LES FICHES DU CLUB ALKINDI
FICHE N° 3

LES CODES ET LES DICTIONNAIRES

1. Les codes qui ne sont pas secrets

Il existe des codes qui ne sont pas secrets et qui sont utilisés par tous. Ces codes sont souvent des moyens de communication ou des systèmes de numération. Ce sont des méthodes pour coder, mais ce n'est pas à proprement parler de la cryptographie puisque tout le monde les connaît et les utilise. Citons quelques grands classiques :

- ***Le code Morse international***, ou alphabet Morse international, dans lequel chaque lettre est représentée par un système de points et de traits. Il permet de transmettre des textes à l'aide de séries d'impulsions courtes et longues. Il est utilisé entre autres pour des émissions à caractère automatique : radiobalises en aviation, indicatif d'appel des stations maritimes, signalisation maritime par des transpondeurs radar.

-***L'alphabet phonétique de l'OTAN*** est un alphabet radio international qui a été normalisé par l'Union internationale des télécommunications. On connaît les célèbres « Alpha, Bravo, Charlie, Delta... » que l'on entend dans les films. Chaque mot désigne une lettre et ces mots ont été spécialement choisis pour qu'ils soient biens distincts les uns des autres quand on les prononce à la radio et qu'il n'y ait aucune ambiguïté sur la lettre considérée.

- ***Le système de numération binaire***
C'est un système de numération, c'est à dire une façon de compter. Mais au lieu d'utiliser le système décimal, c'est à dire les 10 chiffres de 0 à 9, on n'utilise que le 0 et le 1. Les tables de correspondance avec le système décimal se trouvent partout.

Ces chiffres 0 et 1, chiffres de la numération binaire positionnelle, sont appelés communément « *bits* », abréviation de l'anglais *binary digit*, soit « chiffre binaire ». Les bits sont le système de calcul utilisés par les ordinateurs : les 0 et les 1 sont déterminés par la présence ou non d'un courant électrique.

- ***Le système de numération hexadécimal***
est un système de numération en base 16. Il utilise 16 symboles, les chiffres arabes pour les 10 premiers chiffres (de 0 à 9), , puis les lettres A à F pour les 5 suivants. Le nombre 15 (décimal) s'écrit donc F, le nombre 16 s'écrit 10, c'est à dire (1 x 16) + 0 unités, le nombre 17 s'écrit 11, soit (1 x 16) + 1 unité, le nombre 18 s'écrit 12, soit (1 x 16) + 2 unités et ainsi de suite.

Le système de numération hexadécimal est très utilisé en informatique car il permet une conversion sans aucun calcul avec le système binaire, système employé par les ordinateurs, du fait que 16 est une puissance de 2. Un chiffre en base 16 correspond exactement à 4 chiffres en base 2. Recherchez sur Internet si le sujet vous intéresse et si vous souhaitez approfondir ces notions.

Pour visualiser concrètement, le tableau page suivante met en parallèle les premiers nombres écrits dans chaque système.

Décimal	Héxadécimal	Binaire
0	0	0
1	1	1
2	2	10
3	3	11
4	4	100
5	5	101
6	6	110
7	7	111
8	8	1000
9	9	1001
10	A	1010
11	B	1011
12	C	1100
13	D	1101
14	E	1110
15	F	1111
16	10	10000
17	11	10001
18	12	10010

- Le code ASCII (American Standard Code for Information Interchange)

La mémoire d'un ordinateur conserve toutes les données sous forme numérique. Il n'existe pas de méthode pour stocker directement les caractères. Chaque caractère du clavier possède donc son équivalent en code numérique : c'est le code ASCII.

Le code ASCII permet de représenter les chiffre de 0 à 9, les lettres majuscules et minuscules ainsi que des symboles comme %, *, #, § , les symboles de ponctuation, l'espace, en fait tous les chiffres, lettres et symboles que l'on trouve sur un clavier d'ordinateur. Ce code comporte 128 nombres représentés par 7 bits, de 0 à 127 (7 bit permettent d'écrire les nombres de 0 à 127, puisqu'en numération binaire 127 = 1111111, soit 7 caractères).

L'ASCII suffit pour représenter les textes en anglais, mais il est trop limité pour les autres langues, dont le français et ses lettres avec des accents. Les limitations du jeu de caractères ASCII sont encore sensibles actuellement, par exemple dans le choix restreint de caractères généralement offerts pour composer une adresse électronique.

Pour être clair, *voir le tableau page suivante*. On y trouve les 128 caractères du code ASCII représentés par un nombre en système décimal ou hexadécimal. La 1re colonne est le codage en nombres décimaux, la 2ème le codage en hexadécimal, et la 3ème marquée « Char » (caractère) est le symbole qui est à représenter.

Par exemple, le chiffre 4 est codé 52 (en décimal), < est codé 60, @ est codé 64, P majuscule est codé 80, k minuscule est codé 107 etc.

Plusieurs énigmes de ce site utilisent un codage ASCII.

