# Techniques Cryptographiques, Applications et Sécurité

**Notes de Cours de Master Pro en Cybersécurité et Gouvernance Sécuritaire**

**TAYOU DJAMEGNI Clémentin**
**Professeur**

**Chapitre 1 : Introduction**
1.1 Motivation
1.2 Différents modèles de sécurité
1.3. Ressources bibliographiques

**Chapitre 2 :** Introduction à la cryptographie
2.1 Vocabulaire de base
2.2 Notations
2.3 Principe de Kerckhoff
2.4 La publication des algorithmes
2.5 Les principaux concepts cryptographiques
- 2.5.1 Cryptosystème à clé symétrique
- 2.5.2 Cryptosystème à clé publique
- 2.5.3 Fonction de hachage
- 2.5.4 Protocoles cryptographiques
  - 2.5.4.1 Confidentialité
  - 2.5.4.3 Authentification
  - 2.5.4.4 Synthèse

**Chapitre 3 : La cryptographie classique**
3.1 Substitution monoalphabétique
- 3.1.1. Chiffre de César
- 3.1.2. Analyse de fréquences
- 3.1.3. Chiffre affine
  - 3.1.3.1 Cryptanalyse du chiffre affine

3.2 Substitution homophonique : Le Disque de l'Armée Mexicaine
3.3 Chiffrement polygraphique
- 3.3.1. Chiffre de Playfair (1854)
- 3.3.2. Chiffre de Hill (1929)

3.4 Substitutions polyalphabétiques
- 3.4.1 Chiffre de Vigenère (1568)
  - 3.4.1.1 Cryptanalyse du Commandant Bazeries
  - 3.4.1.2 Cryptanalyse de Kasiski
  - 3.4.1.3 Cryptanalyse de Friedman
- 3.4.2 Chiffre de Beaufort
- 3.4.3 Chiffre de Vernam

3.5 Transpositions
- 3.5.1. La scytale spartiate
- 3.5.2. Chiffre de Ubchi

**Chapitre 4 : Cryptographie moderne**
4.1. Chiffrement Symétrique
- 4.1.1. Chiffrement par blocs
  - 4.1.1.1 Approche de la notion de chiffrement par blocs
  - 4.1.1.2. Réseau de Feistel
  - 4.1.1.3. D.E.S. - Data Encryption Standard
  - 4.1.1.4. A.E.S. - Advanced Encryption Standard
- 4.1.2. Chiffrement par flot

4.2. Chiffrement par clé publique
- 4.2.1. Chiffre de Merkle-Hellman
- 4.2.2. Chiffre de Rivest - Shamir - Adleman (RSA)
- 4.2.3. Chiffre de El Gamal
- 4.2.4. Chiffre de Rabin

4.3. Cryptographie à courbe elliptique

**Chapitre 5 : Thèmes des exposés d'ouvertures**
5.1. Quelques types d'attaque
5.2. La biométrie
5.3. La stéganographie
5.4. Protocoles d'authentification
5.5. IPSEC
5.6. Le monde quantique
5.7. Sécurité logicielle
5.8. Sécurité en entreprise
5.9. Les architectures de paiement électronique
5.10. Le code barre et le code QR

# Chapitre 1 : Introduction

## 1.1 Motivation

- Les défis  en matière de sécurité sont  de plus en plus grandissants
- Ces défis portent sur la sécurité des données, des traitements (ou applications) et du matériel (ordinateurs, réseaux informatiques)
- On assiste à une évolution constante des technologies de sécurisation des systèmes informatiques et des techniques d'intrusion des dits systèmes

Parmi celles-ci, on trouve diverses catégories de menaces :

- **Les menaces accidentelles :** Aucune préméditation. Par exemple les bugs logiciels, les pannes matérielles, et autres défaillances "incontrôlables"
- **Les menaces intentionnelles :** Elles reposent sur l'action d'un tiers désirant s'introduire frauduleusement dans le système informatique.
  – **Attaques passives :** L'intrus va tenter de dérober les informations sans laisser de trace, ce qui rend sa détection relativement difficile. Il ne modifie pas les fichiers, ni n'altère les systèmes.
  – **Attaques actives :** L'intrus laisse des traces. Il  modifie les fichiers ou le système en place pour s'en emparer.

Les menaces actives appartiennent principalement à quatre catégories  :

- **Interruption :** Interruption de service. Ce qui peut provoquer la non-disponibilité des données
- **Interception :**  Interception des données sur la ligne de communication (man in the middle) . Ce qui ôse le problème de confidentialité des données
- **Modification :** Modification des données qui circulent sur la ligne de communication. Ce qui pose le problème d'intégrité des données
- **Fabrication :**  Fabrication et introduction de données sur la ligne de communication. Ce qui pose le problème l'authenticité des données

Les auteurs des attaques :

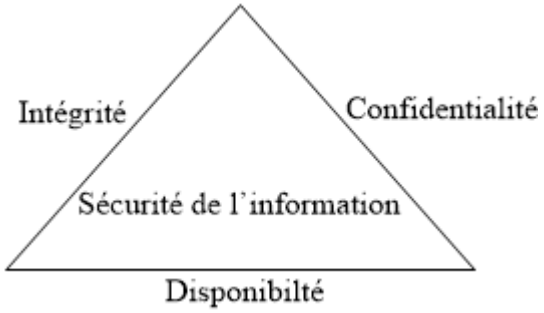
- Les hackers (agissant souvent par défi personnel)
- Les concurrents industriels. Vol d'informations : stratégie de l'entreprise, brevets d'invention, projets
- Les espions, la presse ou encore  les agences nationales .

## 1.2. Différents modèles de sécurité

### 1.2.1. Le Modèle de Sécurité CIA (1987)

CIA intègre trois grands axes de la sécurité. La plupart des autres modèles l'utilisent comme brique de base. Les lettres de CIA sont définies comme suit :

- **Confidentialité :** l'information n'est connue que des entités communicantes
- **Intégrité :** l'information n'a pas été modifiée entre sa création et son traitement (en ce compris un éventuel transfert)
- **Disponibilité :** l'information est toujours accessible et ne peut être bloquée/perdue




Le triangle opposé existe également. Il porte le nom de DAD, pour Disclosure, Alteration, Disruption : Divulgation, Altération et Perturbation.

## 1.2.2. Le contrôle d'accès : Le protocole AAA

Le modèle CIA n'intègre pas explicitement la vérification de l'identité des parties communicantes . Le contrôle d'accès se fait en 4 étapes.

- **Identification :** Qui êtes-vous ?
- **Authentification :** Prouvez-le !
- **Autorisation :** Avez-vous les droits requis ?
- **Accounting/Audit :** Qu'avez-vous fait ?

Dans le protocole AAA, les deux premières étapes sont fusionnées. Dans certaines situations, on scindera la dernière étape. On parlera d'Accounting lorsque le fait de comptabiliser des faits sera demandé, et d'Audit lorsque des résultats plus globaux devront être étudiés.

L'authentification, visant à prouver l'identité d'un individu peut se faire de plusieurs manières :

- **Ce que vous savez :** mot de passe, code PIN, etc.
- **Ce que vous avez :** carte magnétique, lecteur de carte, etc.
- **Ce que vous êtes :** empreintes digitales, réseau rétinien, etc.

