Cours: Sécurité Informatique

2022-2023

Chapitre 02 : Initiation à la cryptographie

01 - Introduction, définitions et historique

Sommaire

- Introduction, définitions de base et historique.
- Principes des crypto-systèmes symétriques : Algorithmes DES, AES.
- Principes des crypto-système Asymétrique: Algorithme RSA.
- Fonctions de hachages, signatures numérique.

01 - Introduction, définitions et historique

Introduction

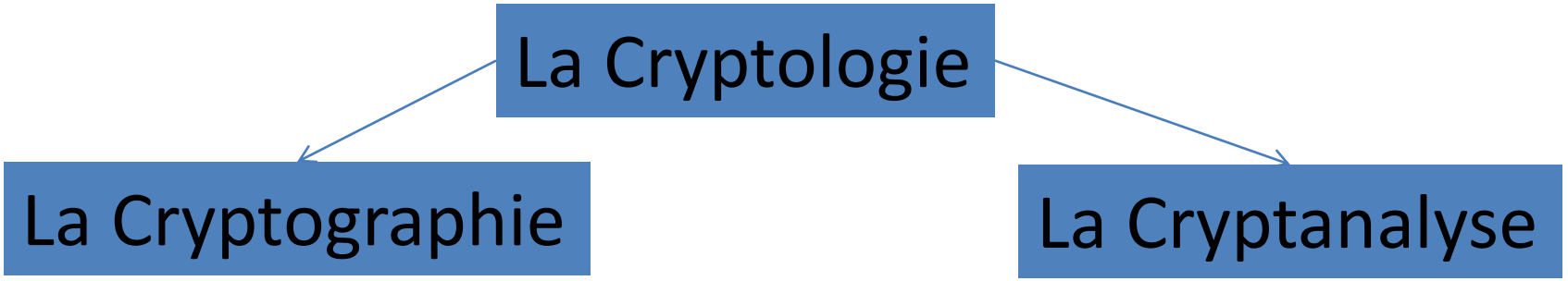
- Depuis fort longtemps, les hommes ont tenté de rendre **sécuritaires** leurs **communications**. Différentes techniques ont été utilisées.
- Au début, il s'agissait seulement de **cacher** l'existence du message. Cette technique s'appelle la **stéganographie**.
- Puis, des techniques de plus en plus **sophistiquées** furent utilisées pour rendre les messages compréhensibles seulement par leurs destinataires légitimes.
- Tout au cours de l'histoire, une difficile bataille eut lieu entre les constructeurs de code (**cryptographes**) et ceux qui essayaient de les briser (les **cryptanalystes**).

Introduction

- Par contre à la **cryptologie**, la stéganographie vise à **dissimuler l'existence** d'un message à transmettre. Au cours de la transmission, personne ne doit savoir qu'un message secret **existe** (ex: encre sympathique).
- Quand à la **cryptologie**, le message à transmettre est codé sous une forme **différente** que **seul** le destinataire peut comprendre.
- Un message en claire est **transformé** en une forme ambigüe, vague et **incompréhensible** par tout le monde. Lors de sa réception, seul le destinataire **légitime** peut renverser le processus et obtenir le message **original** en claire.

Introduction

- La **cryptologie** peut être définie littéralement comme la *science du secret*. Elle se compose de deux grandes branches distinctes:



Étude et **conception** des méthodes et **algorithmes** de **chiffrement** des données claires à transmettre

Analyse des informations chiffrées pour **retrouver** les informations **originales** sans avoir la **permission**