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	|
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

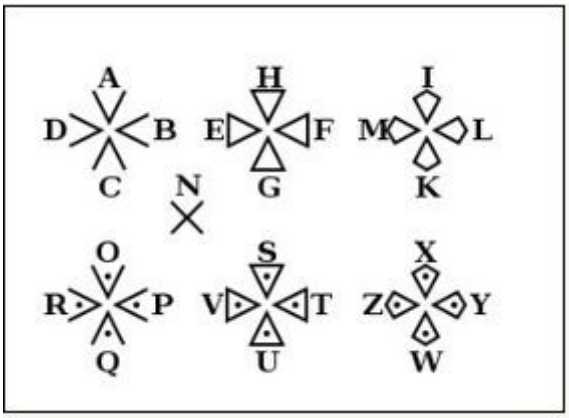
- Comme il a été expliqué en introduction, tous ces types de codage ne sont pas de la cryptographie à proprement parler. Cependant, de nombreuses énigmes de ce site comportent des codages en numérotation binaire, ASCII ou autres. *Il est important de se familiariser avec ceux-ci. Ils sont d'un usage courant en informatique et sont présents dans certaines énigmes du concours Alkindi.* Leur pratique permet de mieux maîtriser d'autres types de codages réellement cryptographiques et de faire marcher ses neurones.

2. Les « vrais » codes secrets

Une rapide promenade dans l'Histoire permettra de juger de l'imagination des hommes pour créer des codes secrets. Citons par exemple :

- Le chiffre des Templiers

Le codage des lettres est le suivant :



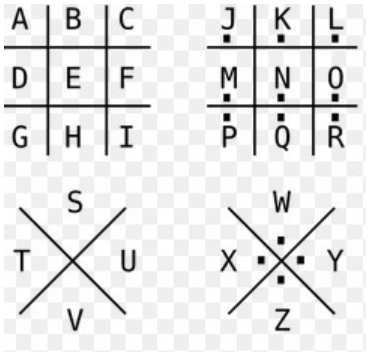
Ces symboles on été créés à partir de la croix des templiers

elle-même issue probablement de la croix de l'Ordre de Saint-Jean de Jérusalem.

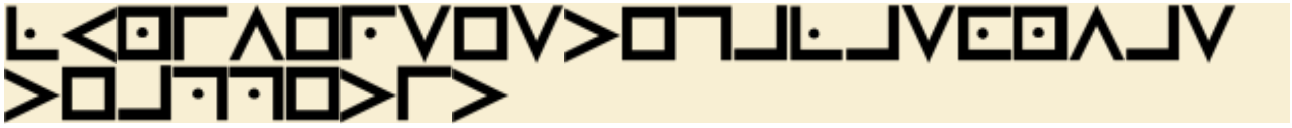
Ainsi le texte : **« Pour l'enfant, amoureux de cartes et d'estampes »** sera chiffré :

- Le chiffre des francs-maçons

Ce mode de chiffrement est inspiré du chiffre des Templiers. Il se présente ainsi



Le texte **« L'univers est égal à son vaste appétit »** deviendra



Les francs-maçons britannique surnommaient ce chiffre Pig-Ben, l'« enclos aux cochons » en anglais, sans doute du fait des lignes régulières et géométriques qui le composait. Il y a également un jeu de mot avec Big Ben, la grosse cloche de l'horloge du palais de Westminster à Londres.

- Le chiffre de Marie Stuart

La fin de la vie de la reine Marie Stuart a été évoquée rapidement dans la fiche n° 1, *A quoi sert la cryptographie ?* Dans le cas de cette malheureuse reine, son chiffre, ou plutôt une mauvaise utilisation de celui-ci, l'a conduite à la mort.
Le chiffre de Marie Stuart peut aussi être appelé une *nomenclature*. Il est en effet constitué d'un alphabet chiffré, de symboles indiquant que la lettre suivante est nulle ou doublée, ainsi que de symboles exprimant les mots les plus courants. Le voici :

Tous ces codes sont des chiffrements par substitution. Ils ne résistent pas à une attaque par l'analyse de fréquence, surtout si les émetteurs des messages commettent les erreurs classiques qui font le bonheur des décrypteurs : nombreux messages codé avec le même système de chiffrement, erreurs de chiffrement, présence de mots probables répétés en début ou en fin de messages. Nous ne reviendrons pas sur l'importance d'une clef, le lecteur se lasserait...

Certaines énigmes de ce site sont constituées à partir de ces types de code.

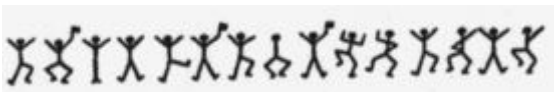
- Les polices de caractères

Il existe également des énigmes qui sont codées tout simplement par une *police de caractères* que l'on trouve dans Word. Ainsi par exemple l'étrange message :

✌♒·🖉 ✈◆♏ ●♏ ❍❑■♎♏ ♏♦♦ ♑❒♋■♎ → ●♋ ♍●♋❒♦🡅 ♎♏♦ ●♋❍❑♏♦·🖉

est écrit en police Wingdings et signifie : « **Ah! Que le monde est grand à la clarté des lampes** !»

