et non pas sur le secret de l'algorithme de chiffrement, qui peut être connu de l'ennemi. »

Ce principe est fondamental. Nous le trouverons constamment tout au long de ces fiches.

*****
***
*