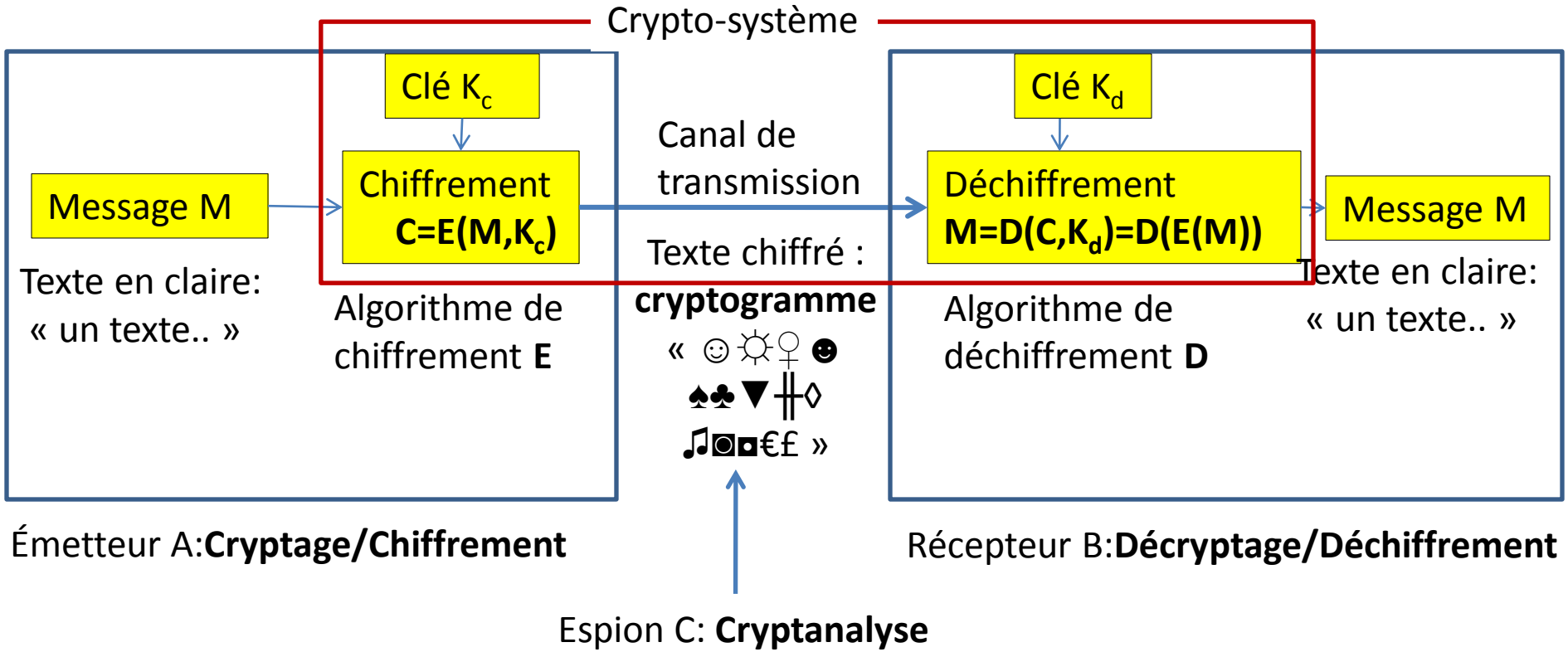
Introduction

- Le mot **cryptographie** est un terme générique désignant l'ensemble des techniques permettant de **chiffrer** des messages, c'est-à-dire permettant de les rendre inintelligibles sans une action spécifique.
- Le fait de coder un message de telle façon à le rendre secret s'appelle **chiffrement**. La méthode inverse, consistant à retrouver le message original, est appelée **déchiffrement**.
- Le résultat de chiffrement (le message chiffré) est appelé **cryptogramme** (en anglais ***ciphertext***) par opposition au message initial, appelé message en **clair** (en anglais ***plaintext***) ;

Terminologie

- Le scénario classique de la cryptographie est qu'un émetteur A veut envoyer un message M à un destinataire B:
- **Texte clair « M »**: information que A souhaite transmettre a B; Ex : texte en français, donnée numérique, etc...
- **Chiffrement** : processus de transformation du message M de telle manière à le rendre incompréhensible
  - ✓ Basé sur une **fonction** de **chiffrement « E »**
  - ✓ 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 **bijectives**)

Terminologie



- Le chiffrement / Déchiffrement se fait généralement par une **information commune** entre l'émetteur et le récepteur (secret commun) appelé **clé** de chiffrement (en anglais: **Key**), sa nature dépend du type du chiffrement (voir plus loin)

Buts de la cryptographie

- La cryptographie est **traditionnellement** utilisée pour dissimuler des messages à certains utilisateurs. Cette utilisation a aujourd'hui un intérêt d'autant **plus grand** :

1. **Confidentialité des informations stockées/manipulées**:

- Le fait de s'assurer que l'information n'est seulement accessible qu‘à ceux dont l'accès est autorisé.
- Algorithmes cryptographiques

  - Chiffrement symétrique : chiffrement par blocs ou par flots.

  - Chiffrement asymétrique.

Buts de la cryptographie

2. Intégrité des informations stockées/manipulées

- L'information ne doit subir aucune altération ou destruction volontaire ou accidentelle, lors de son traitement, conservation ou transmission.
- Détection de l'altération de l'information par une entité non autorisée.
- Algorithmes cryptographiques

  - Fonctions de hachage

Buts de la cryptographie

3. Authentification d'utilisateurs/de ressources

- Vérification de l'identité d'une entité, Authentification de l'origine de l'information et son intégrité.
- Algorithmes cryptographiques

  - Signature numérique

  - Fonction de hachage

Buts de la cryptographie

4. Non-répudiation des informations

- Utilisation d'algorithmes de signatures pour empêcher un utilisateur de se dédire (l'auteur d'un acte ne peut nier l'avoir effectué).
- Empêcher qu'une entité réfute des actions ou engagements antérieurs.
- Vérifier que l'expéditeur et le destinataire sont les entités qui ont envoyé ou reçu l'information.
- Algorithmes cryptographiques

  - Signature numérique.

  - Fonctions de hachage.

Sécurité selon Shannon

- Shannon à définie certains **propriétés** qu'un crypto-système doit respecté pour être plus sécurisé :
- **Confusion :** rendre la relation entre la **clé de chiffrement** et le **texte chiffré** la plus complexe possible par une Substitution ou transformation non-linéaire des symboles du message claire M.
- **Diffusion : la redondance statistique** dans un texte en clair est dissipée dans les statistiques du texte chiffré par Permutation/transposition ou transformations linéaires, « *Modifier un bit du texte clair ou de la clé secrète doit engendrer la modification de chaque bit du texte chiffré avec une probabilité ½ »*(***effet avalanche***)

Classes des crypto-systèmes

- On peut classé les algorithmes de cryptographie selon plusieurs critères :

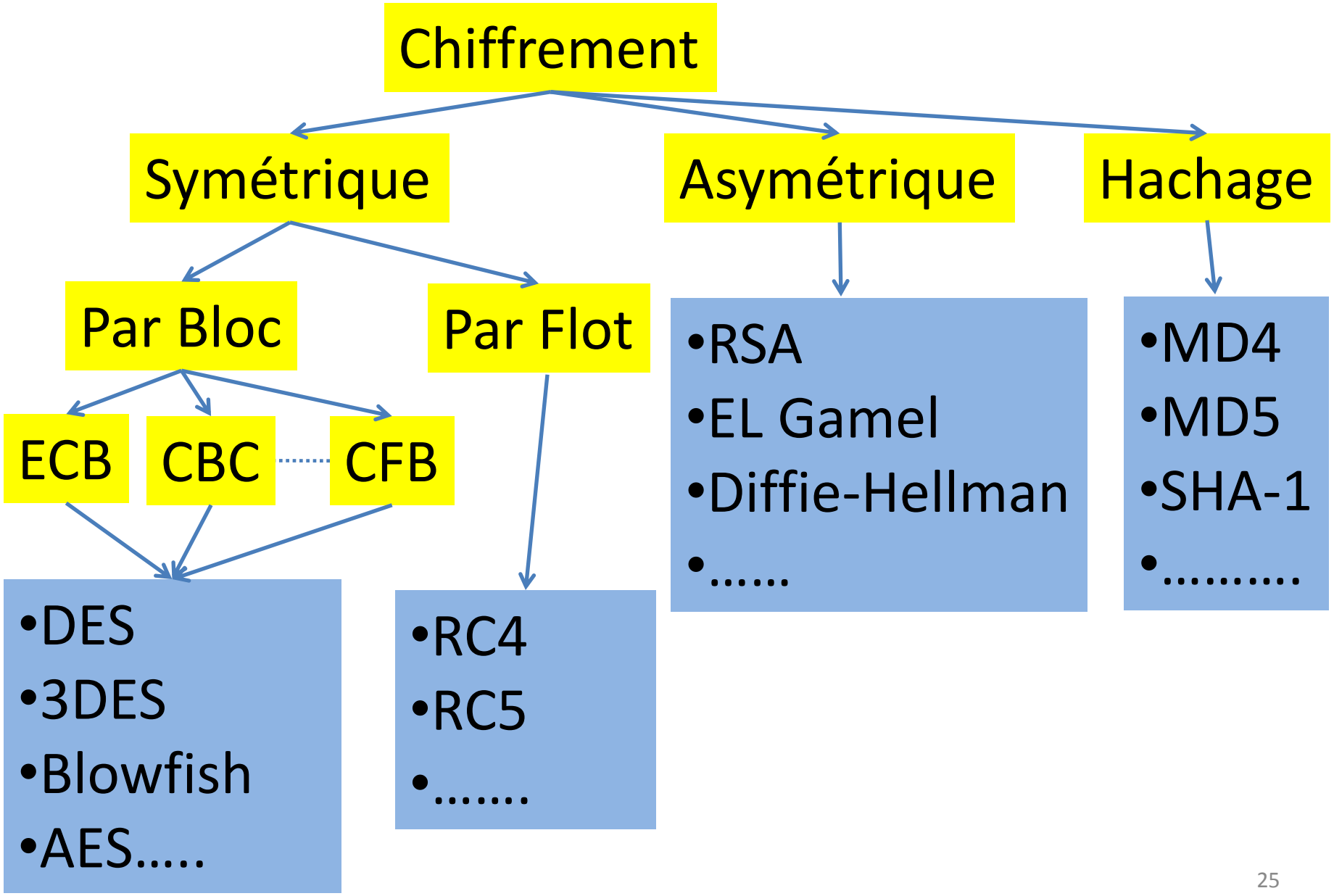
Selon le mode d'utilisation de la clé :