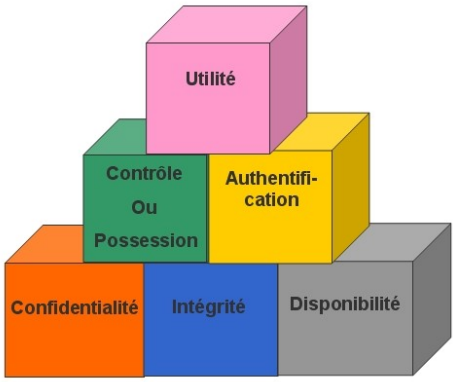
**L'authentification forte** résultera de la combinaison de 2 de ces facteurs.

## 1.2.3. Le pentagone de confiance (2006)

Comparé à CIA, il manque la confidentialité et l'intégrité. Comparé à AAA, il manque l'audit/accounting.
**L'admissibilité :** La machine sur laquelle nous travaillons, à laquelle nous nous connectons, est-elle fiable ? En d'autres termes, peut-on faire confiance à la machine cible ?


## 1.2.4. L'hexagone de Parker (2002)

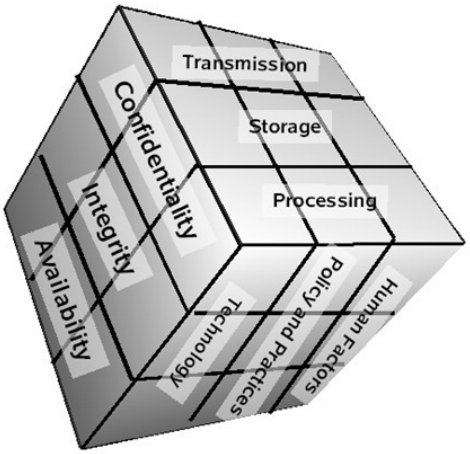
Il ajoute la nuance **d'utilité** (une information chiffrée pour laquelleon a perdu la clé de déchiffrement n'est plus d'aucune utilité, bien que l'utilisateur y ait accès, que cette information soit confidentielle, disponible et intègre)



### 1.2.5 Le Cube (1991)

On y retrouve les trois piliers de la sécurité (CIA), mais deux autres dimensions apparaissent :

- **L'état des données :** le stockage, la transmission, l'exécution
- **Les méthodes :** les principes et règles à adopter pour atteindre le niveau de sécurité souhaité




### 1.2.5 Autres définitions

**La sécurité en parallèle**

Dans ce modèle, plusieurs mécanismes de sécurité protégeant un système possèdent le même rôle. Le niveau de protection du système est équivalent à celui du mécanisme le moins sûr. Par exemple, on peut déverrouiller un ordinateur portable soit par un mot de passe, soit par la biométrie (empreinte digitale, ).

**La sécurité en série**

Dans ce modèle, plusieurs mécanismes de sécurité qui protègent un système et ont des rôles complémentaires. On parlera de « **défense en profondeur** ». Citons par exemple le réseau d'une entreprise où :

- Le réseau est sécurisé par un Firewall hardware,
- Les liaisons entre machines sont protégées,
- Les machines individuelles sont munies d'un Firewall software,
- Les accès aux machines se font par empreinte biométrique,
- Le logiciel à utiliser est accessible par mot de passe,
- etc.

### 1.3. Ressources bibliographiques


http://nomis80.org/cryptographie/cryptographie.html
http://www.01adfm.com/win-xp/hacking/
http://www.apprendre-en-ligne.net/crypto/menu/index.html
http://perso.club-internet.fr/guidovdi/codes/lapagecryptologie.htm
http://www.pro-technix.com/information/crypto/pages/vernam_base.html
http://www.chez.com/nopb/crypto2.html#transposition

http://www.labri.fr/Perso/~betrema/deug/poly/premiers.html
http://www.cryptosec.org/article.php3?id_article=10
http://members.tripod.com/irish_ronan/rsa/attacks.html

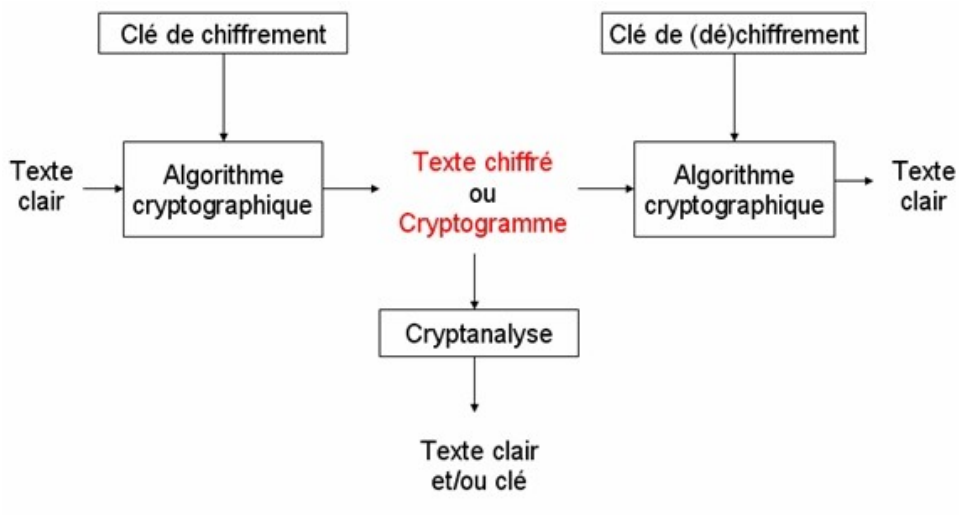
http://www.bibmath.net/crypto/moderne/sigelec.php3
http://www.ssh.fi/support/cryptography/introduction/signatures.html

http://home.ecn.ab.ca/~jsavard/crypto/pk0503.htm
http://home.ecn.ab.ca/~jsavard/crypto/mi0607.htm
http://developer.netscape.com/docs/manuals/security/pkin/contents.htm
http://www.itl.nist.gov/

http://www.hsc.fr/ressources/articles/
http://web.mit.edu/tytso/www/ipsec/index.html

# Chapitre 2 : Introduction à la cryptographie

## 2.1 Vocabulaire de base




**Cryptologie :** Science mathématique comportant deux branches : la cryptographie et la cryptanalyse
**Cryptographie :** Etude des méthodes d'envoie des données de manière confidentielle sur un support.
**Chiffrement:** Il consiste à transformer une donnée (texte, message, ...) afin de la rendre incompréhensible à une tierce personne.
**Déchiffrement :** Il consiste à retrouver le texte clair à partir du texte chiffré
**Texte chiffré :** Appelé également cryptogramme, le texte chiffré est le résultat de l'application d'un chiffrement à un texte clair.
**Clef :** Elle est utilisée pour effectuer des chiffrements/déchiffrements. Dans le cas d'un algorithme symétrique, la clef est identique lors des deux opérations. Dans le cas d'algorithmes asymétriques, elle diffère pour les deux opérations.
**Cryptanalyse :** Opposée à la cryptographie, elle a pour but de retrouver le texte clair à partir de textes chiffrés en exploitant les failles/faiblesses des algorithmes/méthodes de chiffrement.
**Cryptosystème :** Il est défini comme l'ensemble des clés possibles (espace de clés), des textes clairs et chiffrés possibles associés à un algorithme donné.

Le triplet suivant permettrait de garantir la confidentialité :

- Un algorithme générer aléatoirement une clé K
- Un deuxième algorithme pour chiffrer un message M en C
- Un troisième algorithme pour déchiffrer C.

