Université Mohammed Khider de Biskra
Faculté des Sciences et de la Technologie
Département Génie électrique,
Filière Electronique

Master 1 – Système Embarqué

2019-2020

Codage de l'information et sécurité

Saâdia MEDOUAKH

Chapitre 1: introduction de la théorie d'informations

1) Notion de base

1.1) Définition d'information

Déf 1 : L'**information** peut être de n'importe quel type pourvu qu'on puisse en donner une représentation numérique : textes, images, sons, vidéos par exemple. La transmission de ces types de données est omniprésente dans la technologie, et spécialement dans les télécommunications.

Déf 2 : L'**information** se présente sous trois formes : les données, les connaissances et les messages. On a l'habitude de désigner par « système d'information » l'ensemble des moyens techniques et humains qui permet de stocker, de traiter ou de transmettre l'information.

Déf 3 : Une **information** est un message, que l'on désire communiquer. Ce message utilise des **symboles** pris dans un « **dictionnaire** » pour véhiculer les données qu'il contient.

Exemples de messages : un drone (voir fiche système S4) communiquant avec le poste de pilotage basé dans une station au sol échange avec celle-ci des informations (le drone reçoit des commandes et renvoie des observations et des mesures). Le pilotage d'un drone nécessite l'échange d'une grande quantité d'information.

Déf 4 : Information : Toute donnée, réelle ou abstraite, manipulable par la pensée humaine.

Toute **information** est représentable par une suite (éventuellement infinie) de 0 et de 1 (appelés ***bits*** = contraction de ***binary elements***)

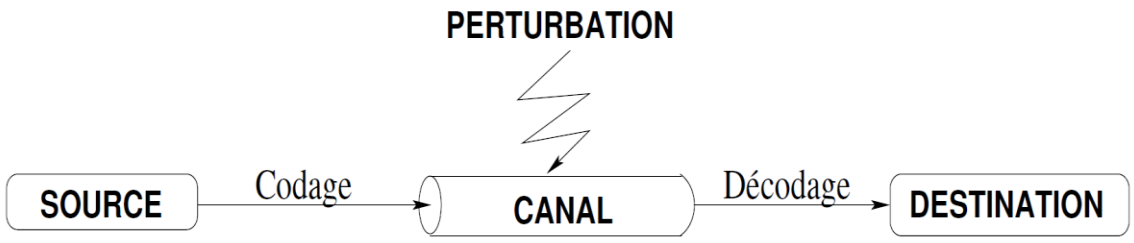
1.2) Codage de l'information

Il est très rare d'avoir un codage canonique pour une information, d'où la nécessite de norme, pour permettre l'échange de données.

Comme plusieurs codages sont possibles pour une même information, on peut choisir le code le mieux adapté à une utilisation donnée

Exemples de qualités souhaitables :

- Vitesse d'encodage et/ou de décodage
- Compacité (Compression de données)
- Robustesse (Résistance aux pertes partielles)
- Sécurité (Cryptographie)



1.3) Sécurité de l'information

Le problème de sécurité se pose dans tout système informatique. Dans un système distribué, les ressources doivent être protégées contre des utilisations abusives et malveillantes. En particulier, le problème de piratage des données sur le réseau de communication. En ces raisons, il est préférable d'utiliser des périphériques ou logiciels licenciés. Outre, les connexions doivent être sécurisées par authentification avec les éléments distants ainsi que les messages circulant sur ce réseau doivent être cryptés en vue d'éviter des conséquences graves.

Le concept de sécurité des systèmes d'information recouvre un ensemble de méthodes, techniques et outils chargés de protéger les ressources d'un système d'information afin d'assurer :

- *la disponibilité des services* : les services (ordinateurs, réseaux, périphériques, applications...) et les informations (données, fichiers...) doivent être accessibles aux personnes autorisées quand elles en ont besoin ;
- *la confidentialité des informations* : les informations n'appartiennent pas à tout le monde ; seuls peuvent y accéder ceux qui en ont le droit ;
- *l'intégrité des systèmes :* les services et les informations (fichiers, messages...) ne peuvent être modifiés que par les personnes autorisées (administrateurs, propriétaires...)
- *la non-répudiation de l'information* est la garantie qu'aucun des correspondants ne pourra nier la transaction.
- *L'authentification* : L'identification des utilisateurs est fondamentale pour gérer les accès aux espaces de travail pertinents et maintenir la confiance dans les relations d'échange

2) Théorie de l'information :

Théorie de l'information <=> Processus de transmission numérique

- La communication d'une information commence par sa formulation par un émetteur, se poursuit par un transit via un canal, et se termine par la reconstitution du message par le destinataire.
- L'information doit être reçue par son destinataire dans son intégralité, en sécurité, et le plus rapidement possible.
  - Efficacité de la transmission : compression des données
  - Intégrité du message : correction d'erreurs
  - Sécurité de l'information : cryptage, authentification

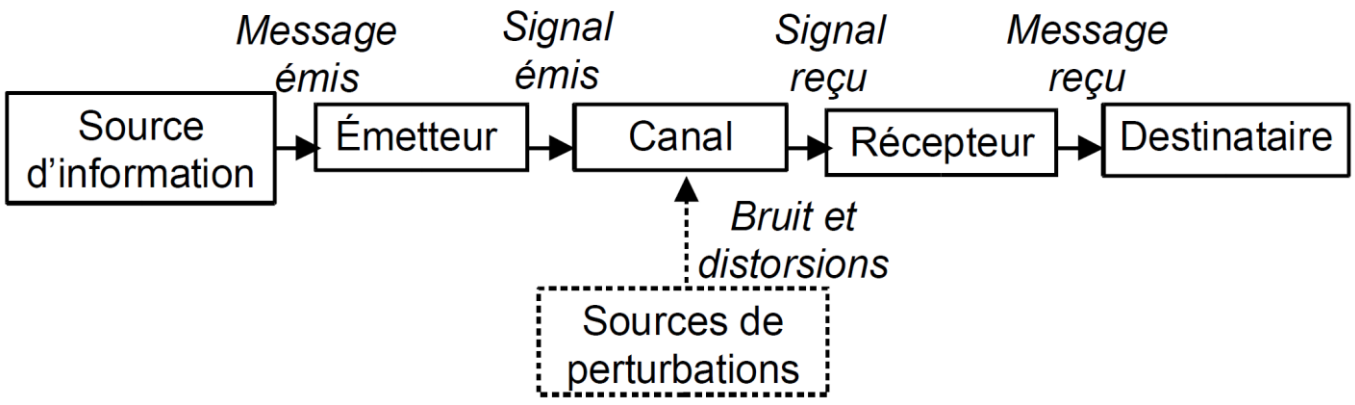


Figure – Schéma fondamental du codage

- La théorie de l'information a été créée par Claude Shannon en 1948. Il s'agit d'une théorie mathématique qui décrit les aspects les plus fondamentaux des systèmes de communication. Elle consiste en l'élaboration et l'étude de modèles pour la source et le canal qui utilisent différents outils comme les probabilités et les automates finis.
- Pour simplifier l'étude des systèmes de communication est séparée en deux parties :

  les codeurs de sources et *les codeurs de canaux*.

  – **Le but du codeur de source** est de représenter la sortie de la source en une séquence binaire, et cela de façon la plus économique possible.

  Le codage de source est une compression des données (augmenter la compacité des signaux (sans ou avec distorsion)) → Eliminer la redondance inutile.

  – **Le but du codeur de canal** et de son décodeur est de reproduire le plus fidèlement possible cette séquence binaire malgré le passage à travers le canal bruité.

  Le codage de canal est une accroître la sécurité de la transmission en présence de bruit →Ajouter de la redondance pour la détection, voire la correction, de principale erreurs.

- Alors l'optimisation d'une chaîne de transmission peut se faire en optimisant séparément les codeurs de source et de canal.

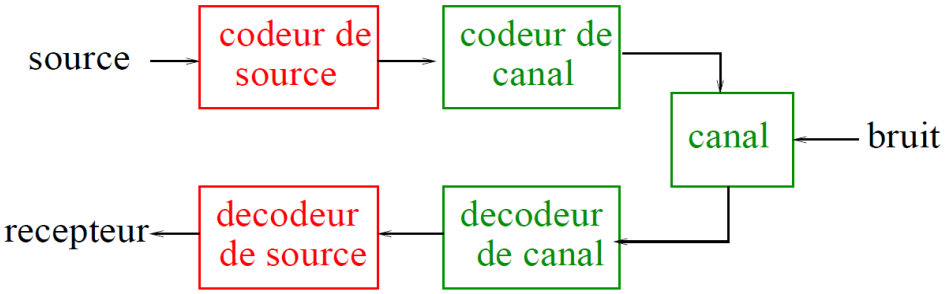


Figure : Codeur de source et codeur de canal

3) Représentation des nombres

3.1) Numération de position

Exemple 1 :
Calculons l'écriture binaire de l'entier n = 2395. Pour cela procédons par divisions euclidiennes successives jusqu'à obtenir un quotient plus petit que b = 2 (autrement dit 1).
2395 = 2 x1197 + 1
1197 = 2 x 598 + 1
598 = 2 x 299 + 0
299 = 2 x 149 + 1
149 = 2 x 74 + 1
74 = 2 x 37 + 0
37 = 2 x 18 + 1
18 = 2 x 9 + 0
9 = 2 x 4 + 1
4 = 2 x 2 + 0
2 = 2 x 1 + 0

L'écriture binaire du nombre n = 2395 est donc 100101011011_2:

Exemple 2 :

Calculons maintenant l'écriture du même nombre en base b = 8 par divisions euclidiennes successives jusqu'à obtenir un quotient inférieur à 8.

$$2395 = 8 \times 299 + 3$$
$$299 = 8 \times 37 + 3$$
$$37 = 8 \times 4 + 5$$

L'écriture octale du nombre n = 2395 est donc 4533_8:

Exemple 3 :
Calculons enfin l'écriture du même nombre en base b = 16 par divisions euclidiennes successives jusqu'à obtenir un quotient inférieur à 16.

$$2395 = 16 \times 149 + 11$$
$$149 = 16 \times 9 + 5$$

L'écriture octale du nombre n = 2395 est donc $95B_{16}$:

(Notons l'emploi du chiffre B pour représenter l'entier 11.)

3.2) Représentation des nombres en informatique

Bases courantes en informatique

binaire, octal, hexadécimal

octal = représentation par paquets de 3 bits (dans les années 1960, ordinateurs travaillant avec des registres de 12 ou 18 bits)

hexadécimal = représentation par paquets de 4 bits (les processeurs actuels travaillent avec des registres de 8, 16, 32 ou 64 bits).

Déc.	Hexa.	Octal	Bin.
00	0	00	0000
01	1	01	0001
02	2	02	0010
03	3	03	0011
04	4	04	0100
05	5	05	0101
06	6	06	0110
07	7	07	0111
08	8	10	1000
09	9	11	1001
10	A	12	1010
11	B	13	1011
12	C	14	1100
13	D	15	1101
14	E	16	1110
15	F	17	1111

Table. Écriture des entiers de 0 à 15 en décimal, hexadécimal, octal et binaire

4) Rappels de la théorie des probabilités discrètes

4.1) Evènements et mesure de probabilité:

Un *évènement* est le résultat possible d'une expérience aléatoire. Par exemple, si l'expérience est un jet de dé à six faces, l'obtention du nombre 6 est un évènement. Les opérateurs sur les ensembles (U, ∩, /) sont utilisés pour les évènements (ils signifient *ou, et, sauf*).

On note Ω l'ensemble de tous les évènements possibles pour une expérience donnée et est appelé ***espace des épreuves*** (espace probabilisé).