- **Chiffrement symétrique**: la même clé est utilisé pour le chiffrement et le déchiffrement;
- **Chiffrement asymétrique**: deux clés sont utilisées , l'une pour chiffré (clé publique) et l'autre pour déchiffré (clé privé).

Selon le mode d'opération :

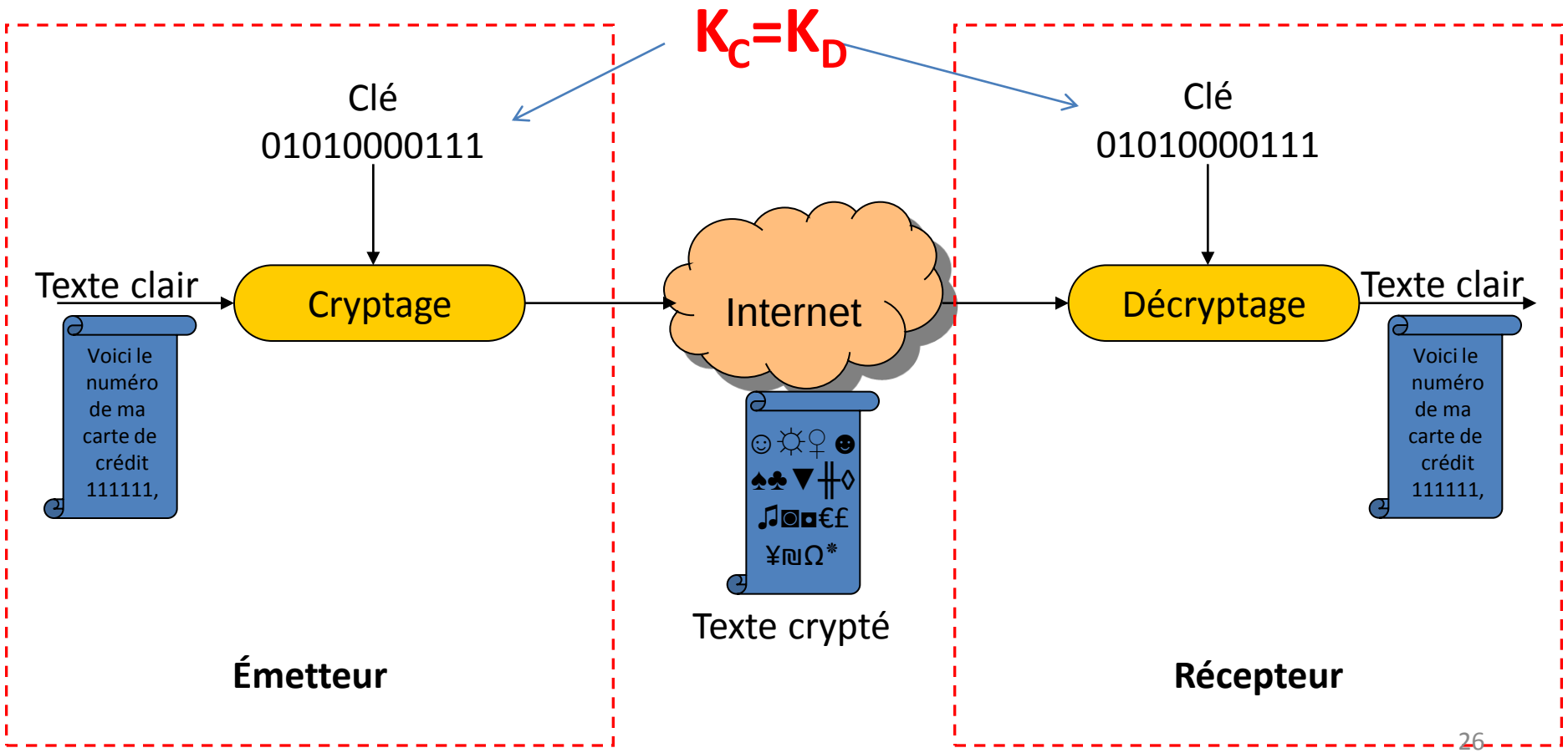
- **Chiffrement par Bloc**: le texte en claire est divisé en blocs de taille identiques
- **Chiffrement par flot**: le texte est considéré comme un flot de bits (chiffrement par bits)