**Remarques :**

- **Décryptage :** Action permettant de retrouver le texte clair sans connaître la clef de déchiffrement.
- **Cryptage :** Action de chiffrer un message.
- **Encryptage et Encryptement :** Ce sont des anglicismes dérivés du verbe "to encrypt"

## 2.2 Notations

En cryptographie, la propriété de base est que le déchiffrement d'un chiffrement de M doit être égal à M. Le déchiffrement doit permettre de retrouver le message initialement chiffré :

$$M = D(E(M))$$

où
– M représente le texte clair,
– C est le texte chiffré,
– K est la clé (dans le cas d'un algorithme à clé symétrique)
– E est la fonction de chiffrement (E(x) désigne le chiffrement, x)
– D est la fonction de déchiffrement (D(x) désigne le déchiffrement de x) .

### 2.3 Principe de Kerckhoff

**La sécurité du chiffrement ne doit pas dépendre de ce qui ne peut pas être facilement changé**

Ce principe implique **aucun secret ne doit résider dans l'algorithme mais plutôt dans la clé**. Sans celle-ci, il doit être impossible de retrouver le texte clair à partir du texte chiffré. Par contre, si on connaît K, le déchiffrement est immédiat.

Il faut distinguer les termes "Secret" et "Robustesse" d'un algorithme.

- **Garder un algorithme secret :** Cacher des concepts et ses operations (fonctions mathématiques).
- **Un algorithme robuste :** Un algorithme qui résiste à diverses attaques .

## 2.4 La publication des algorithmes

Selon l'endroit où réside le secret, dans les opérations de l'algorithme ou dans la clé, on peut parler d'algorithme secret ou d'algorithme publié. Chacun possède ses atouts et inconvénients.

**Algorithme secret**

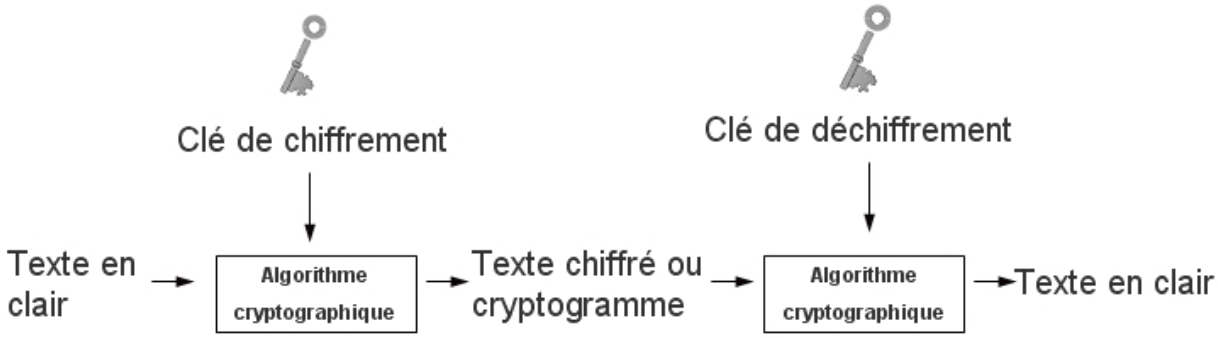
- **Nécessité de protéger la sémantique de l'algorithme contre le reverse-engineering :** Le reverse-engineering est une technologie permettant de retrouver les mécanismes et opérations d'un algorithme à la suite d'une série d'exécution du dit algorithme.
- **La cryptanalyse est olus difficile :** Non seulement, il faut retrouver la clé secrète, il faut aussi retrouver toutes les opérations/traitements/mécanismes secrets de l'algorithme.
- **Petit nombre d'utilisateurs :** Moins il y a de monde l'utilisant, moins il y a d'intérêts à le casser, moins l'algorithme est éprouvé, moins l'algorithme est amélioré.
- **Diffusion limitée de l'algorithme :** Ceci permet de limiter le nombre d'utilisateurs en vue de réduire la probabilité de découverte des secrets de l'algorithme et de limiter les attaques.

**Algorithme publié**

- **Tout le monde à accès à l'algorithme :** Aucun(e) mécanisme/traitement/opération ne peut constituer un secret puisque tout le monde y a accès. De ce fait, la cryptanalyse repose uniquement sur la découverte de la clé.
- **Sécurité améliorée :** Comme tout le monde à accès à l'algorithme, les failles (laissées intentionnellement ou non par les concepteurs) peuvent être plus facilement découvertes et corrigées.
- **Pas besoin de protéger la sémantique de l'algorithme contre le reverse-engineering :** La protection de la sémantique de l'algorithme n'est pas nécessaire puisque tous ses mécanismes, et traitements (opérations) sont connus.
- **Large diffusion de l'algorithme :** Large nombre d'implémentations de l'algorithme dans des logiciels peuvent donc être réalisées.niveau mondial.
- **Standardisation de l'algorithme:** C'est possible si tout le monde utilise la même version publique.

## 2.5 Les principaux concepts cryptographiques

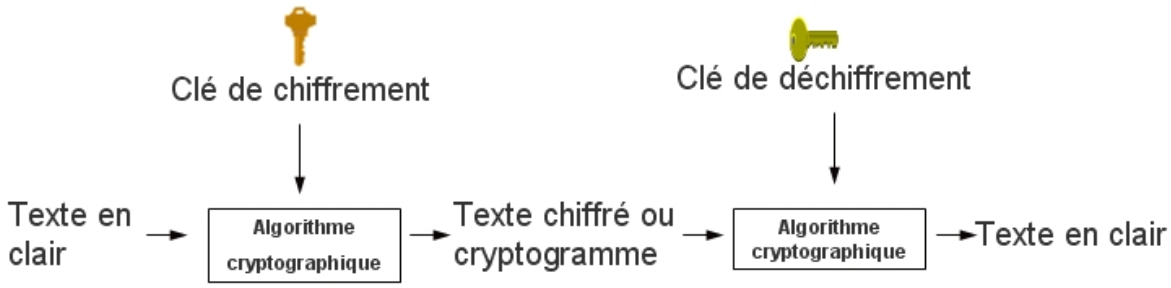
### 2.5.1 Cryptosystème à clé symétrique




Caractéristiques :

- Une seule clé est nécessaire : les clés chiffrement (notée KE) et de déchiffrement (notée KD) sont identiques: KE = KD = K
- La clé K doit rester secrète,
- Les algorithmes les plus répandus sont le DES, AES, 3DES, ...
- Les clés sont générées aléatoirement dans l'espace des clés,
- Les opérations des algorithmes sont des transpositions et substitutions des bits du texte clair en fonction de la clé,
- Le DES en utilise des clés de 56 bits, mais l'AES peut aller jusque 256,
- L'avantage principal de ce mode de type de cryptosystème est sa rapidité dans les opérations de chiffrement et de déchiffrement,
- Le principal désavantage réside dans la distribution des clés : pour une meilleure sécurité, on préfèrera l'échange manuel. Malheureusement, pour de grands systèmes, le nombre de clés peut devenir conséquent. C'est pourquoi on utilisera souvent des échanges sécurisés pour transmettre les clés. En effet, pour un système à N utilisateurs, il y aura autant de clés que de pairs d'utilisateurs, c'est à dire $N.(N-1)/2$ clés.