Pour conclure sur les codes, un exemple amusant de codage par symboles est donné par Arthur Conan Doyle dans *Les hommes dansants*, une aventure de Sherlock Holmes. Le célèbre détective fait une très belle démonstration d'analyse de fréquence (difficile en anglais) et de recherche de mots probables pour décrypter des messages constitués par des petits dessins :



3. Les dictionnaires

Les dictionnaires chiffrés reposent sur le principe de faire correspondre un mot, et non pas seulement une lettre, à un nombre.

Continuons notre approche historique avec :

- Le Grand Chiffre de Louis XIV

C'était un dictionnaire où les mots courants étaient codés. Les mots rares étaient découpés en syllabes. Pour résister à une analyse de fréquence, une même syllabe pouvait être codée par plusieurs nombres différents. C'était donc beaucoup plus subtil qu'un simple chiffrement par substitution. Il comportait au total 587 nombres.

Ce chiffre fut élaboré par le célèbre cryptologue Antoine Rossignol qui avait été au service de Richelieu, Louis XIII et Mazarin. Son nom est resté dans l'Histoire puisque dans le langage courant un *rossignol* désigne un outil qui permet de forcer les serrures. Ce chiffre tomba en désuétude après la mort d'Antoine Rossignol et de son fils, et rapidement plus personne ne sut l'employer. De ce fait, il résista aux briseurs de code pendant 300 ans ! Jusqu'à la fin du XIXe siècle, personne n'était capable de le lire.

En 1890, une nouvelle correspondance de Louis XIV utilisant ce Grand Chiffre fut découverte par un historien. Il la confia à Étienne Bazeries, un spécialiste du service cryptographique de l'armée. Celui-ci mit trois ans à le décrypter entièrement !

Ce déchiffrement permit entre autre la découverte de l'identité d'un personnage rendu célèbre par Alexandre Dumas dans son roman *Le Vicomte de Bragelonne :* l'homme au Masque de fer. Ainsi un nom fut mit sur ce mystérieux personnage : c'était le général de Bulonde. Mais est-ce vraiment la vérité ? Ces lettres cryptées étaient peut-être destinées à être déchiffrées pour orienter les générations suivantes sur une fausse piste et cacher la véritable identité de l'homme au Masque de fer... En tout cas certains l'ont imaginé. Nous sommes dans le monde du secret…

- Le dictionnaire de Sittler

En France, la loi du 13 juin 1866 sur les usages commerciaux autorisa le chiffrement des dépêches privées par télégrammes. A cette époque, tout le monde correspondait par télégrammes envoyés par la Poste, un peu comme les SMS aujourd'hui. Aussitôt, de nombreux codes basés sur des dictionnaires virent le jour.

Un des premiers codes commerciaux fut le code Sittler de 1868. Le dictionnaire fonctionne ainsi : les mots et les expressions courantes sont rangés dans l'ordre alphabétique sur les 100 lignes d'une page. Le dictionnaire comporte 100 pages. L'utilisateur détermine lui-même la numérotation de chaque page de son dictionnaire, dans n'importe quel ordre.
Chaque mot du message est chiffré par un nombre qui commence par le numéro de la page, puis le numéro de la ligne où se trouve le mot. Bien entendu le destinataire des messages possède un dictionnaire identique pour déchiffrer. On peut ajouter un codage supplémentaire au chiffre initial. Cela s'appelle un surchiffrement.
Avec 100 mots par page et un dictionnaire de 100 pages, on peut coder 10 000 mots, ce qui est largement suffisant pour une utilisation courante.

Ce dictionnaire fut très utilisé entre 1890 et 1920 par des particuliers, des entreprises ou par l'État.

Ce système de codage présentait des faiblesses. Mais tout le monde l'utilisait et en particulier les autorités diplomatiques et militaires italiennes et allemandes.

En août 1914, un croiseur allemand, le Magdeburg, s'est échoué près d'une île en mer Baltique. Il fallut évacuer le navire et un officier prit les livres de codes pour les jeter au fond la mer. Mais les bâtiments russes tirèrent au canon sur le navire allemand et firent de nombreuses victimes. Ils repêchèrent l'officier, mort, qui tenait les livres de codes serrés dans ses bras.

Les codes furent donnés aux Britanniques qui les étudièrent et jusqu'en 1916, ceux-ci décryptèrent tous les messages de la Marine allemande. Cette histoire est un peu comparable à celle d'Énigma pendant la Seconde Guerre mondiale.

4. Les codes mystérieux et non-déchiffrés

Il existe des codes célèbres qui n'ont jamais été décryptés, comme le manuscrit de Voynich, le disque de Phaistos ou les chiffres de Beale, un étrange personnage qui a caché un fabuleux trésor. Mais ces histoires appartiennent aux mythes de la cryptographie et sortent largement du cadre d'une initiation. Vous les trouverez sur Internet si elles vous intéressent. Mais de toute façon, on ne les proposera jamais comme énigmes sur le site de Club Akindi !

****
*****
***
*