Une mesure de probabilité *P* est une application d´définie sur Ω, à valeur dans [0, 1], qui vérifie :

1. $P(\Omega) = 1$ et $P(\varnothing) = 0$;

2. quels que soient *A, B* des évènements disjoints ($A \cap B = \varnothing$), $P(A \cup B) = P(A) + P(B)$.

4.2) Espace probabilisé discret

Si l'ensemble des évènements est un ensemble discret ou fini, on parle de ***probabilité discrète***.
Par exemple, si l'expérience aléatoire est le jet d'un dé à six faces.

- L'ensemble des évènements est *{*1, 2, 3, 4, 5, 6*}*, et la probabilité d'occurrence de chacun 1⁄6.
- L'ensemble des valeurs prises par la fonction de probabilité est la *distribution des probabilités*, ou *loi de probabilité.*
- Une distribution est dite *uniforme* si tous les évènements ont une même probabilité d'occurrence.

4.3) Source discrète ⇔ variable aléatoire discrète

Choisissant un signe dans un alphabet de taille finie : $S=\{s_1, s_2, s_3, \ldots, s_k\}$ espace des épreuves
avec les probabilité qui $P(S= s_i) = p_i$ qui satisfont la condition : $\sum_{i=1}^{K} p_i = 1$
$P(s_i)$ est (lois de probabilité) la donnée des probabilités de chacune des résultats possibles.
Notations :

- $S=s_i$ – Événement aléatoire A_i
- $P(S= s_i) = A_i$ – *Proba.* de l'événement aléatoire A_i

4.4) Espace probabilisé joint

Pour modéliser un canal discret, nous considérons l'espace A x B produit des deux ensembles $A = \{a_1, a_2, \ldots, a_n\}$ et $B = \{b_1, b_2, \ldots, b_m\}$. Le produit est formé des couples (a, b) avec a dans A et b dans B.

A	B
a_i	b_j

L'espace de probabilité joint est noté AB, et P_{AB} appelée loi de probabilité jointe de A et B.

La probabilité $P_{AB}(a, b)$ est la probabilité d'avoir simultanément a en entrée et b en sortie on définit une loi de probabilité

$$P_A(a) = \sum_{b \in B} P_{AB}(a, b)$$

A		B
a_1	canal	b_1
a_2		b_2
⋮		⋮
a_n		b_m

La probabilité d'un évènement etont égale à la somme des probabilites des issues réalisant cet évènement, la proba de l' évènement $A = a_k$ est donc

$$P_A(a_k) = \sum_{j=1}^{J} P_{AB}(a_k, b_j)$$

De même la probabilité de l'évènement $B = b_j$

$$P_B(b_j) = \sum_{k=1}^{K} P_{AB}(a_k, b_j)$$

Les deux lois de probabilité P_A et P_A appelées ***lois marginales***

4.5) Probabilité conditionnelle

On dit que deux évènements sont *indépendants* si $P(A \cap B) = P(A).P(B)$. On appelle *probabilité conditionnelle* de l'évènement *A* par rapport à l'évènement *B*, la probabilité que *A* se produise, sachant que *B* s'est déjà produit. Elle est notée $P(A|B)$ et définie par :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

4.6) Incertitude et information propre :

La façon la plus appropriée de décrire un système de communication est d'en donner un modèle probabiliste. En suivant ce modèle, fournir une information à une utilisation consiste à choisir un évènement parmi plusieurs consiste donc à lever une incertitude sur l'issue d'une expérience aléatoire.

La variation de l'incertitude est une fonction décroissante de la probabilité. On a $I(a)$ l'incertitude sur a , encore appelée information propre de a :

$$I(a) = -\log_2 p(a)$$

L'information propre de l'évènement $A = a_k$ est définie par :

$$I(a_k) = -\log_2 (a_k) >= 0$$

$I(a_k)$: la quantité d'information fournée par la réalisation d'un évènement.

Chapitre 2: Codage de l'information

Les informations traitées par les ordinateurs sont de différentes natures : du texte, des nombres, des sons, des images, des vidéos, des instructions...

1) Types d'informations

Une information numérique peut être de deux types :

– une instruction, qui représente une opération réalisée par un organe de calcul (un microprocesseur par exemple) ;

– une donnée, sur laquelle des instructions pourront être faites.

Parmi les données qui ne seront pas traitées de la même manière par les systèmes informatiques, on peut citer :

– des données non numériques (caractère alphanumérique) ;

– des données numériques :

– entiers naturels (0 ; 1 ; 315 ...)

– entiers relatifs (-1578 ; -15 ; -1 ...)

– réels (3.1415 ; 4587.598 ...)

2) Représentation de l'information

Les informations numériques sont transmises par des signaux électriques (une tension électrique, chaque symbole étant représenté par une valeur, par exemple 0 Volt et 5 Volts). Afin d'avoir un langage universel, ces données sont représentées sous forme binaire, c'est à dire une suite de 0 et de 1.
Donc l'information élémentaire est appelé BIT (Binary digIT: chiffre binaire).

3) Codage de l'information

Le codage est un processus nécessaire à l'être humain pour communiquer. On peut définir un code comme un ensemble de symboles (alphabet d'une langue par exemple) représentant des informations utiles.

Le codage de l'information permet d'établir une correspondance qui permet sans ambiguïté de passer d'une représentation (dite externe) d'une information à une autre représentation (dite interne : sous forme binaire) de la même information, suivant ***un ensemble de règle précise.***

- En informatique, ces symboles se résument aux deux objets que sont le « 0 » et le « 1 ». Donc, dans ce domaine, toutes les informations sont représentées sous la forme de configurations binaires.

Exemple :

***** Le nombre 35 : 35 est la représentation externe du nombre trente cinq

* La représentation interne de 35 sera une suite de 0 et 1 (100011)

- En informatique, le codage de l'information s'effectue principalement en trois étapes :
  - L'information sera exprimée par une suite de nombres (Numérisation)
  - Chaque nombre est codé sous forme binaire (suite de 0 et 1)
  - Chaque élément binaire est représenté par un état physique

3.1) Système de numération

Système de numération décrit la façon avec laquelle les nombres sont représentés.

- Un système de numération est défini par : Un alphabet ***A*** : ensemble de symboles ou chiffres,
- Des règles d'écritures des nombres : Juxtaposition de symboles

3.1.1) Numération Romaine

système romain	I	V	X	L	C	D	M
valeur décimal	1	5	10	50	100	500	1000

- Lorsqu'un symbole est placé à la droite d'un symbole plus fort que lui, sa valeur s'ajoute : CCLXXI → 271
- Lorsqu'un symbole est placé à la gauche d'un symbole plus fort que lui, on retranche sa valeur : CCXLIII → 243
- On ne place jamais 4 symboles identique à la suite : 9 s'écrit IX et non VIIII
- La plus grand nombre exprimable est : 3999 ((MMMCMXCIX))
- Système inadapté au calcul

3.1.2) Numération décimale :

- C'est le système de numération le plus pratiqué actuellement.
- L'alphabet est composé de dix chiffres : ***A*** = {0,1,2,3,4,5,6,7,8,9}
- Le nombre 10 est la base de cette numération
- C'est un système positionnel. Chaque position possède un poids.
- Par exemple, le nombre 4134 s'écrit comme :

$$4134 = 4 \times 10^3 + 1 \times 10^2 + 3 \times 10^1 + 4 \times 10^0$$

3.1.3) Système de numération positionnel pondéré à base b

❖ Un système de numérotation positionnel pondéré à base **b** est défini sur un alphabet de b chiffres : ***A*** = $\{c_0, c_1, \ldots, c_{b-1}\}$ avec $0 \leq c_i < b$

❖ Soit $N = a_{n-1}\, a_{n-2} \ldots\ldots a_1\, a_{0\,(b)}$: représentation en base **b** sur **n** chiffres

- a_i : est un chiffre de l'alphabet de poids **i** (position **i**).
- a_0 : chiffre de poids 0 appelé le chiffre de poids faible
- a_{n-1} : chiffre de poids n-1 appelé le chiffre de poids fort

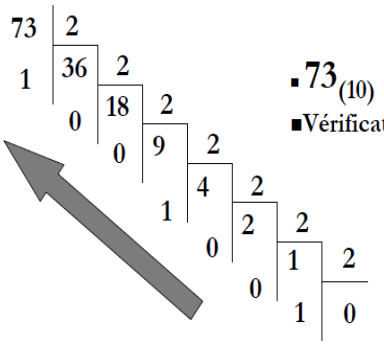
❖ La valeur de N en base 10 est donnée par:

❖

$$\mathbf{N = a_{n-1}.b^{n-1} + a_{n-2}.b^{n-2} + \ldots + a_0.b^0{}_{(10)}} = \sum_{i=0}^{n-1} a_i\, b^i$$

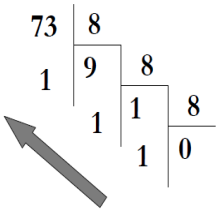
Bases de numération (Binaire, Octale et Hexadécimale)

- Système *binaire* (b=2) utilise deux chiffres : {0,1}
  - C'est avec ce système que fonctionnent les ordinateurs



▪ $73_{(10)} = 1001001_{(2)}$
▪Vérification

- Système *Octale* (b=8) utilise huit chiffres :{0,1,2,3,4,5,6,7}
  - Utilisé il y a un certain temps en Informatique.
  - Elle permet de coder 3 bits par un seul symbole.

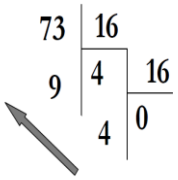


▪ $73_{(10)} = 111_{(8)}$
▪Vérification

- Système *Hexadécimale* (b=16) utilise 16 chiffres :

{0,1,2,3,4,5,6,7,8,9,A=$10_{(10)}$,B=$11_{(10)}$,C=$12_{(10)}$,D=$13_{(10)}$,E=$14_{(10)}$,F=$15_{(10)}$}

  - Cette base est très utilisée dans le monde de la micro-informatique.
  - Elle permet de coder 4 bits par un seul symbole.



▪ $73_{(10)} = 49_{(16)}$
▪Vérification

3.2) Codage des nombres entiers naturels

►Un des rares cas où il existe code canonique

►Principe : Convertir la valeur entière en base 2

►Pb : Stockage varie en fonction de la valeur

- 1 octet (byte) = 0 à 255 = 0 à 2^8 - 1
- 2 octets (word ou short) = 0 à 65 535 = 2^{16} - 1
- 4 octets (int) = 0 à 4 294 967 295 = 2^{32} - 1
- 8 octets (long) = 0 à 18 446 744 073 709 551 615 (18 milliards de milliards)

3.3) Codage des nombres entiers relatifs

►Il n'existe plus de code canonique

- Comment coder le signe, les intervalles < 0 et > 0 ?

►Principe : Intervalle symétrique + bit de signe

- 1 octet (byte) = -128 à 127 = -2^7 à 2^7 - 1
- 2 octets (short) = -32 768 à 32 767 = -2^{15} à 2^{15} - 1
- 4 octets (int) = -2^{31} à 2^{31} - 1
- 8 octets (long) = -2^{63} à 2^{63} – 1

3.4) Codage des nombres réels

Deux modes de codages principaux :

1) Codage en virgule fixe

● Principe : Chaque réel est encodé par 2 entiers, la partie entière et la partie décimale séparés par une virgule. Exemple : 54,25(10) ; 10,001(2) ; A1,F0B(16)

● Inconvénient : La taille de l'intervalle des nombres représentable est très petit

2) Codage en virgule flottante

● Principe : Basé sur notation scientifique des réels

➢ **Notation scientifique en base 10 :**

● Exemple : $123.456 = 1.23456 \times 10^{2}$

● Plus généralement : $N = (-1)^{S} \times M \times 10^{E}$

Avec: $S \in \{0,1\}$ appelé "signe" , $M \in [1,10[$ appelé "mantisse" , $E \in \mathbb{Z}$ appelé "exposant

➢ **Notation scientifique en base 2 :**

● Exemple : $1010.101 = 1.010101 \times 2^{3}$

● Plus généralement : $N = (-1)^{S} \times M \times 2^{E}$

où : M : est la mantisse (virgule fixe) et E : l'exposant (signé).