### 2.5.2 Cryptosystème à clé publique




Caractéristiques :

- Deux clés sont nécessaires : une clé publique PK (symbolisée par la clé verticale) et une clé privée secrète SK (symbolisée par la clé horizontale),
- Propriété mathématique 1: La connaissance de la clé PK ne permet pas de déduire la clé SK ,
- Propriété mathématique 2: Le déchiffrement via la clé privée SK du chiffrement d'un message M par la clé public PK est égal à M : $D_{SK}(E_{PK}(M)) = M$ ,
- L'algorithme de cryptographie asymétrique le plus connu est le RSA,

- Le principe de ce genre d'algorithme est qu'il s'agit d'une fonction unidirectionnelle à trappe. Une telle fonction à la particularité d'être facile à calculer dans un sens, mais difficile voire impossible dans le sens inverse. La seule manière de pouvoir réaliser le calcul inverse est de connaître une trappe. Une trappe pourrait par exemple être une faille dans le générateur de clés. Cette faille peut être soit intentionnelle de la part du concepteur (définition stricte d'une trappe) ou accidentelle.
- Les algorithmes se basent sur des concepts mathématiques tels que l'exponentiation de grands nombres premiers (RSA), le problème des logarithmes discrets (ElGamal), ou encore le problème du sac à dos (Merkle-Hellman).
- La taille des clés s'étend de 512 bits à 2048 bits en standard. Dans le cas du RSA, une clé de 512 bits n'est plus sûre au sens "militaire" du terme, mais est toujours utilisable de particulier à particulier.
- Au niveau des performances, le chiffrement par voie asymétrique est environ 1000 fois plus lent que le chiffrement symétrique.
- Cependant, à l'inverse du chiffrement symétrique où le nombre de clés est le problème majeur, ici, pour un système à N utilisateurs, seules n paires sont nécessaires. En effet, chaque utilisateur possède une paire de clés (SK , PK) et tous les transferts de message ont lieu avec ces clés.
- La distribution des clés est grandement facilitée car l'échange de clés secrètes n'est plus nécessaire. Chaque utilisateur 2 conserve sa clé secrète sans jamais la divulguer. Seule la clé publique devra être distribuée.

Ces cryptosystèmes sont basés sur des problèmes dits NP-complets. Ce sont des problèmes pour lesquels il n'existe pas d'algorithme de performance raisonnable pour les résoudre. Leur résolution implique une consommation prohibitive de ressource (temps et mémoire). La cryptanaylse, le déchiffrement sans la connaissance de la clé privée, implique la résolution d'un problème NP-Complet. RSA repose le problème d'exponentiation de grands nombres premiers. Le cryptosystème d'ElGamal repose sur le problème des logarithmes discrets, celui de Merkle-Hellman repose sur le problème du sac à dos. Tous ces problèmes sont NP-Complets.

### 2.5.3 Fonction de hachage

Il s'agit de la troisième grande famille d'algorithmes utilisés en cryptographie. Le principe est de transformer un message clair de longueur quelconque en un message de longueur fixe et inférieure à celle de départ. Le message réduit portera le nom de "**Haché**" ou de "**Condensé**". L'intérêt est **d'utiliser ce condensé comme empreinte digitale du message original afin que ce dernier soit identifié de manière univoque.** Deux caractéristiques (théoriques) importantes sont les suivantes :

- **Ce sont des fonctions unidirectionnelles :** A partir du haché du message M, noté H(M), il est impossible de retrouver M.
- **Ce sont des fonctions sans collisions :** Deux messages distincts ont des hachés distincts. A partir de H(M ) et M il est impossible de trouver M' ≠ M tel que H(M') = H(M ).

Il est bien entendu que le terme "impossible" n'est pas toujours à prendre au pied de la lettre ! Il s'agit ici de concepts théoriques. La réalité est quelque peu différente. Ainsi, pour le caractère "sans collision", dans les faits, cela est "très difficile" dans le meilleur des cas, mais jamais impossible, comme le bon sens le laisse penser

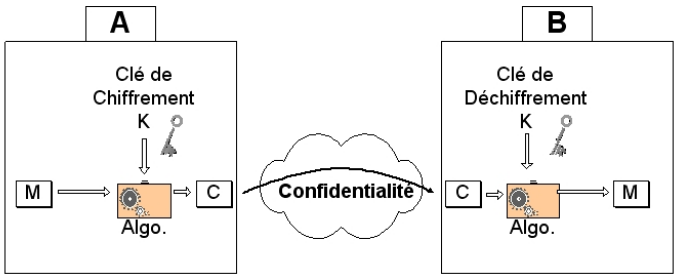
### 2.5.4 Protocoles cryptographiques

Un protocole est un ensemble de règles qui régissent la communication entre plusieurs entités. Un protocole est dit cryptographique s'il utilisent des opérations cryptographiques pour sécuriser les communications.

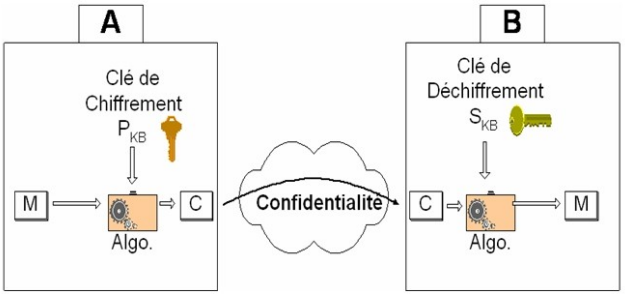
La sécurisation des échanges implique au moins trois services : la confidentialité, l'intégrité et l'authentification. Signalons **la distinction entre "services"** (confidentialité, intégrité, authentification, etc.) **et "mécanismes"** (les moyens utilisés : chiffrement, déchiffrement, signature, hachage, etc.).

### 2.5.4.1 Confidentialité

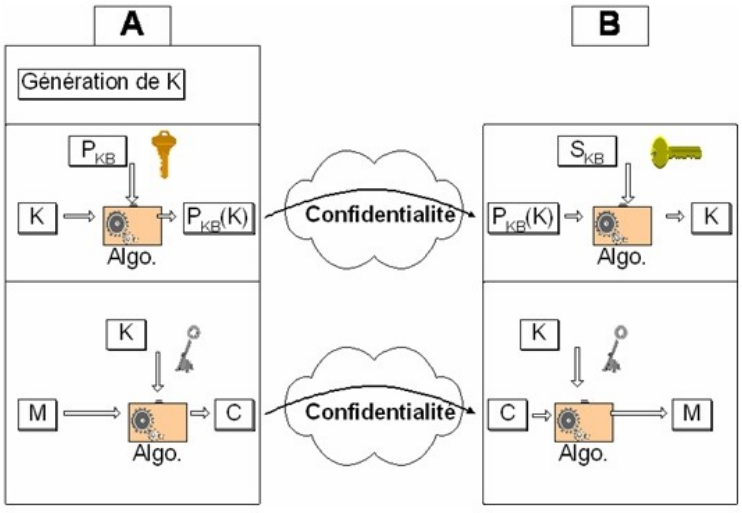
**Système symétrique :** Les mécanismes de chiffrement et déchiffrement permettent d'implémenter le service de confidentialité. Elle est amenée par le chiffrement du message. Dans le cas de systèmes à clés symétriques, la même clé est utilisée pour les fonctions de chiffrement et de déchiffrement : $E_K(M)$ et $D_K(C)$. Ce type de chiffrement nécessite un échange sûr préalable de la clé K entre les entités A et B.