Classes des crypto-systèmes



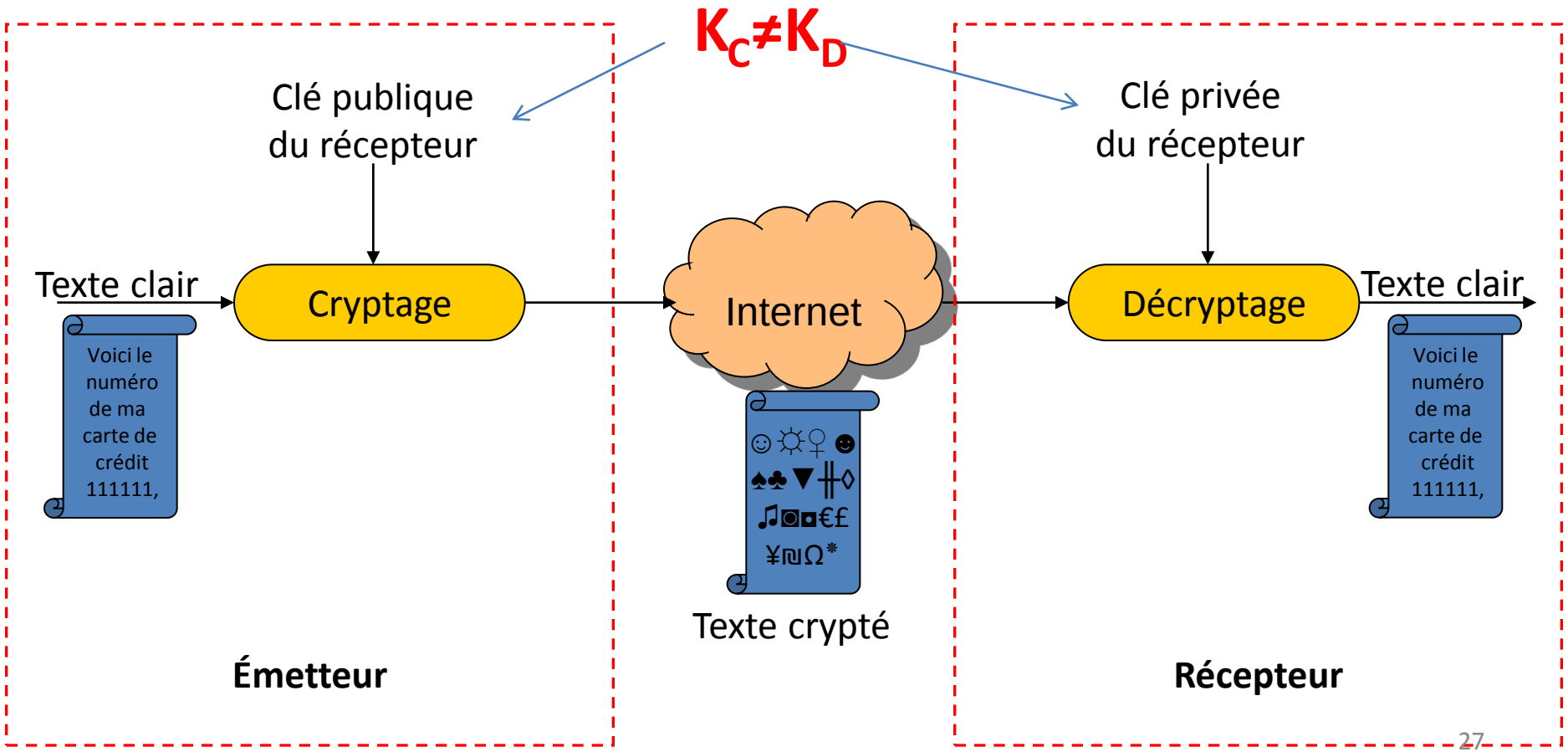
Crypto-systèmes Symétrique: clé secrète

- La même clé est utilisé pour le chiffrement et pour le déchiffrement. L'échange de la clé doit se faire sur un canal sécurisé. La sécurité repose totalement sur la confidentialité de la clé.



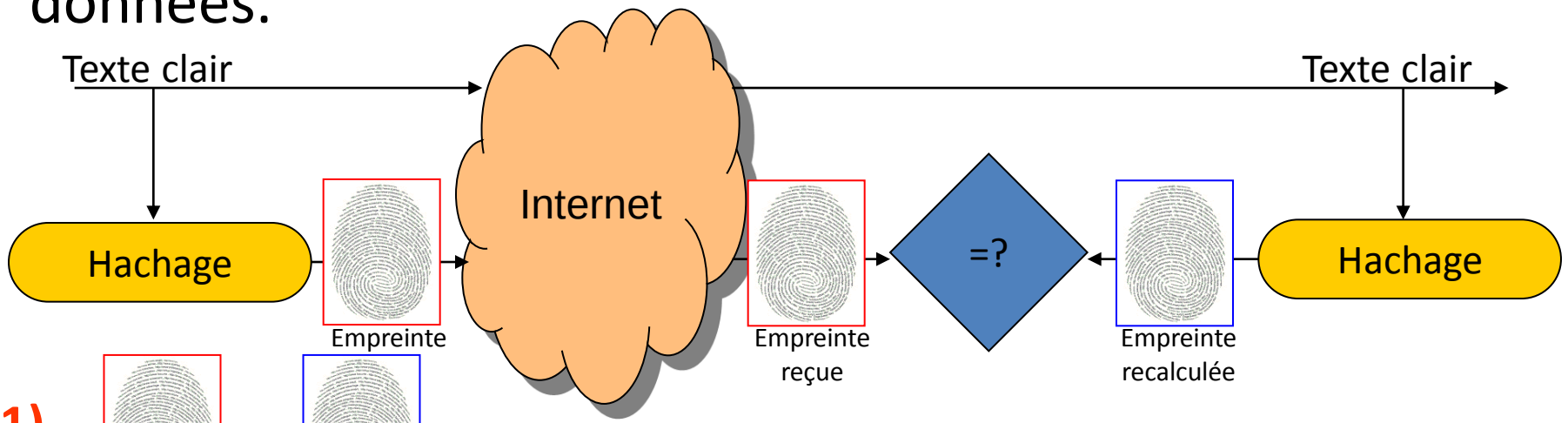
Crypto-systèmes Asymétrique: clé publique

- La clé du chiffrement diffère de celle du déchiffrement, la première est publique (connue par tout le monde), et la deuxième est secrète.