Le codage en base 2, format virgule flottante, revient à coder le signe, la mantisse et l'exposant.

Exemple 1 : Changement de base 10->2

4,25(10) = ? (2) format virgule fixe
4(10) = 100(2)
0,25 x 2= 0,5 -> 0
0,5 x 2 = 1,0 -> 1

donc : 4,25(10) = 100,01(2)

Exemple 2: Codage en base 2, format virgule flottante, de (3,25)

3,25(10) = 11,01(2) (en virgule fixe)

$= 1,101 . 2^{1}(2)$

$= 110,1 . 2^{-1}(2)$

● Pb : différentes manières de représenter E et M

➢ **La norme IEEE 754 (1985)**

La norme IEEE 754 définit la façon de coder un nombre réel. Cette norme se propose de coder le nombre sur 32 bits et définit trois composantes :

• le signe est représenté par un seul bit, le bit de poids fort (le signe – : bit 1, le signe + : bit 0)
• l'exposant est code sur les 8 bits consécutifs au signe
• la mantisse (les bits situes après la virgule) sur les 23 bits restants. Exemple : 11,01 -> 1,101 donc M =101

$N = (-1)^{S} \times 1,M \times 2^{Eb}$

Exemple:

Conversion décimale - IEEE754
(Codage d'un réel)

$35{,}5_{(10)} = ?_{\text{(IEEE 754 simple précision)}}$

Nombre positif, donc SM = 0

$35{,}5_{(10)} = 100011{,}1_{(2)}$ **(virgule fixe)**

$= 1{,}000111 \cdot 2^{5}{}_{(2)}$ **(virgule flottante)**

Exposant = Eb-127 = 5, donc Eb = 132

1,M = 1,000111 donc M = 00011100...

SM M

$0\,10000100\,00011100000000000000000_{\text{(IEEE 754 SP)}}$

Eb

Conversion IEEE754 - Décimale
(Evaluation d'un réel)

SM M

$0\,10000001\,11100000000000000000000_{\text{(IEEE 754 SP)}}$

Eb

S = 0, donc nombre positif

Eb = 129, donc exposant = Eb-127 = 2

1,M = 1,111

$+ 1{,}111 \cdot 2^{2}{}_{(2)} = 111{,}1_{(2)} = 7{,}5_{(10)}$

3.5) Codage des caractères

Caractère = Symbole graphique atomique utilisé pour la transcription écrite d'une langue orale

● Différentes classes de symboles :

— Symboles littéraires : lettres, signes de ponctuation

— Symboles calculatoires : chiffres, opérations arithmétiques

— Symboles indentatoires : espace, tabulation, ligne

● Le nombre de symboles dépend de la langue

— Anglais ~ 50 symboles, Chinois ~ 3000 symboles

● Il n'existe pas de relation naturelle entre un caractère et une représentation en binaire

► Convention : Codage = Table de caractères

● Il existe plusieurs logiciels traitant du texte de différentes langues. Comment sont codés les caractères pour permettre d'afficher une langue et aussi plusieurs langues? En fait, il s'agit de coder les lettres de l'alphabet y compris les chiffres et les caractères de contrôle du clavier. Le choix d'un code dépend du pays et de la langue.

- Différents codes sont utilisés :
  - ASCII (7 bits)
  - ASCII étendue (8 bits)
  - Unicode (16 bits)
  - Etc.

1/ Le codage ASCII

Le codage ASCII (American Standard Code for Information Interchange) a été initialement conçu en 1963, puis modifié en 1967 pour inclure les lettres minuscules. Il est devenu *la norme ISO 646 en 1983*.

- 7 bits pour représenter 128 caractères (0 à 127)
- 48 à 57 : chiffres dans l'ordre (0,1,…,9)
- 65 à 90 : les alphabets majuscules (A,…,Z)
- 97 à 122 : les alphabets minuscule (a,…z)

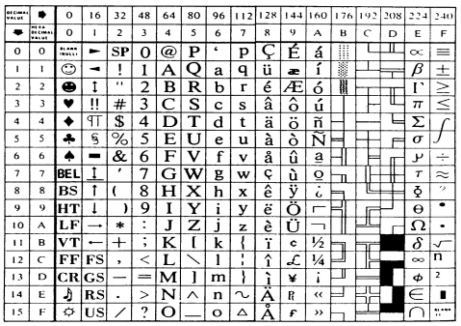
Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	$	68	44	D	100	64	d
5	5	[ENQUIRY]	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]

Figure : code ASCII

2/ Code ASCII Etendu

- 8 bits pour représenter 256 caractères (0 à 255)
- Code les caractères accentues : à, é,…etc.
- Compatible avec ASCII

Code ASCII Etendu



3/ Code Unicode (mis au point en 1991)

- 16 bits pour représenter 65 536 caractères (0 à 65 535)
- Compatible avec ASCII
- Code la plupart des alphabets : Arabe, Chinois, ….
- On en a défini environ 50 000 caractères pour l'instant
- Ce standard est lié à la *norme ISO/CEI 10646* qui décrit une table de caractères équivalente. La dernière version, Unicode 12.0, a été publiée en *mars 2019*.

3.6) Codage des images

On distingue 2 catégories de codage des images :

- ***Les images vectorielles*** : l'image est codée par un ensemble de formules mathématique
- ***Les images bitmap ou matricielles*** : l'image est codée comme un tableau de points. le codage matriciel permet de représenter numériquement les photos.

Bitmap signifie «carte de bits : l'image est décrite point par point. Les points d'une image sont appelés des pixels («*picture elements*»). Chaque pixel est décrit par un nombre indiquant sa couleur. L'image est donc représentée par une série de nombres.

- Le codage de l'image se fait en écrivant successivement les bits correspondant à chaque pixel, ligne par ligne, en commençant par le pixel en bas à gauche.

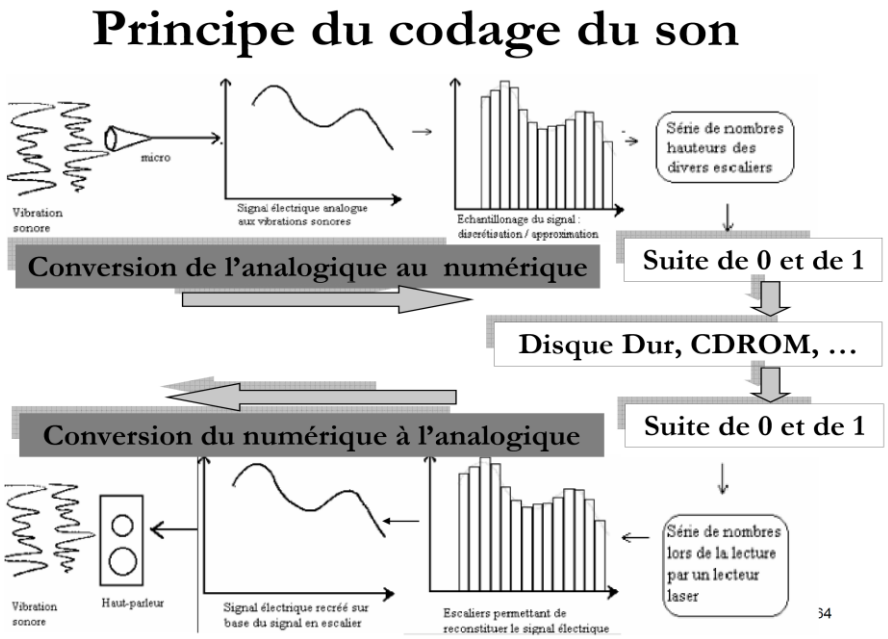
- Le codage est simple mais l'image bitmap occupe beaucoup de mémoire : plus les pixels sont petits, plus nombreux ils sont ! Ce qui explique la nécessité de compression.
- Trois paramètres définissent une image bitmap :
  - Le nombre de colonnes
  - Le nombre de lignes
  - Le nombre de couleur par pixel

Les 2 premiers paramètres déterminent ce qu'on appel par définition de l'image.
Par exemple 800x600 pixels. Le dernier paramètre détermine ce qu'on appelle par profondeur de l'image.

3.7) Codage du son

Le son se représente naturellement sous la forme d'un signal analogique. Il s'agit d'une onde acoustique issue de la vibration d'une membrane entrainant des mouvements ondulatoires de l'air environnant. Un haut-parleur ou les cordes vocales d'un homme peuvent être à l'origine de ces vibrations.

Le codage d'un tel signal consiste à le représenter en binaire afin de pouvoir le transmettre à l'ordinateur qui ne peut stocker et traiter les informations qu'en binaire.



3.8) Les codes-barres

Un code-barres est la représentation d'une donnée numérique ou alphanumérique sous forme d'u symbole constitue de barres et d'espaces dont l'épaisseur varie en fonction des données ainsi codées.

Ils sont destinés à une lecture automatisée par un capteur électronique.

Lorsque ces barres sont remplacées par des petites carres ou des points, on parle de code en deux dimensions.

3.8.1) Codes-barres EAN

Le numéro EAN (European Article Numbering) identifie des articles ou des unités logistiques de façon unique, codé sous forme de codes-barres. L'EAN est composé de 8 ou 13 chiffres représentés sous forme de séquences de barres noires et blanches formant un code-barres. Ce type de code-barres se trouve sur la presque totalité de produits courants (alimentation, vêtements, droguerie, papeterie, électroménager, etc.). Le code est lu lors du passage aux caisses des commerces.

Il existe des codes EAN-8 des codes EAN-13, composés respectivement de 8 ou 13 chiffres :

- les codes EAN-8 sont réservés à l'usage sur des produits de petite taille ;
- les codes EAN-13 sont utilisés sur tous les autres produits.

Le système des codes EAN, comme tous les systèmes de codes-barres, fait appel à des notions d'arithmétique modulaire. Sa structure tient compte des contraintes physiques liées aux conditions de leur impression et de leur lecture. En effet, la reconnaissance d'un code-barres nécessite :

- de pouvoir séparer et mesurer les largeurs des barres, à des distances de lecture variable, sur différents types de capteurs et en l'absence de toute horloge ou mesure de référence,
- de pouvoir le faire en outre dans n'importe quel sens de lecture et quelles que soient les couleurs effectivement utilisées.

Structure du code EAN-13

• Le caractère de **Début** code 101 (1 = noir, 0 = blanc)
• Le second caractère du **Préfixe** (6). Le premier caractère du préfixe n'est pas code.
• Les cinq caractères du **Numéro de Participant** (12345)
• Le **Séparateur Central** est code 01010
• Les cinq caractères du **Numéro d'Article** (67890)
• Le **Check Digit** (0)
• Le caractère de **Fin** code 101.

3.8.2) QR Codes

Le code QR (ou ***QR code*** en anglais) est un code-barres en deux dimensions (ou code à matrice) constitue de modules noirs disposes dans un carre a fond blanc. Ce cours contient un QR code à chaque début de chapitre. Le nom *QR* est l'acronyme de l'anglais *Quick Response*, car son contenu de données peut être décode rapidement.

Le code QR a été créé par l'entreprise japonaise Denso-Wave en 1994 pour le suivi des pièces de voiture dans les usines de Toyota.
Les codes QR peuvent mémoriser des adresses web, du texte, des numéros de téléphone, des SMS ou autres types de données lisibles par les smartphones et les téléphones mobiles équipes d'une application de lecture (lecteur de code QR ou *QR reader* en anglais).