**Système asymétrique :** Contrairement aux systèmes symétriques, l'échange préalable de clé n'est pas nécessaire. Chaque entité possède sa propre paire de clés. On aura donc la paire ($P_{KA}$ , $S_{KA}$) pour l'entité A et la paire ($P_{KB}$ , $S_{KB}$) pour l'entité B.

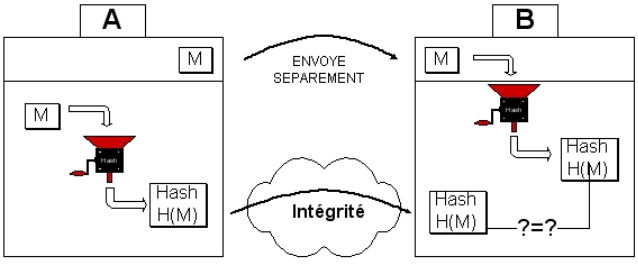



**Système Hybride :** Il repose comme son nom l'indique sur les deux systèmes précédents. Par l'intermédiaire du système asymétrique, on sécurise l'échange de la clé symétrique K. Ensuite, les deux parties ayant acquis de manière sécurisée K basculent en mode symétrique (utilisent K pour chiffrer et déchiffrer les messages).




### 2.5.4.2 Intégrité

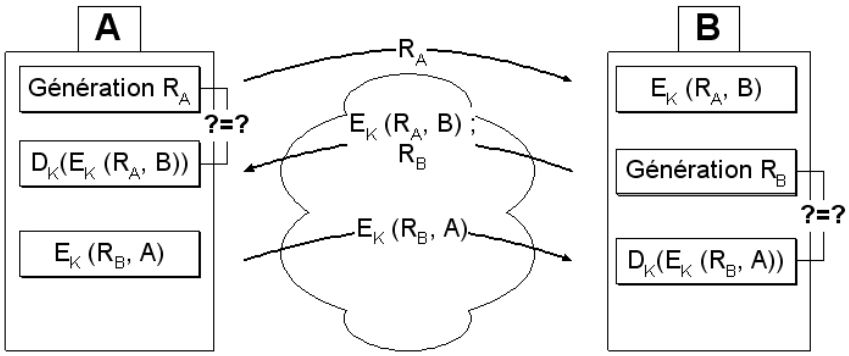
Les mécanismes de hachage permettent d'implémenter le service d'intégrité. Il faut vérifier si le message n'a pas subi de modification durant la communication. Ce schéma prête uniquement l'attention à la vérification de l'intégrité. Il ne tient pas compte de la confidentialité.



### 2.5.4.3 Authentification

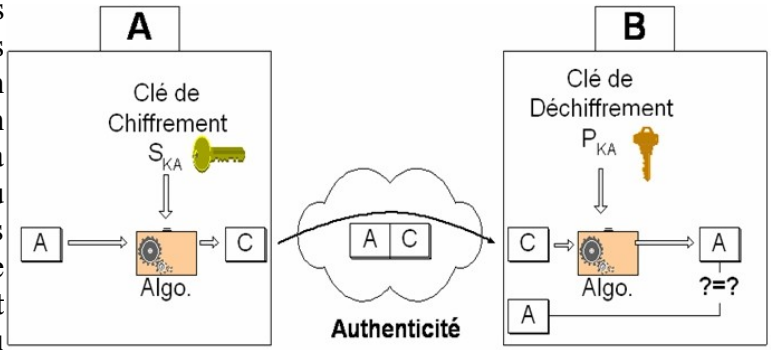
L'authentification a lieu à plusieurs niveaux : au niveau des parties communicantes et au niveau du message envoyé. Les parties communicants doivent être authentiques (pas d'usurpation d'identité) et les messages doivent être authentiques (non falsifiés)

**Au niveau des parties communicantes :** Chaque entité impliquée dans la communication voudrait se rassurer qu'il communique bien avec son homologue et qu'il n'y a pas usurpation d'identité.



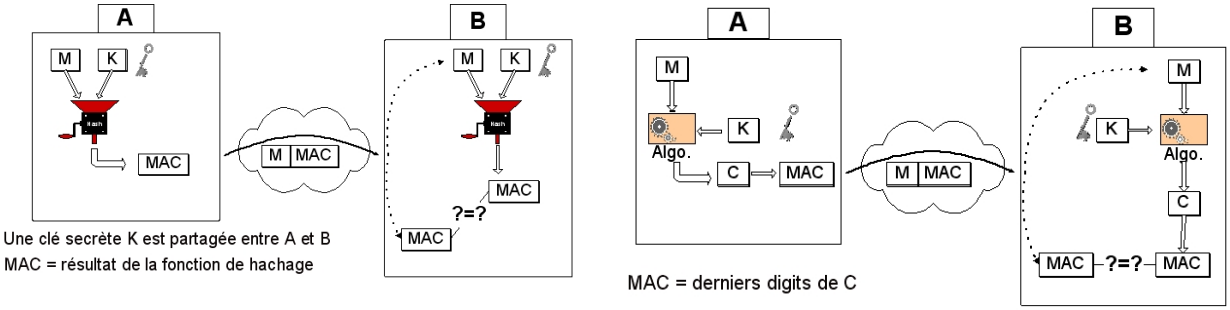

**Cas d'un système symétrique :** Les lettres A et B représentent des identificateurs personnels. $R_A$ et $R_B$ sont des nonces (nombres aléatoires), liées respectivement aux utilisateur A et B. **Trois opérations :** (1) Chacun envoie son nonce en clair à l'autre. (2) Chacun chiffre le nonce reçu et l'envoi. (3) Chacun vérifie si le nonce chiffré et reçu est égal au sien. **S'il y a égalité, A et B sont convaincus que les nonces ont été chiffrés avec leur clé sécrète,** puisqu'on ne peut pas obtenir les mêmes chiffrements avec une clé distincte. A est convaincu qu'il communique avec B et B est convaincu qu'il communique avec A.

**Cas d'un système asymétrique :**Les lettres A et B représentent des identificateurs personnels. A chiffre son identifiant avec sa clé privée et envoi son identifiant en clair et la version chiffrée à B. B déchiffre la partie chiffrée du message avec la clé publique de A, puis vérifie s'il est égal à la partie non-chiffrée du message reçu. **S'il y a égalité, B est convaincu que la deuxième partie du message reçu a été chiffré avec la clé privé (sécrète) de A,** puisqu'on ne peut pas obtenir le même chiffrements avec une clé distincte**.** Cependant, la confidentialité est nulle puisque le message envoyé pourra être lu par toute personne possédant la clé publique, c'est-à-dire, n'importe qui.




**Au niveau du message :**

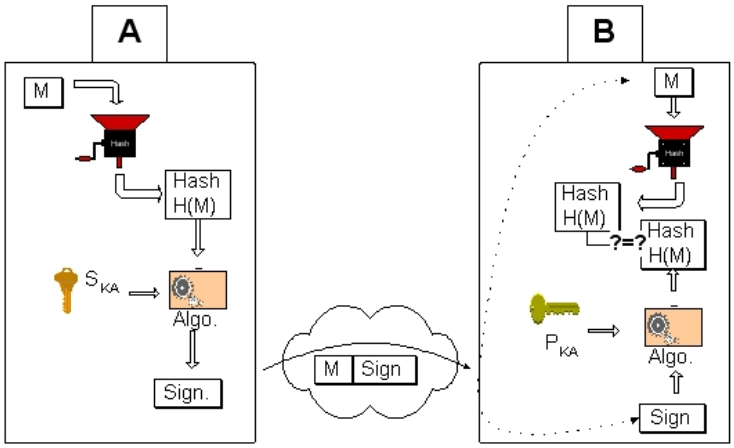
**Par l'utilisation d'un MAC (Message Authentication Code)** généré à l'aide d'un cryptosystème à clé symétrique où le MAC est constitué des derniers digits de C, ou généré à l'aide d'une fonction de hachage, la clé secrète K utilisée étant partagée par les deux entités A et B.