Fonctions de Hachage

- Fonctions à sens unique : pour un entier x, il est simple de calculer H(x), mais étant donné H(x), il est pratiquement impossible de déterminer x. La fonction de hachage permet d'extraire une empreinte de taille fixe qui caractérise les données.



1) Empreinte reçue = Empreinte recalculée — **Le texte reçu est intègre**

2) Empreinte reçue ≠ Empreinte recalculée — **Le texte reçu est altéré**

Cryptanalyse: Définition et méthodes

Définition:

- On appelle **cryptanalyse** est la science qui consiste à décrypter un message chiffré, c'est-à-dire tenter de déchiffrer ce message sans posséder la clé de chiffrement.
- Ainsi, tout crypto-système doit nécessairement être résistant aux méthodes de cryptanalyse.
- Lorsqu'une méthode de cryptanalyse permet de déchiffrer un message chiffré à l'aide d'un crypto-système, on dit alors que l'algorithme de chiffrement a été « **cassé** ».
- Les techniques de cryptanalyse **exploitent des faiblesses** dans l'algorithme de chiffrement pour pouvoir casser le crypto-système, et déchiffrer le message sans avoir besoin de la clé, ou bien déduire la clé à partir seulement du message chiffré.

Historique de la cryptographie

- Historiquement , la plupart des méthodes de chiffrement reposent sur deux principes essentiels :
  - **la substitution**
  - **la transposition (permutation)**
- **Substituer** signifie qu'on remplace certaines lettres par d'autres, ou par des symboles.
- **Transposition** signifie qu'on permute les lettres du message afin de le rendre inintelligible.
- Ces deux approches on été combinées pour crée la majorité des méthodes (algorithmes) de chiffrement/ déchiffrement à travers l'histoire.

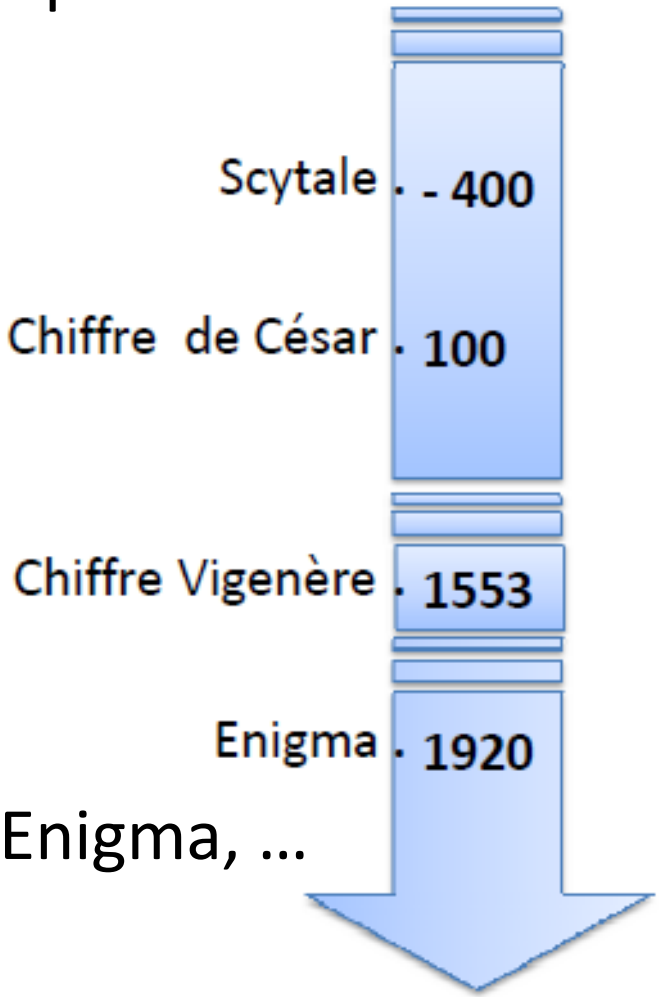
Historique de la cryptographie

La cryptographie est utilisée depuis l'antiquité

– Il y a 4000 ans par les Égyptiens

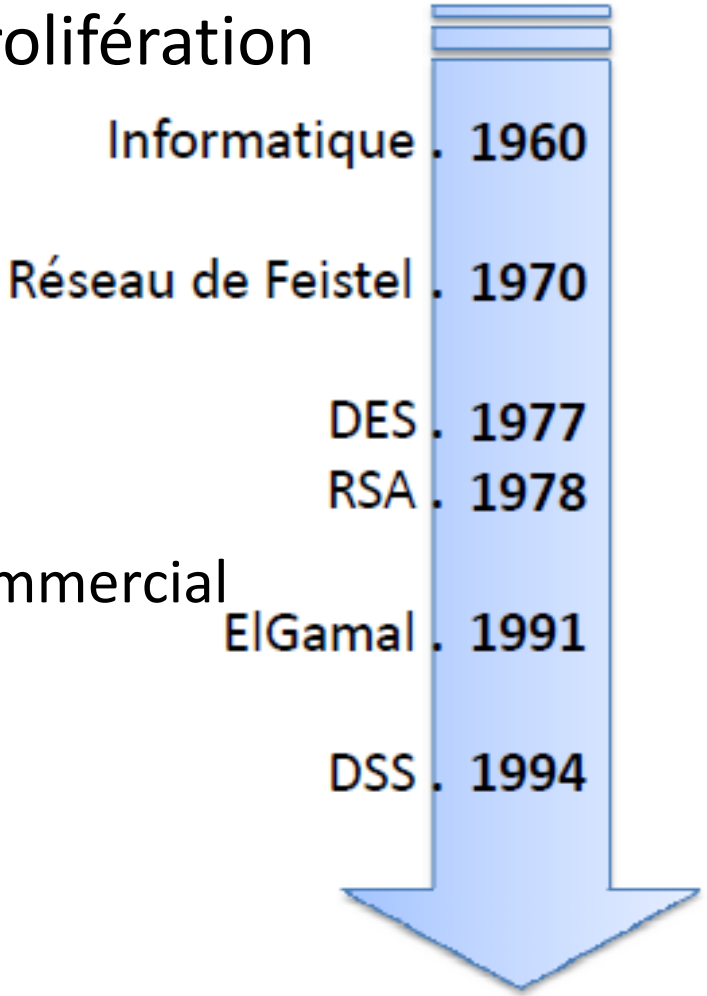
- Cryptographie ancienne
  - – Alphabet de la langue

    Ex: Français : 26 lettres
- Chiffrement de documents
  - – Domaine militaire et diplomatique
  - – Chiffrement symétrique
- Chiffre de César, de Vigenère, Scytale, Enigma, …



Historique de la cryptographie

- Cryptographie moderne
  - L'apparition de l'informatique et prolifération des systèmes de communication
  - Alphabet = {0, 1}
  - Protection de l'information numérique
    - Domaine militaire, diplomatique, commercial
    - Protection de la vie privée
  - Chiffrement symétrique
  - Chiffrement asymétrique
    - Distribution de clés
    - Signature numérique



la scytale

- 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.