4) Sources

4.1) Sources d'information

Une source d'information est caractérisée par :

- un alphabet de source $S = \{s_1, s_2, ..., s_m\}$ décrivant quels sont les symboles s_i que cette source peut émettre ;
- une distribution de probabilité $P = \{p_1, p_2, ... , p_m\}$, chaque p_i étant un réel compris entre 0 et 1 donnant la probabilité que la source délivre le symbole $s_i \in S$ et telle que $\sum_{i=0}^{m} p_i = 1.$

Plus formellement, une source est une variable aléatoire S prenant ses valeurs dans un ensemble fini *S* et dont la loi de probabilité est donnée par P. Ainsi pour tout symbole $s_i \in S$

$$\Pr(S = s_i) = p_i.$$

Nous noterons les sources d'information (S; P) sous la forme du couple (alphabet, distirbution de probabilité).

- Les caractères utilisés dans différentes langues. L'alphabet peut être le même, mais les distributions de probabilité diffèrent (par exemple le caractère le plus fréquent en français est le ***e*** alors qu'en turc il s'agit du ***a***).
- Les caractères utilisés dans différents langages de programmation. L'alphabet est toujours le même, mais les distributions de probabilité diffèrent.
- Source uniforme : source pour laquelle la distribution de probabilité est uniforme, c'est-à dire pour laquelle la probabilité de chaque symbole est la même, ou autrement dit, pour tout symbole $s \in S$, on a : $\Pr(S = s_i) = 1/m$

4.2) Quantité d'information

La quantité d'information contenue dans un symbole $s \in S$ d'une source $S = (S; P)$, de probabilité non nulle est

$$I(s) = -\log_2 \Pr(S = s).$$

Par extension, pour un symbole s de probabilité nulle, on pose

$$I(s) = +\infty.$$

L'unité de la quantité d'information est le bit.

Propriétés :

1. Pour toute source *S* et tout symbole s de cette source, on a 0 <= *I(s).*

2. De plus, *I(s)* = 0 si et seulement si Pr(S = s) = 1, autrement dit si la source émet le symbole s avec certitude, et jamais les autres symboles.

3. La quantité d'information dans un symbole est infinie si et seulement si la probabilité de ce symbole est nulle, autrement dit si ce symbole n'est jamais produit par la source.

4. La quantité d'information contenue dans un symbole augmente lorsque la probabilité de ce symbole diminue.

Exemple :
Prenons un dé à 8 faces non pipé. L'information véhiculée par « Le dé a fait 5 » est de 3 bits.

4.3) Entropie d'une source

L'entropie est une notion fondamentale pour la manipulation d'un code. C'est en effet une mesure, à la fois de la quantité d'information qu'on peut attribuer à une source (ce qui sera utile pour la compression des messages), et du degré d'ordre et de redondance d'un message, ce qui est une information cruciale pour la cryptographie.
L'entropie d'une source *S* est la valeur moyenne, ou espérance mathématique, de la quantité d'information contenue dans chacun de ces symboles. On la note *H(S).*

$$H(S) = \sum_{s \in \mathcal{S}} \Pr(S = s) I(s).$$

On peut aussi exprimer l'entropie par

$$\begin{aligned} H(S) &= \sum_{i=1}^{m} p_i I(s_i) \\ &= -\sum_{i=1}^{m} p_i \log_2 p_i \end{aligned}$$

L'unité de l'entropie d'une source est le bit.

Propriétés :
1. L'entropie d'une source ne dépend que des valeurs pi des probabilités des symboles et non des symboles eux-mêmes.
2. Pour toute source on a

$$0 \leq H(S) \leq \log_2 m.$$

3. L'entropie d'une source est nulle si et seulement si un symbole de la source a une probabilité égale à 1, les autres ayant tous une probabilité nulle.
4. L'entropie d'une source est maximale, c'est-à-dire égale à log2m si et seulement si la distribution de probabilité est uniforme.

Exemple :

1. Soit la source définie par l'alphabet $S = \{s_1, s_2, s_3, s_4\}$ et la distribution de probabilité

s	s_1	s_2	s_3	s_4
$\Pr(S = s)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

La quantité d'information contenue dans chacun des symboles est la même et vaut

$$I(s) = -\log_2 \frac{1}{4} = \log_2 4 = 2.$$

L'entropie de la source est donc égale à

$$H(S) = 2.$$

2. Avec la distribution de probabilités

s	s_1	s_2	s_3	s_4
$\Pr(S = s)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

,

les quantités d'information contenues dans chacun des symboles sont

$$
\begin{aligned}
I(s_1) &= -\log_2 \frac{1}{2} = 1 \\
I(s_2) &= -\log_2 \frac{1}{4} = 2 \\
I(s_3) &= -\log_2 \frac{1}{8} = 3 \\
I(s_4) &= -\log_2 \frac{1}{8} = 3
\end{aligned}
$$

et l'entropie de la source vaut

$$H(S) = \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \frac{1}{8} \times 3 = \frac{7}{4}.$$

3. Enfin avec la distribution de probabilités

s	s_1	s_2	s_3	s_4
$\Pr(S = s)$	1	0	0	0

,

les quantités d'information contenues dans chacun des symboles sont

$$
\begin{aligned}
I(s_1) &= -\log_2 1 = 0 \\
I(s_2) &= +\infty \\
I(s_3) &= +\infty \\
I(s_4) &= +\infty
\end{aligned}
$$

et l'entropie de la source vaut

$$H(S) = 1 \times 0 + 0 + 0 + 0 = 0.$$

5) Codages optimaux

Étant donnés une source *S = (S; P)* et un alphabet cible ***A***, on souhaite coder « au mieux » les symboles de S par des mots de ***A***.

5.1) Longueur moyenne d'un codage de source

La longueur moyenne d'un codage c'est la moyenne des longueurs des mots utilisés dans le codage, longueurs coefficientées par la probabilité des symboles de source correspondant. Elle s'exprime par

$$\overline{n}_{\mathbf{c}} = \sum_{i=1}^{m} p_i |\mathbf{c}(s_i)|.$$

On peut aussi exprimer cette longueur moyenne comme étant l'espérance mathématique de la longueur des mots associés à chaque symbole de la source

$$\overline{n}_{\mathbf{c}} = \sum_{s \in \mathcal{S}} \Pr(S = s) |\mathbf{c}(s)|.$$

Exemple :
Considérons à nouveau l'alphabet source *S = {s₁, s₂, s₃, s₄ }*, et les deux codages binaires des symboles de S définis par

s	s_1	s_2	s_3	s_4
$\mathbf{c_1}(s)$	00	01	10	11
$\mathbf{c_2}(s)$	0	11	100	101

Il est évident que la longueur moyenne du codage c1 est égale à 2 quelque soit la distribution de probabilités de la source.

$$\overline{n}_{\mathbf{c_1}} = 2.$$

- ❖ Pour le second codage, sa longueur moyenne dépend de la distribution de probabilités.

1. Avec la distribution uniforme

s	s_1	s_2	s_3	s_4
$\Pr(S=s)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

,

la longueur moyenne est

$$\overline{n}_{\mathbf{c_2}} = \frac{1}{4}(1+2+3+3) = \frac{9}{4}.$$

2. Avec la distribution

s	s_1	s_2	s_3	s_4
$\Pr(S=s)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

,

la longueur moyenne est

$$\overline{n}_{\mathbf{c_2}} = \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3 = \frac{7}{4}.$$

3. Et avec la distribution

s	s_1	s_2	s_3	s_4
$\Pr(S=s)$	1	0	0	0

,

la longueur moyenne est

$$\overline{n}_{\mathbf{c_2}} = 1.$$

5.2) Codage optimal d'une source

Comme l'exemple précédent le montre, les différents codages d'une même source n'ont pas la même longueur moyenne. Parmi tous les codages d'une source donnée, il en existe dont la longueur moyenne est minimale. De tels codages sont dits ***optimaux***.

Un codage ***c*** d'une source *S = (S; P)* dans un alphabet cible *A* est dit optimal, si pour tout autre codage ***c'*** de la même source sur le même alphabet cible, on a :

$$\overline{n}_{\mathbf{c}} \leq \overline{n}_{\mathbf{c'}}.$$

Exercice :

Quelle est l'entropie d'une source qui é un caractère 1 *avec une probabilité* 0.1 *et le caractère* 0 *avec une probabilité* 0.9 *?*

Pourquoi une faible entropie est-elle un bon augure pour la compression ?

Sol :

On trouve une entropie de 0,14 (à comparer avec l'entropie de la source équiprobable 0,3. Le 0 sera très courant dans le message, on peut s'attendre à de longues suites de 0. On peut donc choisir de coder efficacement ces longues suites.

Chapitre 3 : Cryptologie

1) Définition (Cryptologie)

- Etude de la *protection de l'information sous forme numérique*contre des accès ou manipulations non-autorisés.
- La **cryptologie** : Mécanisme permettant *de camoufler des messages i.e., de le rendre incompréhensible pour quiconque n'est pas autorisé.* Elle fait partie d'un ensemble de théories et de techniques liées à la transmission de l'information (théorie des ondes électromagnétiques, théorie du signal, théorie des codes correcteurs d'erreurs, théorie de l'information, théorie de la complexité,...).
- Ensemble de techniques permettant d'assurer la sécurité des systèmes d'information.

- **La cryptologie** est une science mathématique qui comporte 2 branches : la cryptographie et la cryptanalyse.

cryptologie= cryptographie + cryptanalyse

- **cryptographie**: conception des algorithmes cryptographiques
- **cryptanalyse**: évaluation de la sécurité des algorithmes cryptographiques

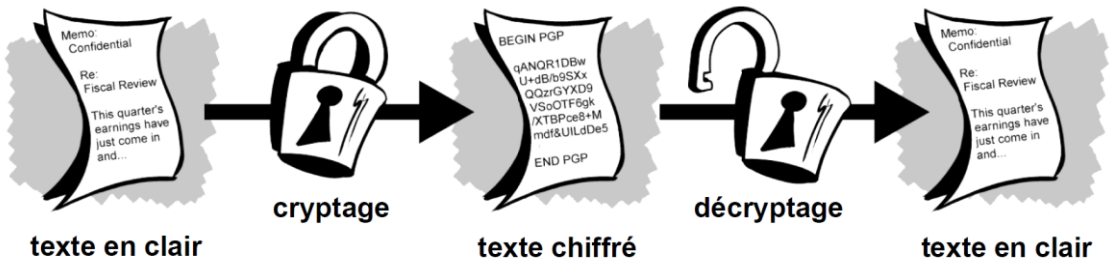


Figure : Cryptage et décryptage

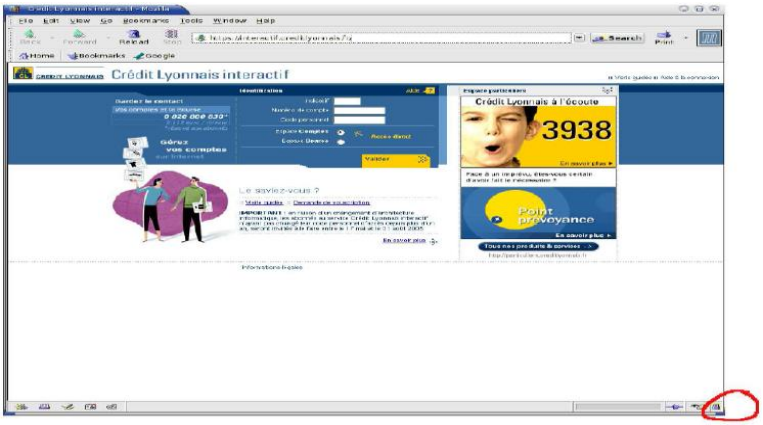
2) Quelques définitions

- **Cryptogramme:** Message chiffré ou codé.
- **Cryptographie:** Discipline incluant les principes, les moyens et les méthodes de transformation des données, dans le but de masquer leur contenu, d'empêcher leur modification ou leur utilisation illégale.
- **Cryptologie:** Science des messages secrets. Se décompose en cryptographie et cryptanalyse. Le mot cryptologie est souvent utilisé comme synonyme de cryptographie.
- **Chiffre:** Ensemble de procédés et ensemble de symboles (lettres, nombres, signes, etc.) employés pour remplacer les lettres du message à chiffrer. On distingue généralement les chiffres à transposition et ceux à substitution.
- **Chiffrer=Crypter:** Transformer un message afin qu'il ne soit lisible qu'à l'aide d'une clef.
- **Décrypter:** Parvenir à restaurer des données qui avaient été chiffrées, donc à leur faire retrouver leur état premier ("en clair"), sans disposer des clefs théoriquement nécessaires.