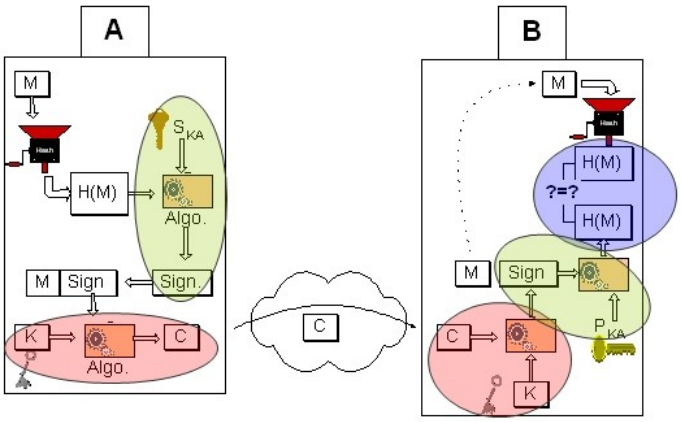
Dans le cas symétrique, si le MAC du message en clair reçu (M) est égal au MAC reçu, B est convaincu que le MAC reçu a été fabriqué avec la clé symétrique qu'il partage uniquement avec A.
Dans le cas asymétrique, **en cas d'égalité entre le MAC reçu et la MAC fabriqué par B, ce dernier est convaincu que le MAC reçu a été fabriqué avec la clé avec la clé privée de A.**

**Par l'utilisation d'une signature digitale.** Parmi les propriétés remarquables de ces signatures, on peut dire qu'elles doivent être authentiques, infalsifiables, non-réutilisables, non-répudiables, et inaltérables. Ici, on fait abstraction de la confidentialité. C'est l'authentification qui importe.

A fabrique un haché du message M, signe ledit haché avec sa clé privée, puis envoie M et la signature (sign) à B. Ce dernier refabrique le haché du message reçu et le compare au déchiffrement de la signature. **En cas d'égalité, il est convaincu que le message a été signé avec la clé privée de A.**




#### 2.5.4.4 Synthèse




## 2.5.5. La Signature Électronique

**Qu'est-ce que la signature électronique ?**

La signature électronique est un mécanisme permettant de garantir l'intégrité d'un document électronique et d'en authentifier l'auteur, par analogie avec la signature manuscrite d'un document papier. Elle a **la même valeur légale qu'une signature manuscrite**. Elle se différencie de la signature écrite par le fait qu'elle n'est pas visuelle, mais correspond à une suite de caractères.
A l'inverse, la signature numérique l (dessiné par le signataire ou l'insertion d'une image) et la signature scannée n'ont pas la même force probante, elle ne permet pas de rapporter la preuve du consentement.
**L'objectif** majeur de la signature électronique est triple :

- **garantir l'intégrité d'un document**, c'est-à-dire s'assurer que le document n'a pas été altéré entre sa signature et sa consultation ;
- **authentifier son auteur**, c'est-à-dire s'assurer de l'identité de la personne signataire ;
- **rapporter la preuve du consentement**.

Pour cela, elle doit avoir les **caractéristiques** suivantes :

- **authentique** : l'identité du signataire doit pouvoir être retrouvée de manière certaine

- **infalsifiable** : une personne ne peut pas se faire passer pour un autre
- **non réutilisable** : la signature fait partie du document signé et ne peut être déplacée sur un autre document
- **inaltérable** : une fois que le document est signé, on ne peut plus le modifier
- **irrévocable** : la personne qui a signé ne peut le contester

La signature électronique permet de signer en quelques secondes et sans contact physique des documents essentiels au bon fonctionnement des entreprises, tels que :

- les contrats de travail
- les factures
- les bons de commande
- les mandats et les compromis de vente
- les devis
- les documents comptables
- les documents juridiques
- les actes notariés

**Quel type de signature électronique utiliser ?**
Le choix du niveau de signature, tel que défini par la norme eIDAS, dépend de l'usage, et de l'enjeu du document à signer : en cas de litige, plus votre signature aura un niveau de fiabilité fort, plus il sera difficile de contester la validité de l'acte signé et les engagements qu'il contient. Selon les cas on choisira le niveau de sécurité adapté.

## La signature électronique standard (niveau 1)

La signature électronique manuscrite est utilisée par exemple lorsque vous tapez le code secret d'une carte de crédit, quand vous faites une signature manuscrite sur un appareil électronique, ou encore quand vous scannez une signature manuscrite, que vous apposez sur un document pour l'envoyer par mail. Elle est parfois appelé une signature numérique.
**Sa valeur juridique est limitée**, car elle ne garantit pas l'intégrité des données signées ni l'identité du signataire, etc. Elle peut toutefois valoir commencement de preuve par écrit. Sa vocation est de **simplifier des processus internes** où la signature est indispensable (autorisations, accusés de réception, commandes, contrats, etc.).

## La signature électronique avancée (niveau 2)

C'est **la plus couramment utilisée par les entreprises**. **Grâce à l'utilisation d'une clé privée accessible seulement à la personne qui signe et seulement à elle** (son smartphone par exemple), elle permet :

- d'identifier la ou le signataire
- de lier la signature à son auteur
- de garantir l'intégrité de l'acte signé.

Concrètement, le signataire télécharge sa pièce d'identité sur la plateforme du prestataire de signature électronique qui peut ainsi procéder à des contrôles et l'authentifier.
Dans la pratique, c'est la signature électronique avancée, qui est la plus couramment utilisée. Ce type de signature est par exemple beaucoup utilisé pour signer une facture dématérialisée, un contrat de travail, un compromis de vente immobilier ou un contrat d'assurance vie. Elle nécessite toutefois l'acquisition d'un certificat de signature électronique répondant aux exigences de la norme eIDAS.

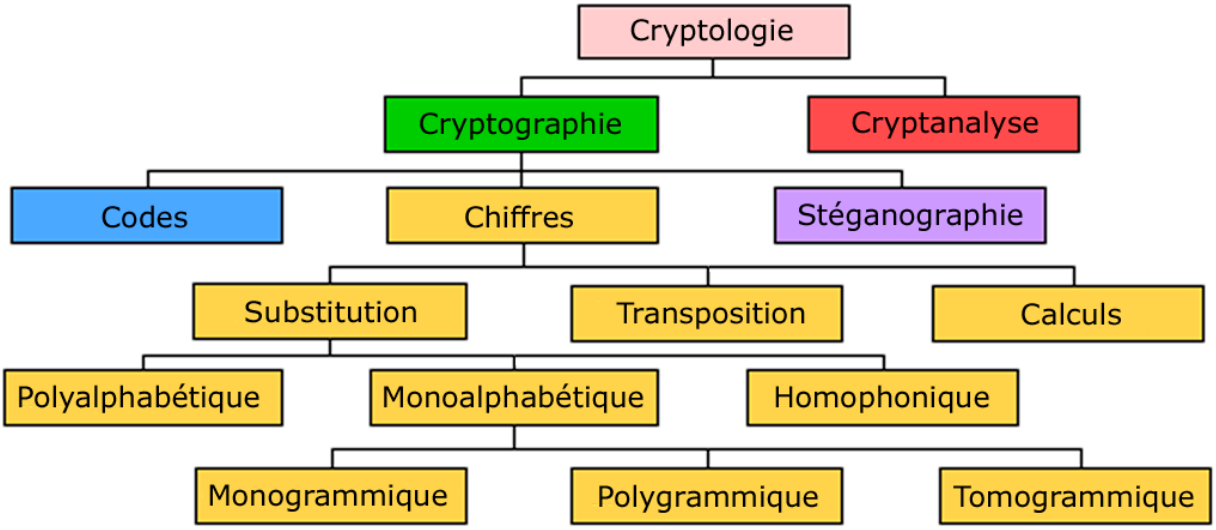
**La signature électronique qualifiée (niveau 3)**
Elle est la signature **la plus robuste sur les plans technique et juridique**. Ce type de signature exige que :