Chiffrement de Polybe

- L'historien grec **Polybe (150 av. J.-C.)** est à l'origine du premier procédé de chiffrement par **substitution**. C'est un système de transmission basé sur un carré de 25 cases.

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

- En français, on supprime le W, qui sera remplacé par V. Il existe une variante où ce sont I et J qui se partagent la même case.
- Chaque lettre peut être ainsi représentée par **un groupe de deux chiffres** : celui de sa ligne et celui de sa colonne. Ainsi e=(1;5), u=(4;5), n=(3;3)...

Chiffrement de César

- **60-50 av. J.-C.** **Jules César** décale les lettres de l'alphabet d'une quantité fixe dans les communications du gouvernement.
- Le chiffre de César est une des plus simples méthodes de cryptage connues.
- C'est une technique de codage par **substitution**, c'est-à-dire que chaque lettre du texte en clair est remplacée par une autre lettre à distance **fixe** dans l'alphabet.
- Par exemple, si l'on utilise un décalage de 3, A serait remplace par D, B deviendrait E, et ainsi de suite. Cette méthode doit son nom a Jules César, qui utilisait cette technique pour certaines de ses correspondances.

Chiffrement de César

- Cette technique de chiffrement est-elle sécuritaire?

  On intercepte par exemple le message

  FAGEMYREMPURZV_EMZR_R FMNMDAZR

- Essayons différents décalages…

  1: **E_FDLXQDLOTQYUZDLYQZQZELMLC_YQ**

  2: **DZECKWPCKNSPXTYCKXPYPYDKLKBZXP**

  3… 4… 5… 6… 7… 8… 9… 10… 11… 12…

  13: **TOUS_LES_CHEMINS_MENENT_A_ROME**

- Après 13 essais, le message est parfaitement déchiffré sans avoir au préalable la valeur du décalage.
- Clairement, le chiffrement de César **n'est pas sécuritaire**.

Substitution mono-alphabétique

- Vue cette faiblesse, une amélioration a été apportée: Essayons autre chose:

  _ 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

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

- Chaque lettre est remplacer par une autre prédéfinie (table de correspondance).

  TOUS_LES_CHEMINS_MENENT_A_ROME devient

  FQLJRPAJRHCAE_ZJREAZAZFRDRNQEA

- Ce type de codage est appelé **substitution mono-alphabétique**. Le décodage devrait être plus difficile. Peut-on essayer tous les décodages possibles?
- Il y a 26!=403291461126605635584000000 possibilités…

Substitution mono-alphabétique

- Le premier usage révélé de chiffrement par substitution dans un usage militaire est rapporté par Jules César dans **la guerre des Gaules**. César utilisait fréquemment ce type de chiffrement.
- La substitution mono-alphabétique fut la technique de chiffrement **la plus utilisée** durant le **premier millénaire**.
- Nombreux savants de l'antiquité tenaient cette technique pour inviolable.
- Ce sont les **Arabes** qui réussirent à **briser ce code** et qui inventèrent la cryptanalyse au 9ième siècle (**Al-Kindi**), La technique est appelée **analyse des fréquences** rédigées dans un traité intitule ≪ *Manuscrit sur le déchiffrement des messages cryptographiques* ≫. رسالة الكندي في كشف المعمى

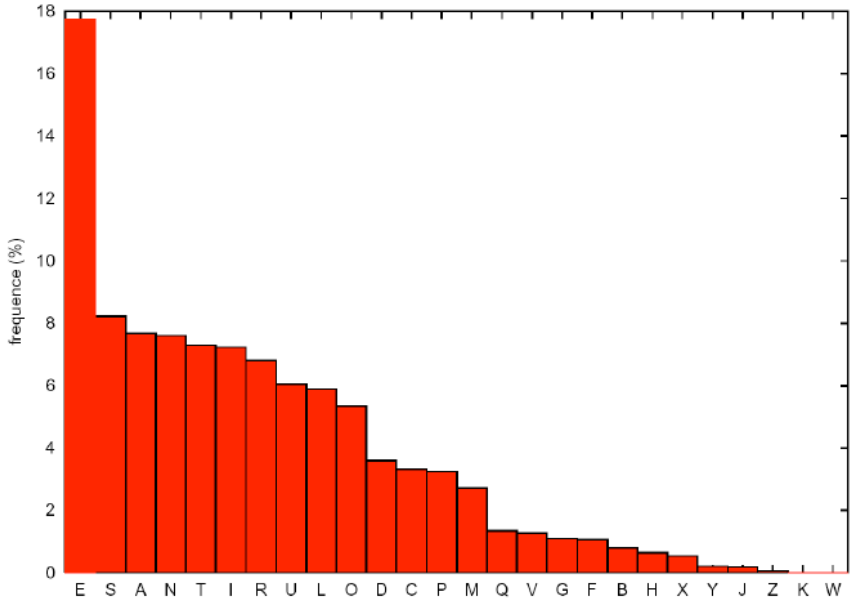
Substitution mono-alphabétique

- Par exemple soit le texte chiffré suivant:

nvxlbgi avxw n ctxnbw ubn dvttbn r bhxqacyb awbggbgi rbn cueciwv lcnibn vqnbcxz rbn tbwn hxq nxqlbgi qgrvubgin mvtacygvgn rb lvscyb ub gclqwb yuqnncgi nxw ubn yvxoowbn ctbwn c abqgb ubn vgi qun rbavnbn nxw ubn aucgmdbn hxb mbn wvqn rb u ckxw tcucrwvqin bi dvgibxz ucqnnbgi aqibxnbtbgi ubxwn ywcgrbn cqubn eucgmdbn mvttb rbn clqwvgn iwcqgbw c mvib r bxz mb lvscybxw cqub mvttb qu bni ycxmdb bi lbxub uxq gcyxbwb nq ebcx hx qu bni mvtqhxb bi ucqr u xg cycmb nvg ebm clbm xg ewxubyxbxub u cxiwb tqtb bg evqicgi u qgoqwtb hxq lvucqi ub avbib bni nbteuceub cx awqgmb rbn gxbbn hxq dcgib uc ibtabib bi nb wqi rb u cwmdbw bzqub nxw ub nvu cx tquqbx rbn dxbbn nbn cqubn rb ybcgi u btabmdbgi rb tcwmdbw