- **Clef:** Dans un système de chiffrement, elle correspond à un nombre, un mot, une phrase, etc. qui permet, grâce à l'algorithme de chiffrement, de chiffrer ou de déchiffrer un message.
- **Texte en clair :** c'est le message à protéger.
- **Texte chiffré :** c'est le résultat du **chiffrement** du **texte en clair**.
- **Chiffrement :** c'est la méthode ou l'algorithme utilisé pour transformer un **texte en clair** en **texte chiffré**.
- **Déchiffrement** c'est la méthode ou l'algorithme utilisé pour transformer un **texte chiffré** en **texte en clair**.
- **Clé :** c'est le secret partagé utilisé pour **chiffrer** le **texte en clair** en **texte chiffré** et pour **déchiffrer** le **texte chiffré** en **texte en clair**. On peut parfaitement concevoir un algorithme qui n'utilise pas de **clé**, dans ce cas c'est l'algorithme lui-même qui constitue **la clé**, et son principe ne doit donc en aucun cas être dévoilé.
- **Cryptographie :** cette branche regroupe l'ensemble des méthodes qui permettent de **chiffrer** et de **déchiffrer** un **texte en clair** afin de le rendre incompréhensible pour quiconque n'est pas en possession de la **clé** à utiliser pour le **déchiffrer**.
- **Cryptanalyse :** c'est l'art de révéler les **textes en clair** qui ont fait l'objet d'un **chiffrement** sans connaître la **clé** utilisée pour **chiffrer** le **texte en clair**.
- **Cryptologie :** il s'agit de la science qui étudie les communications secrètes. Elle est composée de deux domaines d'étude complémentaires : la **cryptographie** et la **cryptanalyse**.
- **Décrypter :** c'est l'action de retrouver le **texte en clair** correspondant à un **texte chiffré** sans posséder la clé qui a servit au chiffrement. Ce mot ne devrait donc être employé que dans le contexte de la cryptanalyse.
- **Coder, décoder :** c'est une méthode ou un algorithme permettant de modifier la mise en forme d'un message sans introduire d'élément secret. Le Morse est donc un code puisqu'il transforme des lettres en trait et points sans notion de secret. L'ASCII est lui aussi un code puisqu'il permet de transformer une lettre en valeur binaire.

3) Ou utilise-t-on la cryptographie ?

- Internet (conventualité, anonymat, authentification (s'agit-il bien de ma banque ?))



- Signature électronique (vérifiable, authentique, non-répudiation (je n'ai jamais signé ce texte ...))

- Vote électronique (le résultat reflète le vote, chaque vote est confidentiel, on ne peut pas connaître des résultats partiels, seuls les électeurs peuvent voter et une seule fois)

- Paiement par carte bancaire (Est-ce qu'il s'agit d'une vraie carte ? Est-ce que le montant débité sera égal au montant crédité ? Est-ceque le code secret est bien protégé ?)

- Décodeurs (vérification de l'abonné, impossibilité de retransmettre les données décodées à une tierce personne, mise à jour de l'abonnement)

- Porte monnaie électronique (pas de création de fausse monnaie, pas de création de faux porte-monnaie)

- Bases de données sécurisées (ex : carte vitale. seules les personnes habilitées ont accès à la vue partielle à laquelle elles ont droit, les données peuvent être échangées entre un médecin, un laboratoire, un hôpital, mise à jour possible des données)

4) Cryptographie historique

4.1) La technique grecque

En 487, les grecs emploient un dispositif appelé la **"scytale"** - un bâton autour duquel une bande longue et mince de cuir était enveloppée et sur laquelle on écrivait le message. Le cuir était ensuite porté comme une ceinture par le messager. Le destinataire avait un bâton identique permettant d'enrouler le cuir afin de déchiffrer le message.

- Méthode de transposition grecque.
- Le diamètre du bâton est la clé.

4.2) Le code de César

Jules César employait une substitution simple (substitution mono-alphabétique) avec l'alphabet normal (il s'agissait simplement de décaler les lettres de l'alphabet d'une quantité fixe) dans les communications du gouvernement.
Le code de César est utilisé dans l'armée romaine durant la guerre des Gaules.

- Méthode : Décalage alphabétique de 4 caractères.
- clair : ABCDEFGHIJKLMNOPQRSTUVWXYZ
- chiffré : EFGHIJKLMNOPQRSTUVWXYZABCD

Exemple : Sile texte chiffré est "VIRHSRW E GIWEV GI UYM IWX E GIWEV",

le texte en clair est : **rendons a cesar ce qui est a cesar**

Un autre algorithme, nommé **ROT13, basé sur le même fonctionnement a existé au tout début de l'informatique. Le décalage était de treize lettres, et c'est donc le même algorithme pour le chiffrement et le déchiffrement.

4.3) Le chiffrement de Vigenère

Inventé par Blaise de Vigenère en 1586.

Chiffrement

- Sélectionner la colonne (texte) et la ligne (clé).
- La clé est répétée autant que nécessaire.

Déchiffrement

- Pour chaque lettre de la clé (répétée) on cherche dans la colonne correspondante la lettre chiffrée.
- Le résultat est la lettre en 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

Si le texte à chiffrer est "**rendons avigenere ce qui est avigenere**", avec le mot clé "**RAYMOND**",

le texte chiffré est : **IELPCAV R VGSSAHIE AQ EHL VSR M JVJVNCDS"**

Exemple (Chiffrement de Vigenère)

Considérons encore une fois des messages faits des **27 lettres** de l'anglais (**espace blanc inclus**). La clé est alors une séquence de caractères, par exemple k ='INFORMATION'.

- Comment le message 'VIGENERE CIPHER IS ALSO QUITE SIMPLE' est-il codé ?

Si nous décidons que la lettre 'A' correspond à '1' et **l'espace à 27**, la lettre 'I' correspond alors à 9, et donc la première lettre du message , 'V', est donc codée en 'V'+9='D', la deuxième lettre du message 'I' est codée en 'I'+'N'='I'+14='W', la troisième lettre 'G' en 'G'+'F'='G'+6='M', etc.

Voici le codage complet :

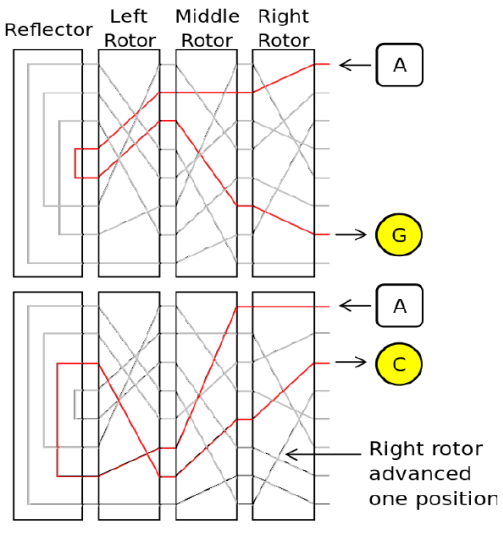
VIGENERE CIPHER IS ALSO QUITE SIMPLE
INFORMATIONINFORMATIONINFORMATIONINF
DWMTERSYIRWYVKFRVTTJ FXNWI FFTAX YZK

c.-à-d. que le message codé est 'DWMTERSYIRWYVKFRVTTJ FXNWI FFTAX YZK'.

4.4) La machine Enigma (Machine électromécanique portable)

La machine Enigma est née en 1918 dans l'entreprise de l'inventeur Arthur Scherbius en Allemagne. Elle a été utilisée par l'armée allemande durant la seconde guerre mondiale.

Il s'agit d'une machine à chiffrer et à déchiffrer mécanique qui allie à la fois les méthodes de substitution et de transposition.



4.5) Le chiffrement aujourd'hui

Jusqu'au milieu des années 70, les seuls cryptosystèmes connus étaient symétriques (ondit aussi conventionnels ou à clé secrète) : la **clé de chiffrement** *KC* était la mêmeque **la clé de déchiffrement** *KD* (ou du moins, la connaissance de la **clé dechiffrement** permettait d'en déduire la clé de déchiffrement) ce qui obligeait à gardersecrète la clé *KC* elle aussi.
En 1976, W. Diffie et M. Hellman introduisirent le concept de **cryptographie à clépublique** (ou asymétrique). Dans ce type de système, la **clé de chiffrement** estpublique, c'est à dire connue de tous. Seule la clé de déchiffrement reste secrète.
En 1978, le premier système de chiffrement à clé publique fut introduit par R. Rivest,A. Shamir et L. Adleman : le système RSA. Ce système est un des seuls qui aientrésisté à la cryptanalyse.
1990 : Premières publications sur la cryptographie quantique.
1992 : MD5 (fonction de hachage).
1994 : DSA (signature numérique).
2000 : AES : nouveau standard dans le chiffrement à clé privée.

5) Terminologie

- Protagonistes traditionnels :
  - Alice et Bob : souhaitent se transmettre des informations
  - Oscar : un opposant qui souhaite espionner Alice et Bob
- Objectif fondamental de la cryptographie
  - permettre à Alice et Bob de communiquer sur un canal peu sûr
  - Oscar ne doit pas comprendre ce qui est échangé.
- Texte clair : information qu'Alice souhaite transmettre à Bob
  - Ex : texte en français, donnée numérique etc...
- Chiffrement : processus de transformation d'un message M de tellemanière à le rendre incompréhensible
  - Basé sur une fonction de chiffrement
  - On génère ainsi un message chiffré C = E(M)
- Déchiffrement : processus de reconstruction du message clair à partir du message chiffré
  - Basé sur une fonction de déchiffrement D
  - On a donc D(C) = D(E(M)) = M (D et E sont injectives)

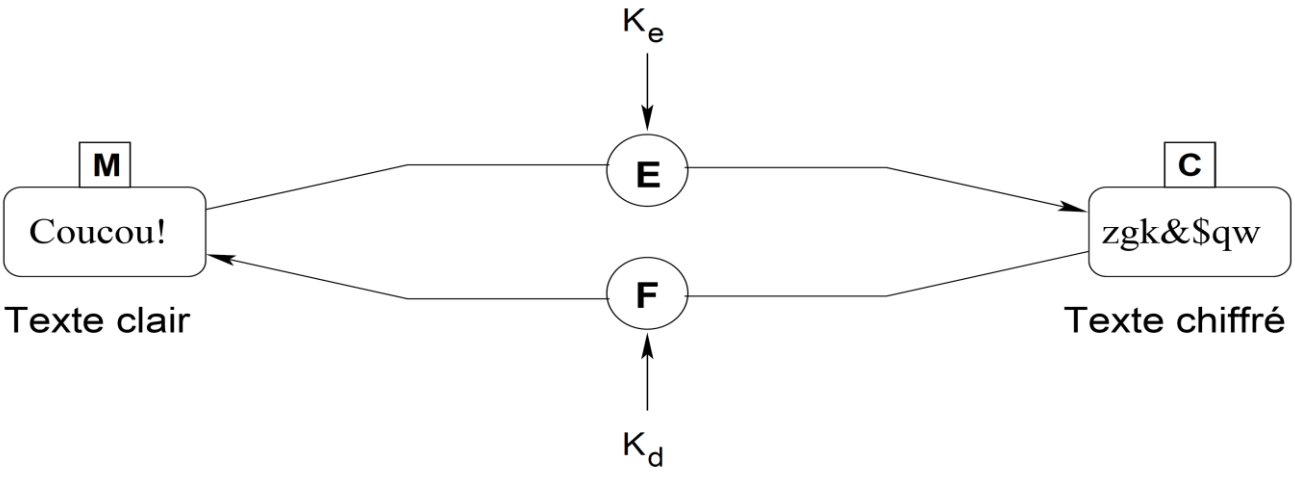


Figure : Relation fondamentale de la cryptographie.