- l'identité du signataire soit validée en amont (en physique ou à distance selon certaines conditions), et ce par une autorité de certification ou un prestataire de service de certification électronique ;
- une clé de signature, un dispositif qualifié de création de signature électronique. Ce token physique (clé USB, carte à puce…), est délivré à uniquement à une personne physique. Une entreprise ne peut signer qu'au travers d'un représentant, une personne physique, dûment habilitée.

Selon le code civil, **seule cette signature est l'équivalent de la signature manuscrite**. Plus lourde à mettre en œuvre et plus onéreuse, la signature qualifiée est généralement **réservée aux documents pour lesquels l'authentification est fondamentale**, par exemple, dans le cas de production d'actes notariés (notaires, huissiers…) ou dans le contexte des marchés publics (de l'appel d'offre à la facture). Elle nécessite l'acquisition d'un certificat de signature électronique et un dispositif de création de signature électronique.

**Fonctionnement de la signature électronique avancée (niveau 2)**
**Pour signer numériquement un contrat de location ou même un achat immobilier et que cela ait une valeur légale, il faut passer par un tiers de confiance.** Ces entreprises habilitées à effectuer des opérations de sécurité juridique d'authentification, de transmission et de stockage sont nombreuses. Si elles proposent chacune leur solution, plus ou moins élaborées ou faciles d'utilisation, leur fonctionnement est relativement similaire : la procédure ressemble un peu à un achat en ligne, avec une authentification par code secret via SMS. Le processus est le suivant.

- Vous vous connectez sur le site du tiers de confiance en ligne à l'aide de vos identifiants, voire de votre clef électronique dans le cas d'une authentification qualifiée (niveau 3)
- Vous ajoutez les documents (word, PDF, etc…) que vous souhaitez faire signer.
- Vous invitez des signataires après avoir renseigné leurs coordonnées (en particulier leurs numéros de téléphone portable).
- Chaque signataire reçoit par mail une notification pour signer ainsi qu'un code par SMS permettant de sécuriser la signature.




## 3.1. Substitution monoalphabétique

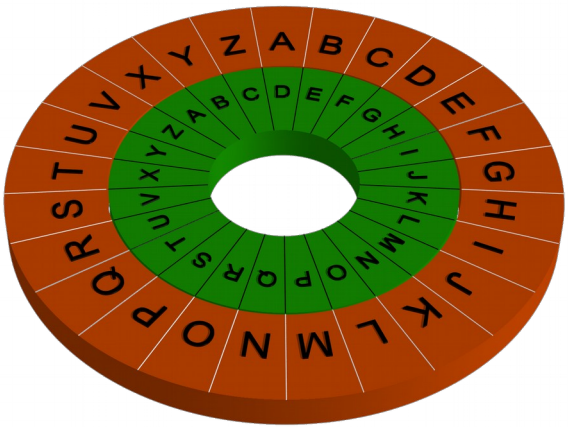
### 3.1.1. Chiffre de César (50 av. J-C)

Jules César a-t-il vraiment prononcé la célèbre phrase : « DOHD MDFWD HVW » ou bien comme le disent deux célèbres Gaulois : « Ils sont fous ces romains ! ». En fait César, pour ses communications importantes à son armée, cryptait ses messages. Ce que l'on appelle le chiffrement de César est un **décalage circulaire** des lettres. Dans les formules ci-dessous, p est l'indice de la lettre de l'alphabet, k est le décalage.

**La formule de chiffrement est p : C = E(p) = (p + k) mod 26**
**La formule de déchiffrement est de C: p = D(C) = (C − k) mod 26**
On rappelle que **a mod b** est égal au reste de la division de a par b
Les indices des lettres se suivent. L'indice de A est 0, celui de B est 1, celui de C est 3…

| 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chiffré | 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 |

Par exemple, si k=3, la valeur du décalage est 3. Pour crypter un message, A devient D, B devient E, C devient F ,…: A → D, B −→ E , C→ F . . . W → Z , X → A, Y → B, Z → C. La lettre a chiffrer est en rouge et son chiffrement est en vert. Le chiffrement de « BONJOUR » est « ERQMRXU ».
Pour déchiffrer le message de César, il suffit de décaler les lettres dans l'autre sens, D se déchiffre en A, E en B,...

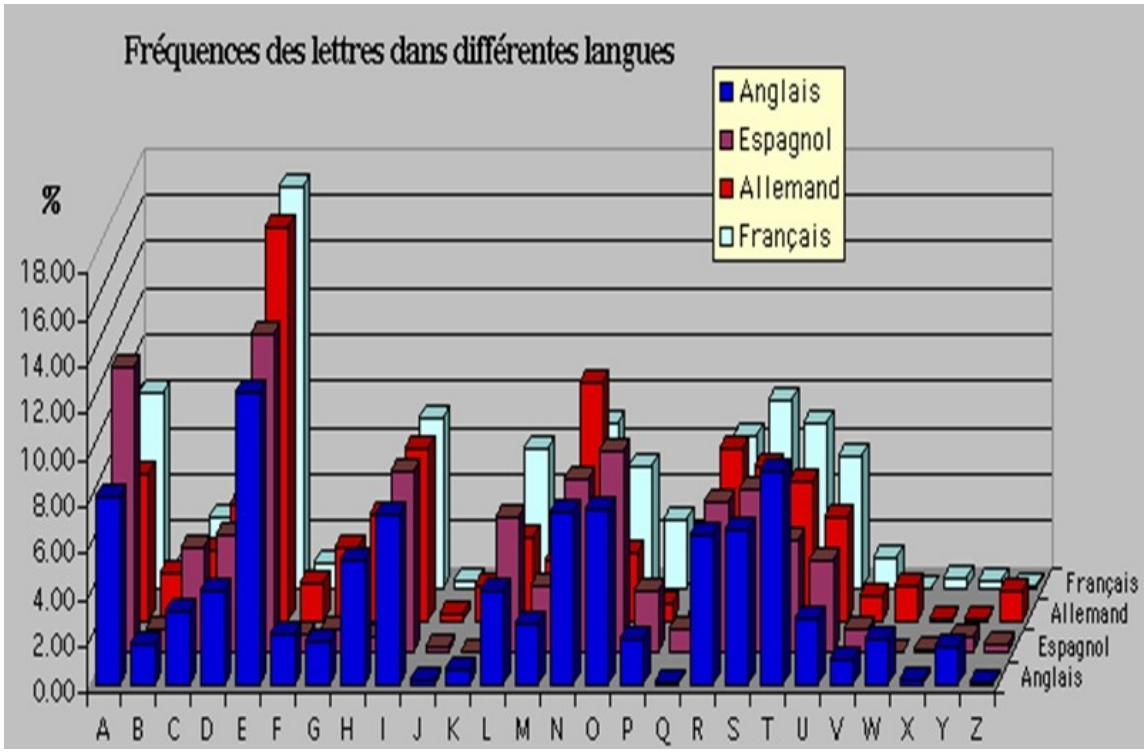


Supposons maintenant que E(x) = x + 11. Le chiffrement de « COUCOU » est : COUCOU → 2 14 20 2 14 20 → **13 25 5 13 25 5** → NZFNZF. Le Déchiffrement de « NZFNZF. » est « COUCOU » .