Substitution mono-alphabétique

- En utilisent les fréquences des lettres en français, on remarque que :



Dans le **chiffré :**

B	N	C	U	X	Q	G	I	W	V
18,7	9,91	7,78	6,90	6,72	6,37	5,84	5,84	5,30	4,60

En **français :**

E	S	A	N	T	I	R	U	L	O
17,8	8,23	7,68	7,61	7,30	7,23	6,81	6,05	5,89	5,34

Substitution mono-alphabétique

- On déduit donc que : B→E, N →S, C →A, ce qui donne :

svxlegi avxw s atxsew ues dvttes r ehxqaaye aweggegi res aueaiwvs lasies vqseaxz res tews hxq sxqlegi qgrvuegis mvtaaygvgs re lvsaye ue galqwe yuqssagi sxw ues yvxoowes atews a aeqge ues vgi qus reavses sxw ues auagmdes
hxe mes wvqs re u akxw tauarwvqis ei dvgiexz uaqssegi aqiexsetegi uexws ywagres aques euagmdes mvtte res alqwvgs iwaqgew a mvie r exz me lvsayexw aque mvtte qu esi yaxmde ei lexue uxq gayxewe sq eeax hx qu esi mvtqhxe ei uaqr u xg ayame svg eem alem xg ewxueyxexue u axiwe tqte eg evqiagi u qgoqwte hxq lvuaqi ue aveie esi seteuaeue ax awqgme res gxees hxq dagie ua ietaeie ei se wqi re u awmdew ezque sxw ue svu ax tquqex res dxees ses aques re yeagi u etaemdegi re tawmdew

Substitution mono-alphabétique

- On peut utiliser ensuite les statistiques sur les bigrammes:

Bigrammes les plus fréquents dans le chiffré :

ES	U**E**	GI	R**E**	**E**G	**E**X	I**E**	**SE**	QU	T**E**	U**A**	**E**W	**A**G	**A**Q	HX
25	17	13	12	9	8	8	8	8	8	8	7	7	7	7

Bigrammes les plus fréquents en français :

ES	L**E**	**E**N	D**E**	R**E**	NT	ON	**E**R	T**E**	**SE**	**E**T	**E**L	QU	**A**N	N**E**	OU	**A**I

- On peut déduire que : U$\rightarrow$ L, R $\rightarrow$ D, G $\rightarrow$ N, Q $\rightarrow$ I

Bigrammes les plus fréquents dans le chiffré (après subst.):

ES	LE	NI	DE	EN	EX	IE	SE	IL	TE	LA	EW	AN	AI	HX
25	17	13	12	9	8	8	8	8	8	8	7	7	7	7

- On peut déduire que: I $\rightarrow$ T

Substitution mono-alphabétique

- Le texte chiffré devient après substitution :

svxlent avxw s atxsew les dvttes d ehxiaaye awennent des aleatwvs lastes viseaxz des tews hxi sxilent indvlents mvtaaynvns de lvsaye le naliwe ylissant sxw les yvxoowes atews a aeine les vnt ils deavses sxw les alanmdes hxe mes wvis de l akxw taladwvits et dvntexz laissent aitexsetent lexws ywandes ailes elanmdes mvtte des aliwvns twainew a mvte d exz me lvsayexw aile mvtte il est yaxmde et lexle lxi nayxewe si eeax hx il est mvtihxe et laid l xn ayame svn eem alem xn ewxleyxexle l axtwe tite en evitant l inoiwte hxi lvlait le avete est setelaele ax awinme des nxees hxi dante la tetaete et se wit de l awmdew ezile sxw le svl ax tiliex des dxees ses ailes de yeant l etaemdent de tawmdew

- Quelque mots apparaissent :

indvlent, vnt (V → O); oiseaxz (X → U, Z → X); a aeine (A → P) ; leuws (W → R); taladroits (T → M); yrandes (Y → G)

Substitution mono-alphabétique

- Le texte devient donc:

soulent pour s amuser les dommes d ehuipage prennent des aleatros lastes oiseaux des mers hui suilent indolents mompagnons de losage le nalire glissant sur les gouoores amers a peine les ont-ils deposes sur les planmdes hue mes rois de l akur maladroits et donteux laissent piteusement leurs grandes ailes elanmdes momme des alirons trainer a mote d eux me losageur aile momme il est gaumde et leule lui naguere si eeau hu il est momihue et laid l un agame son eem alem un erulegueule l autre mime en eoitant l inoirme hui lolait le poete est semelaele au prinme des nuees hui dante la tempete et se rit de l armder exile sur le sol au milieu des duees ses ailes de geant l empemdent de marmder

- On peut facilement continué le processus et trouver le texte complet (ex: L $\rightarrow$ V, D $\rightarrow$ H, h $\rightarrow$ Q…..)

Substitution Poly-alphabétique

- L’espace des clés du chiffrement mono-alphabétique est immense, mais le fait qu’une lettre soit toujours cryptée de la même façon représente une trop grande faiblesse.
- La substitution poly-alphabétique consiste à substituer une lettre du message en clair, par une autre choisie en fonction d'un **état du crypto-système**, et non plus de manière fixe comme pour la mono-substitution.
- Ce changement de lettre tout au long du processus, s'obtient à l'aide d'une clé, qui indique le nombre de décalage à réaliser à ce moment. Pour chiffrer la lettre suivante on utilise alors le caractère suivant de la clé et ainsi de suite.