En pratique E et D sont paramétrées par des clefs Kd et Ke

- Deux grandes catégories de systèmes cryptographiques
  - Systèmes à clefs secrètes (symétriques) : Ke=Kd=K
  - Systèmes à clefs publiques (asymétriques) : Ke ≠ Kd
- Deux types de fonctionnement
  - Par flot : chaque nouveau bit est manipulé directement
  - Par bloc : chaque message est découpé en blocs

6) A quoi sert la cryptographie ?

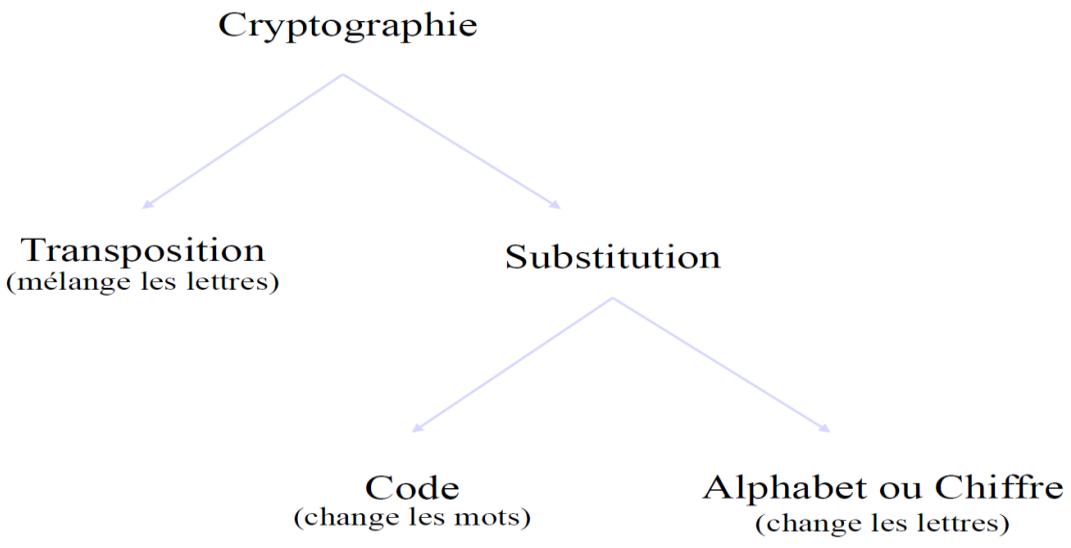
Les communications échangées entre Alice et Bob sont sujettes à un certain nombre de menaces. La cryptographie apporte un certain nombre de fonctionnalités permettant de pallier ces menaces, résumées dans le sigle **CAIN**, pour Confidentialité, Authentification, Intégrité, Non-répudiation :

CAIN

(Confidentialité - Authentification - Intégrité - Non-répudiation)

- **Confidentialité** des informations stockées/manipulées
  - utilisation d'un algorithme de chiffrement.
  - empêcher l'accès aux infos pour ceux qui ne sont pas autorisés.
- **Authentification** d'utilisateurs/de ressources
  - utilisation d'algorithmes d'authentification.
  - Alice s'identifié à Bob en prouvant qu'elle connaît un secret S,(ex : un mot de passe).
- **Intégrité** des informations stockées/manipulées
  - vérifier que les infos transmises n'ont pas subie d'altérations
- **Non-répudiation** (signature) des informations
  - utilisation d'algorithmes de signatures
  - empêcher un utilisateur de se dédire

6.1) Cryptographie ancienne



6.1.1) Transposition

Chiffrement de type anagramme : mélange les lettres du message

• Sécurité théorique

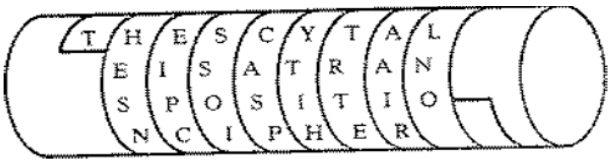
– Message de 35 lettres : 35! chiffrés possibles

• Problèmes

– Confusion sur la syntaxe mais … chaque lettre conserve sa valeur

– Clé de chiffrement « complexe »

– Ex: Scytale spartiate (5ème siècle av JC)



6.1.2) Substitution

• Chiffrement en changeant d'alphabet

– Kama sutra : mlecchita-vikalpà (art de l'écriture secrète, 4ème siècle av JC)

• Sécurité théorique

– Alphabet de 26 lettres : 26! alphabets possibles

• Problèmes

– Confusion sur l'alphabet mais … chaque lettre conserve sa place d'origine

– Ex: Chiffrement de Jules César (1er siècle av JC)

Alphabet clair : abcdefghijklmnopqrstuvwxyz
Alphabet chiffré : DEFGHIJKLMNOPQRSTUVWXYZABC

Texte clair : errare humanum est, perseverare diabolicum
Texte chiffré : HUUDUH KXPDQXP HVW, SHUVHYHUDUH GLDEROLFXP

7) Algorithmes de cryptographie

Propriétés théoriques nécessaires :

1. **Confusion** : Aucune propriété statistique ne peut être déduite du message chiffré
2. **Diffusion** : Toute modification du message en clair se traduit par une modification complète du chiffré

- **Relation fondamentale**
  - En pratique : E et D sont paramétrées par des clés K_e et K_d : E_{Ke} (M) = C et D_{Kd} (C) = M
  - Ke ; Kd ∈ espace des clés.
  - Définit deux catégories de systèmes cryptographiques :
    - Systèmes à **clé secrète** (ou **symétriques**) ($K_e = K_d = K$)
    - Systèmes à **clé publique** (ou **asymétriques**) ($K_e \neq K_d$)

8) Les grands types de menaces

8.1) Attaques passives/actives

On distingue déjà les attaques passives, où Oscar se contente d'écouter les messages échangés entre Alice et Bob, et les attaques actives, dans lesquelles Oscar peut modifier le message au cours de sa transmission. Le premier type menace la confidentialité des informations, tandis que le second peut entraîner l'altération des informations ou des usurpations d'identité.

8.1.1) menaces passives

- Oscar ne fait qu'écouter le message.
- menace la confidentialité
- une information sensible parvient également à une autre personne que son destinataire légitime.

8.1.2) menaces actives

- Oscar peut modifier le contenu des messages échangés.
- menace l'intégrité de l'information.
- Exemple d'attaques actives :
  - l'usurpation d'identité (de l'émetteur ou du récepteur)
  - l'altération / modification du contenu des messages ;
  - la destruction de messages/ le retardement de la transmission ;
  - la répétition de messages (jusqu'_a engorgement)
  - la répudiation de message : l'émetteur nie avoir envoyé le message.

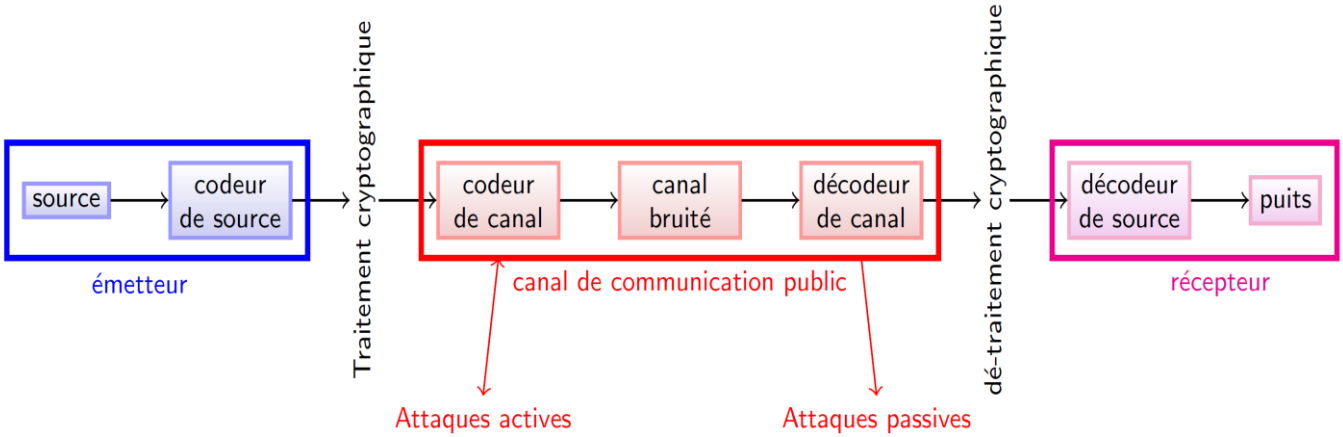


Figure : Modèle d'un système de communication cryptographique

8.2) Modélisation de l'adversaire

On veut modéliser un attaquant :

- le plus intelligent possible ---> il peut faire toutes les opérations qu'il souhaite
- qui dispose d'un temps limité.
  - on ne souhaite pas considérer les attaques faisables en 2^{80} ans
  - sinon, l'adversaire peut toujours énumérer toutes les clefs (temps exponentiel en $2^{\text{taille(clefs)}}$)

8.3) Cryptanalyse et les attaques sur un chiffrement

- Cryptanalyse : étude de la sécurité des procédés de chèrement utilisés en cryptographie
- Niveaux d'attaques possibles :
  - Texte chiffré connu : seul C est connu d'Oscar
  - Texte clair connu : Oscar connaît C et M correspondant
  - Texte clair choisi : ∀ M, Oscar peut obtenir C
  - Texte chiffré choisi : ∀ C, Oscar peut obtenir M
- garantir la confidentialité ➔ Oscar ne peut pas :
  - trouver M à partir de E(M)
  - trouver la méthode de déchiffrement D à partir d'une séquence $\{M_i ; E(M_i)\}$.

8.3.1) Algorithmes d'attaques

- Attaque brutale
  - Enumérer toutes les valeurs possibles de clefs
  - 64 bits) 2^{64} clefs = 1.844 x10^{19} combinaisons
    Un milliard de combinaisons/s) 1 an sur 584 machines
- Attaque par séquences connues
  - Deviner la clef si une partie du message est connue. ex : en-têtes de standard de courriels
- Attaque par séquences forcées
  - Faire chiffrer par la victime un bloc dont l'attaquant connaît le contenu, puis on applique l'attaque précédente ...
- Attaque par analyse différentielle
  - Utiliser les faibles différences entre plusieurs messages (ex : logs) pour deviner la clef

8.3.2) Exemple : Cryptanalyse d'un code type Jules César

Texte chiffré : « *Mp iwx jegmpi hi fvmwiv gi qiwweki* »

- Indices
  - Trois mots de deux lettres :
    - MP : la 2ème lettre est trois plus loin ; OR ou IL
    - HI : la 2ème suit la première ; DE ou TU
    - GI : la 2ème est deux plus loin ; CE ou SU
  - Un seul doublon ww ne nous aide pas
  - On sait que l'algorithme est « Jules César »
- Deux indices suggèrent que I--> E, donc n=4
- Texte clair : « *Il est facile de briser ce message* »

9. Cryptographie à clé secrète (ou symétrique)

- En cryptographie conventionnelle, également appelée cryptage de clé secrète ou de clé symétrique, ***une seule clé*** suffit ***pour le cryptage et le décryptage***.
- Méthode permettant à deux personnes possédant ***une clé secrète commune*** de s'échanger des messages de façon sécurisée.
- La confidentialité des messages dépend de la confidentialité de cette clé et de la robustesse de l'algorithme utilisé.
- L'échange de la clé secrète est la principale difficulté de la mise en œuvre de ces systèmes. En conséquence, il faut une clé différente par couple de correspondants.