Chiffrement de Vigenère

- Au 16ième siècle, **Blaise de Vigenère** (1523-1596), inventa un système de chiffrement **poly-alphabétique**. Il s'agit d'une amélioration du chiffre par décalage.
- Vigenère est le premier à avoir introduit la notion de **clé**, on choisit un **mot de code** on l'utilise pour chiffrer. Il est répété autant de fois que la taille du texte clair, ensuite chaque lettre du texte est décalée en fonction de la valeur numérique correspondante au symbole de la clé associe.
- Exemple : clé=ALAIN={0,11,0,8,13}

Clair: **LE_CODE_DE_VIGENERE_EST_IL_INDECHIFFRABLE**

AL_AINA_LA_INALAINA_LAI_NA_LAINALAINALAIN

Chiffré: **LP_CWQE_OE_DVGPNMEE_PSB_VL_TNLRCSINSRLBTR**

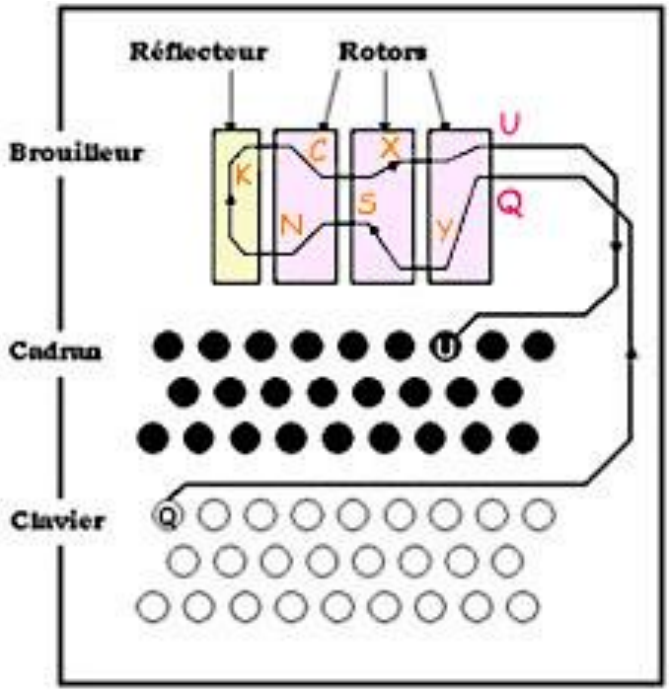
Chiffrement de Vigenère

- Mathématiquement, on considère que les lettres de l'alphabet sont numérotées de 0 à 25 (A=0, B=1 ...). La transformation lettre par lettre se formalise simplement par : Chiffré[i] = (Texte [i]+ Clé[i]) **modulo** 26
- (Texte + Clé) modulo 26 correspond au « **reste de la division entière** de (Texte + Clé) par 26 »
- Remarquez que si l'on utilise la clé avec un texte **rempli uniquement** avec des A on retrouve assez facilement la clé
- « A » + LettreInconnue = LettreInconnue, soit du point de vue mathématique : 0 + x = x.
- Le chiffre de Vigenère est-il indéchiffrable?

La Machine ENIGMA

- La cryptologie a joué un rôle **décisif** pendant la Seconde Guerre mondiale.
- La guerre à permis une grande évolution de l'art de la cryptographie. Plusieurs techniques ont été élaborés, dont la plus fameuse est la machine **ENIGMA**.
- C'est une machine conçue par les Allemands pour chiffrer leurs messages. Cette machine peut être considérée comme la **première machine électromagnétique** traite de l'information, elle a permis de lancer l'informatique après la guerre à travers les travaux de Alain Turing.

La Machine ENIGMA



La Machine ENIGMA

- Brièvement, la machine **Enigma** chiffre les informations en réalisant le passage d'un **courant électrique** à travers une série de composants.
- Le courant est transmis en pressant une **lettre** sur le clavier. Après sa traversée dans un réseau **complexe** de fils, une **lampe indique la lettre chiffrée**. Le premier composant est une série de roues adjacentes, appelées « **rotors** », qui contiennent les fils électriques utilisés pour coder le message. Les rotors tournent, variant la configuration complexe du réseau chaque fois qu'une lettre est tapée. La machine Enigma utilise habituellement une autre roue, nommée « **réflecteur** », et un composant, appelé **pupitre de connexion**, permettant de complexifier encore plus le processus de chiffrement.
- **Simulateur:** http://enigmaco.de/enigma/enigma.html

Chiffrement Affine

- On dit qu'une fonction est affine lorsqu'elle est de la forme $x \rightarrow a*x + b$, c'est-à-dire un polynôme de degré 1.
- L'idée est d'utiliser comme fonction de chiffrement une fonction affine du type $y = (a*x + b) \bmod 26$, où a et b sont des constantes, et où x et y sont des nombres correspondant aux lettres de l'alphabet (A=0,B=1,...). On peut remarquer que si a = 1, alors on retrouve le chiffre de César où b est le décalage (le k du chiffre de César).
- Pour le chiffre affine, la clé est constituée de (a, b) où $a, b \in [0, 25]$ et telle que pgcd(a, 26) = 1.
- Le chiffrement est donné par:

$$c_i = f(m_i) = (a * m_i + b) \bmod 26.$$

- Pour le déchiffrer : $m_i = f^{-1}(c_i) = a^{-1}*(c_i - b) \bmod 26$.

Chiffrement Affine

- Par le chiffre affine, on obtient 312 clés possibles. En effet, pour obéir à la propriété de a, il n'y a que 12 choix possibles. Et puisque b peut prendre n'importe valeur dans [0, 25], il vient 12*26 = 312.
- Le chiffrement affine n'est pas du tout sécurisé, car le nombre de clé est très réduit et peut être facilement cassé par une attaque brute. Il est généralement utilisé pour des fins pédagogiques.

Conclusion

- Les algorithmes de cryptographie sont trop nombreux (par milliers), mais ceux qui sont vraiment fiables et sécurisés sont vraiment trop peu.
- La sécurité d'un algorithme de chiffrement repose sur plusieurs facteurs tels la taille de la clé, la distribution statistique du texte chiffré, et la complexité algorithmique de calcul…
- la résistance à un type particulier d'attaques n'exclut pas la vulnérabilité à d'autres attaques possibles.
- Dans les cours qui suivent, on essayera d'étudier les principaux algorithmes de chiffrement moderne, leurs points forts et faibles, leurs **vulnérabilités** ainsi que leurs complexités algorithmiques.