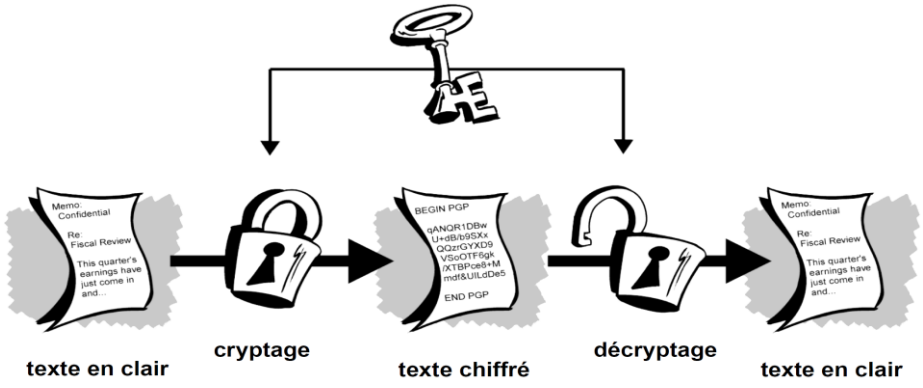


Figure : Cryptage et décryptage conventionnels (de clé secrète).

- Le chiffrement à clé secrète repose sur le principe *Ke = Kd = K*. Autrement dit, Alice et Bob conviennent secrètement d'une clef secrète *K* qui est donc utilisée à la fois pour le chiffrement et le déchiffrement. Ils conviennent également d'un algorithme cryptographique de chiffrement et déchiffrement.

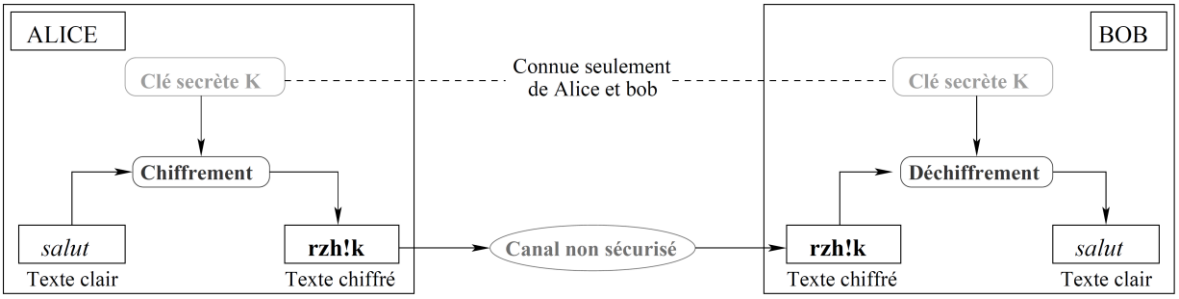


Figure : Principe du chiffrement à clef secrète.

La cryptographie ancienne, depuis le système de Jules Cesar, jusqu'à la machine enigma utilisée durant la dernière guerre mondiale de 1939-1945 est une cryptographie à clé secrète.
Dans le code secret de Jules César, **l'algorithme** constitue *à décaler les lettres de l'alphabet* et **la clé** correspond au *nombre de caractères de décalage.*

9.1) Gestion de clé et chiffrement à clé secrète

Le chiffrement à clé secrète a des avantages.

- ✓ Il est très rapide.
- ✓ Il est particulièrement utile pour chiffrer des données qui ne vont aller nulle part.

Problème de la cryptographie à clé secrète

- Le chiffrement à clé secrète seul en tant que moyen de transmission de données sécurisées peut être assez onéreux simplement en raison de la difficulté de la distribution sécurisée de la clé.
- Ne pas utiliser la même clé trop longtemps ⇒ Problème de l'échange de clé. Transmission d'une nouvelle clé oblige les deux parties à se rencontrer.
- Pour qu'un expéditeur et un destinataire communiquent de façon sûre en utilisant un chiffrement à clé secrète,
  - ils doivent se mettre d'accord sur une clé et la garder secrète entre eux.
  - S'ils sont dans des lieux géographiques différents, ils doivent faire confiance à un messager, au Bat Phone, ou à un autre moyen de communication sûr pour empêcher la divulgation de la clé secrète pendant la transmission.
- Quiconque a entendu par hasard ou intercepté la clé en transit peut plus tard lire, modifier, et contrefaire toutes les informations chiffrées ou authentifiées avec cette clé.

➢ *Donc, le problème continuel avec le chiffrement à clé secrète est la distribution de la clé: comment donnerez-vous la clé au destinataire sans que personne ne puisse l'intercepter?*

9.2) Classes de chiffrements à clé secrète

Le chiffrement à clé secrète peut se classer en deux catégories :

1/ Les systèmes de Chiffrement par blocs

Un système de chiffrement par bloc utilise une transformation sur des blocs de texte clair M de taille fixe, et renvoie des blocs de texte chiffré C de taille fixe (en général de la taille du bloc d'entrée). On doit considérer que cette transformation est publique à l'exception de la clé secrète K et bien entendu du texte clair auquel elle s'applique.

Fonctionnement

- On sépare le texte en blocs de même taille (ex: 64 bits ou 128 bits).
- Le chiffrement se fait bloc par bloc.
- Génération de sous-clés à partir de la clé privée.
- Une fonction de tour itérée autant de fois qu'il y a de sous-clés.
- Utilisation d'opérations simples (en particulier +, XOR, ...).

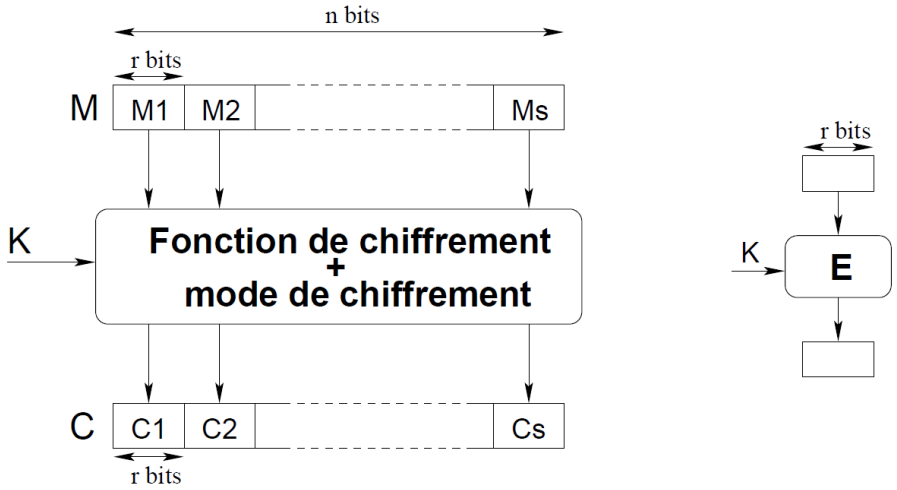


Figure : Chiffrement symétrique par bloc

<u>Parmi les exemples</u> les plus connus de chiffrement par bloc DES (Data Encryption Standard), AES (Advanced Encryption Standard), IDEA (International Data Encryption Algorithme), RC6, BLOWFISH, … .

L'algorithme DES :

Le Data Encryption Standard (standard de chiffrement de données) a été publié en 1977, et fut ainsi le premier algorithme cryptographique à petite clé secrète (56 bits) à avoir été rendu public. L'algorithme DES est un algorithme symétrique de chiffrement consiste en un réseau de Feistel de 16 tours, le message à chiffrer est découpé en blocs de 64 bits fonctionnant avec une clé de 56 bits. Il fonctionne sur 16 rondes et lors de chacune des rondes, le bloc de 64 bits est découpé en 2 blocs de 32 bits.

- A l'heure actuelle, trois jours suffisent aux ordinateurs pour le percer, et ce grâce à des attaques exhaustives !

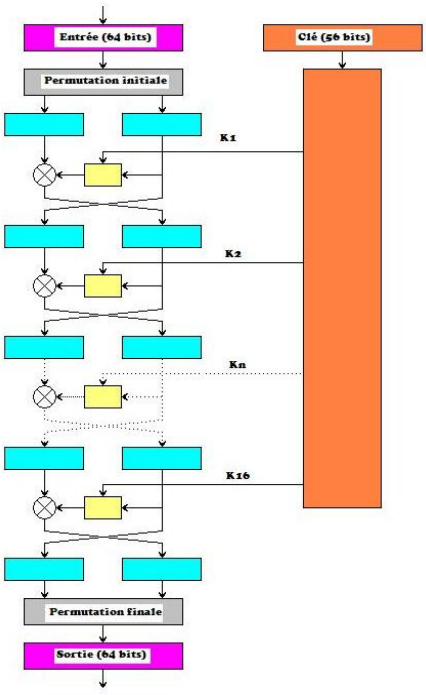


Figure : Vue général du DES

2/ Les systèmes de Chiffrement par flots

Un système de chiffrement à flot opère sur les symboles individuels du texte clair (traité bit par bit) par une transformation dépendant de la clé et de la position du symbole dans le flot de données.
Leur utilisation repose sur un générateur de nombres pseudo-aléatoires et un mécanisme de substitution bit-à-bit rapide comme l'opération 'ou exclusif' (XOR $\oplus$).

Fonctionnement

- Également appelé chiffrement de flux ou chiffrement à la volée.
- Par opposition au chiffrement par bloc, ici aucun découpage n'est effectué.
- Dès qu'un symbole est lu il est chiffré et écrit (sans avoir besoin de la lecture du symbole suivant).
- Généralement : un (XOR $\oplus$) entre un générateur pseudo-aléatoire et le message.
- Chiffrement rapide.
- Sécurité difficile.

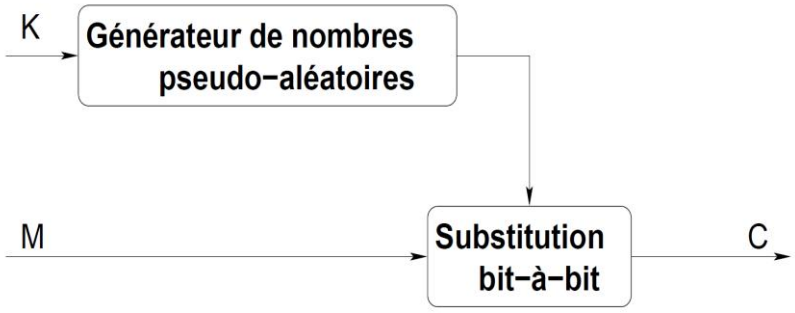


Figure : Chiffrement symétrique par flot

Parmi les exemples les plus connus de chiffrement par flot, RC4 (notamment utilisé dans le protocole SSL pour protéger les communications Internet ou encore dans le protocole WEP utilisé pour sécuriser les connexions WiFi), Py, E0 (utilisé dans les communications par Bluetooth) et A5/3 (utilisé dans les communications par GSM).

10) Cryptographie à clé publique (ou asymétrique)

Les problèmes de distribution des clés sont résolus par la cryptographie de clé publique. Ce concept a été introduit par Whitfield Diffie et Martin Hellman en 1975.

Le chiffrement à clé publique repose sur le principe Ke ≠ Kd (clés publique/privée). Autrement dit, Diffie et Hellman posèrent les bases des systèmes cryptographiques à clef publique, par analogie avec une boîte aux lettres dont Bob est le seul à posséder la clef :

- toute personne peut envoyer du courrier à Bob ;
- seul Bob peut lire le courrier déposé dans sa boîte aux lettres.

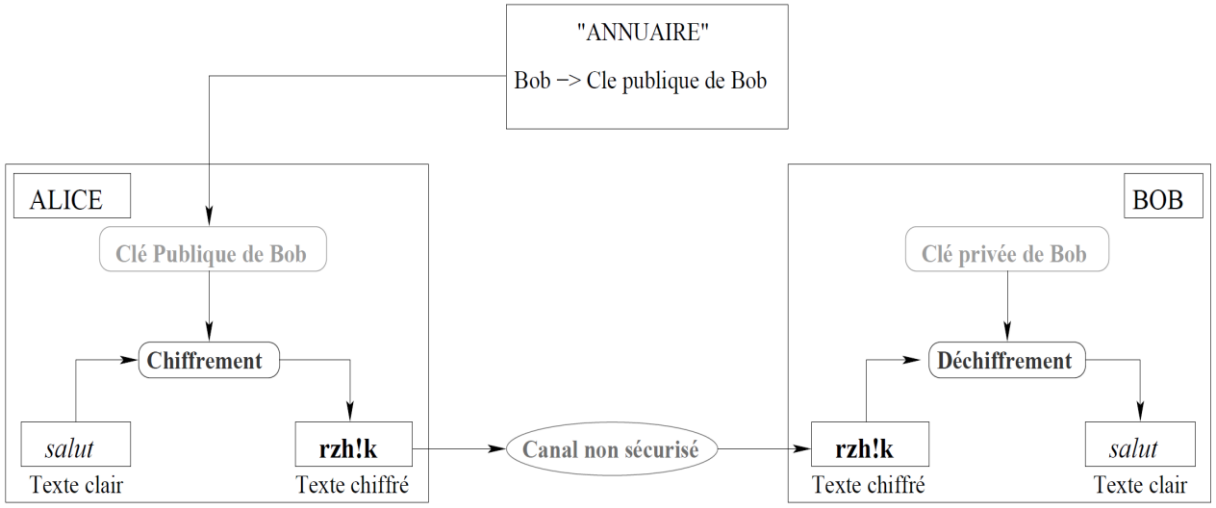


Figure : Principe du chiffrement à clé publique.

La cryptographie de clé publique est un procédé asymétrique utilisant une paire de clés pour le cryptage :

- Une clé privée qui doit être gardée secrète.
- Une clé publique qui est disponible pour tous les autres utilisateurs.

Ces deux clés sont mathématiquement liées. Dans la pratique, ***la clé publique*** sert à ***crypter*** les messages, et ***la clé privée ou secrète*** sert à ***les décrypter***.

- Vous pouvez ainsi publier votre clé publique tout en conservant votre clé privée secrète.
- Tout utilisateur possédant une copie de votre clé publique peut ensuite crypter des informations que vous êtes le seul à pourvoir lire.
- Même les personnes que vous ne connaissez pas personnellement peuvent utiliser votre clé publique.
- D'un point de vue informatique, il est impossible de deviner la clé privée à partir de la clé publique.
- Tout utilisateur possédant une clé publique peut crypter des informations, mais est dans l'impossibilité de les décrypter.
- Seule la personne disposant de la clé privée correspondante peut les décrypter.

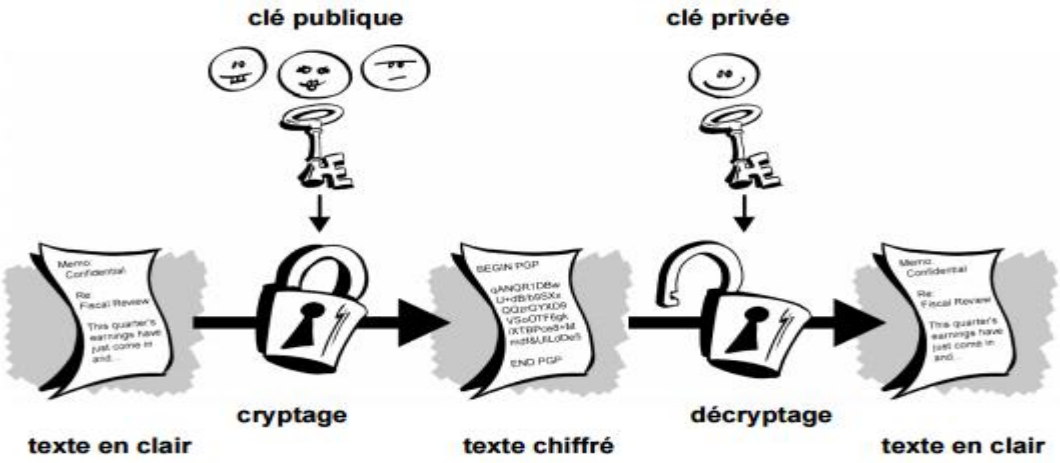


Figure : Cryptage et décryptage à clé publique

Exemple 1 :

Un ami doit vous faire parvenir un message par la poste, mais vous n'avez pas confiance en le facteur. Une solution :

- Vous envoyez à votre ami un cadenas sans sa clé et non verrouille.
- Votre ami met son message dans une boite qu'il ferme à l'aide du cadenas.
- Le facteur ne peut pas ouvrir la boite, et surtout la clé n'a pas été échangée.

Exemple 2 :

La cryptographie asymétrique repose sur cette idée: on possède un couple de clés privée/publique (Kd, Ke).

- Ke est utilisée pour le chiffrement, tout le monde peut y avoir accès.
- Kd est utilisée pour le déchiffrement, elle est privée.
- Il doit être impossible (dans l'idéal) de trouver Kd à partir de Ke

Problèmes :

- La difficulté est de générer ce couple de clés.
- Algorithmes lents (parfois un facteur 1000 avec le chiffrement symétrique).

Solution :

Utiliser le meilleur des deux mondes : chiffrer la clé de façon asymétrique, et le message de façon symétrique.

10.1) Les avantages de la cryptographie de clé publique

La cryptographie de clé publique présente un avantage majeur : en effet,

- ✓ Elle permet d'échanger des messages de manière sécurisée sans aucun dispositif de sécurité.
- ✓ L'expéditeur et le destinataire n'ont plus besoin de partager des clés secrètes via une voie de transmission sécurisée.
- ✓ Les communications impliquent uniquement l'utilisation de clés publiques et plus aucune clé privée n'est transmise ou partagée.

Parmi les exemples d'algorithmes de systèmes de cryptographie de clé publique.

- El Gamal (d'après le nom de son inventeur, Taher Elgamal),
- RSA (d'après le nom de ses inventeurs, Ron Rivest, Adi Shamir et Leonard Adleman),
- Diffie-Hellman (également d'après le nom de ses inventeurs)
- DSA, l'algorithme de signature numérique (élaboré par David Kravitz),

10.2) L'algorithme RSA

Le premier système à clé publique solide à avoir été inventé, et le plus utilisé actuellement, est le système RSA. Publié en 1977 par Ron Rivest, Adi Shamir et Leonard Adleman de l'Institut de technologie du Massachusetts (MIT), le RSA est fondé sur la difficulté de factoriser des grands nombres, et la fonction à sens unique utilisée est une fonction "puissance".
L'algorithme RSA repose sur une fonction a sens unique ou plutôt une fonction très difficilement réversible. Cette fonction difficilement réversible est la factorisation d'un nombre en produit de facteurs premiers. En effet, autant, il est très facile de choisir deux nombres (48 et 52 par exemple) et de calculer leur produit (2496), autant il est long et pénible d'extraire les facteurs premiers du nombre 2397 qui sont 47 et 51.

Remarque :
En effet, à l'heure actuelle, on utilise des cryptosystèmes hybrides, qui couplent les avantages des deux principes, alliant la souplesse de gestion des clés de la cryptographie asymétrique et les performances (vitesse) de la cryptographie symétrique.
Il existe deux types de cryptosystèmes hybrides :

- Soit, la cryptographie à clé publique sécurise le transport d'une clé symétrique
- Soit, les entités émettrice et destinatrice se mettent publiquement d'accord sur un secret commun et l'utilisent ensuite pour chiffrer les données grâce à un algorithme symétrique classique.

10.3) Utilisation de la cryptographie

1/ Les cartes bancaires

Les banques font partie des premiers utilisateurs de systèmes cryptographies. Les cartes bancaires possèdent trois niveaux de sécurité :

- Le code confidentiel : c'est la suite de chiffres à mémoriser et à saisir à l'abri des regards indiscrets.
- La signature RSA : permet de vérifier l'identité de la carte sans avoir besoin de secrets; en d'autres termes, cette vérification peut se faire sans connexion à un centre distant.
- L'authentification DES : apporte une preuve de légitimité de la carte, et se fait par connexion à un centre distant.

2/ Les navigateurs Web

Les navigateurs, ou *broswers*, tels que Mozilla Firefox ou Internet Explorer, utilisent le protocole de sécurité SSL (*Secure Sockets Layers*), qui repose sur un procédé de cryptographie par clé publique : le RSA.

11) Authentification, Intégrité

11.1) L'intégrité des informations

Une bonne cryptographie doit pouvoir offrir une garantie de l'intégrité des informations. En effet, il ne doit pas être possible de pouvoir modifier des informations cryptées de façon totalement transparente. Un processus de vérification de l'intégrité du message (crypté et en clair) doit être mis en place. Ce processus est réalisé par une fonction de hachage. Le résultat d'un hachage (hash en anglais) est une sorte de condensé du message original.

11.2) Fonctions de Hachage

Une fonction de hash (anglicisme) ou fonction de hachage est une fonction qui associe à un grand ensemble de données un ensemble beaucoup plus petit (de l'ordre de quelques centaines de bits) qui est caractéristique de l'ensemble de départ. Cette propriété fait qu'elles sont très utilisées en informatique, en particulier pour accéder rapidement à des données grâce à des "Tables de hachage". En effet, une fonction de hachage permet d'associer à une chaîne de caractères un entier particulier. Ainsi, si nous connaissons l'empreinte des chaînes de caractères stockées, nous pouvons rapidement vérifier si une chaîne se trouve ou non dans cette table. Les fonctions de hachage sont aussi extrêmement utiles en cryptographie pour accélérer le cryptage.

11.3) L'authentification des correspondants

Un aspect à ne pas négliger lorsque l'on désire faire des transactions sécurisées est l'authentification des correspondants : La personne à qui j'envoie un message crypté est-elle bien celle à laquelle je pense ? La personne qui m'envoie un message crypté est-elle bien celle à qui je pense ?
Le principe de l'authentification met en œuvre un prouveur (celui qui prétend être, qui s'est identifié) et un vérifieur (le fournisseur du service) : le vérifieur soumet un challenge au prouveur que ce dernier doit réaliser. Cela suppose qu'au préalable prouveur et vérifieur se sont entendus sur le partage d'un secret.

11.3.1) La signature digitale

C'est un code électronique unique qui permet de signer un message codé. Cette signature permet d'identifier l'origine du message : elle a la même fonction qu'une signature "à la main". C'est la clé privée qui permet de signer, et la clé publique qui permet de vérifier cette signature.

11.3.2) Le certificat digital

C'est un document électronique qui fait correspondre une clé avec une entité (personne, entreprise, ordinateur...). Cette correspondance est validée par une autorité de certification (Certificate Authority : CA). Ces certificats sont utilisés pour identifier une entité. Ce certificat est normalisé (norme X.509v3). Concrètement, les données utilisateur (identité du propriétaire de la clé, la clé publique et l'usage de la clé) sont-elles même signées par l'autorité de certification, en y incluant certaines données propres (période de validité du certificat, l'algorithme de cryptage utilisé, numéro de série, etc...).