Si on connait l'algorithme utilisé (ici César), **la cryptanalyse par force brute est très facile**. En effet, dans le cas du chiffre de César, seules 25 clés sont possibles. L'intrus va appliquer successivement les différentes clés possibles au message jusqu'à obtenir un message compréhensible.

### 3.1.2. Analyse de fréquences

Lorsque la langue de départ et la technique de chiffrement sont connus, on peut exploiter les régularités du langage par le principe d'analyse de la fréquence d'une lettre. Cette technique ne fonctionne bien que si le message chiffré est suffisamment long pour avoir des moyennes significatives.




Pour le Français, l'Allemand, l'Anglais et l'Espagnol, la lettre la plus fréquente est « e ». L'intrus exploite ces statistiques en considérant que la lettre de plus grande fréquence dans le texte chiffré corresponds à « e ». dans le texte en clair.

Cependant, il existe également des cas où cette analyse ne fonctionne pas, comme pour « De Zanzibar à la Zambie et au Zaïre, des zones d'ozone font courir les zèbres en zigzags zinzins »..

Pour éviter ce type d'attaque sur un texte chiffré, il existe différents moyens. On peut par exemple chiffrer le message par digrammes, trigrammes, etc. Cependant l'intrus peut exploiter les statistiques sur les bigrammes et trigrammes.

Les 20 bigrammes les plus fréquents

| Bigrammes | ES | DE | LE | EN | RE | NT | ON | ER | TE | EL | AN | SE | ET | LA | AI | IT | ME | OU | EM | IE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nombres | 3318 | 2409 | 2366 | 2121 | 1885 | 1694 | 1646 | 1514 | 1484 | 1382 | 1378 | 1377 | 1307 | 1270 | 1255 | 1243 | 1099 | 1086 | 1056 | 1030 |

Les 20 trigrammes les plus fréquents

| Trigrammes | ENT | LES | EDE | DES | QUE | AIT | LLE | SDE | ION | EME | ELA | RES | MEN | ESE | DEL | ANT | TIO | PAR | ESD | TDE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nombres | 900 | 801 | 630 | 609 | 607 | 542 | 509 | 508 | 477 | 472 | 437 | 432 | 425 | 416 | 404 | 397 | 383 | 360 | 351 | 350 |

Pour échapper à l'analyse de fréquences, une solution consiste à remplacer une lettre non pas par un symbole unique, mais par un symbole choisi au hasard parmi plusieurs. Dans sa version la plus sophistiquée, on choisira un nombre des symboles proportionnel à la fréquence d'apparition de la lettre; on parle alors de renversement des fréquences. Ce type de substitution est appelé **substitution homophonique** (on dit aussi **substitution à représentations multiples**). On peut situer l'âge d'or de la substitution homophonique entre 1500 et 1750.

## 3.1.3. Chiffre affine

Le chiffre affine est un chiffre de substitution simple. Il est placé ici dans le cours car on peut le voir comme la version unidimensionnelle du chiffre de Hill. L'idée est d'utiliser **une fonction de chiffrage affine du type c(x) = (ax + b) mod 26**, **où gcd(a , 26) = 1,** a et b sont des constantes, et où x est la lettre à chiffrer. La clé sécrete est constituée du couple (a, b).

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). Si a = 1 et b = 0, aucun décalage n'a lieu, l'alphabet de départ se retrouve chiffré par lui même, et donc ne subit aucune modification.

La fonction de déchiffrement de **c(x) = (ax + b) mod 26** est **$c^{-1}(y) = a^{-1} * (y - b) \bmod 26$,** où $a^{-1}$ désigne l'inverse de a modulo 26, c'est à dire que **$a*a^{-1} \bmod 26 = 1$**..

**Exemple :**
Soient la clé (a , b) = (3, 11).
Transformation de chiffrement est : c(x) = (3x + 11) mod 26
$a^{-1} = 3^{-1} \bmod 26 = 9$ car $3 * 9 \bmod 26 = 27 \bmod 26$ 1=1.
Transformation de déchiffrement est :**$c^{-1}(y) = 9\ (y - 11) \bmod 26$,**

Ainsi, le chiffrement de la suite de lettres NSA est YN ; 'NSA' → 13 18 0 → 24 13 11 →'YNL

### 3.1.3.1 Cryptanalyse du chiffre affine

Il s'agit de retrouver les paramètres (a et b) de la fonction de chiffrement à partir d'un ou de plusieurs chiffrements. La cryptanalyse peut se faire en quatre étapes :

**Étape 1 :** Etablir la fréquence relative de chaque lettre du texte chiffré, par analyse de fréquence. Supposons que le texte chiffré est ; « HGAHY RAEFT GAGRH DGAGM OEHIY RAAOT ZGAGJ GKFDG AZGSB INNTG KGRHE NNIRG ». On dénombre 12 fois la lettre G et 8 fois la lettre A. Supposons que le langage original du texte est le français.

**Etape 2 :** Dériver les équations correspondantes sur base de l'analyse de fréquences
Sur la base de l'analyse de fréquences de la section 3.1.2, E, A, S, I, N sont les lettres les plus fréquentes en français, On peut donc upposer que E a été codé par G et S par A : E → G et S → A. Ce qui signifie que

$$c(E) = G$$
$$c(S) = A$$

où c est la fonction de chiffrement. En remplaçant chaque lettre par son indice, il découle que :

$$c(4) = 6$$
$$c(18) = 0$$

**Etape 3 :** Résoudre les équations pour que retrouver la clé (a,b) .

c(4) = 6 et c(18) = 0 → $(4a + b) \bmod 26 = 6$ et $(18\ a + b) \bmod 26 = 0$
→ $(4a + b) \equiv 6 \pmod{26}$ et $(18\ a + b) \equiv 0 \pmod{26}$
→ $14a \equiv -6 \bmod 26$
→ a = 7
→ b = 4.

**Étape 4 :** Fabriquer la fonction de déchiffrement et déchiffrer le message
Rappelons l'expression générale de la fonction de déchiffrement : $\mathbf{c^{-1}(y) = a^{-1} * (y - b) \bmod 26}$, **où** $\mathbf{a^{-1}}$ **désigne l'inverse de a modulo 26, c'est à dire que** $\mathbf{a*a^{-1} \bmod 26 = 1}$. En remplaçant $a^{-1}$ et b par leur expression, il vient que :

$$\mathbf{c^{-1}(y)} = 15 * (y - 4) \bmod 26.$$

En appliquant cette fonction au message chiffré, on obtient le message initial (message en clair) :

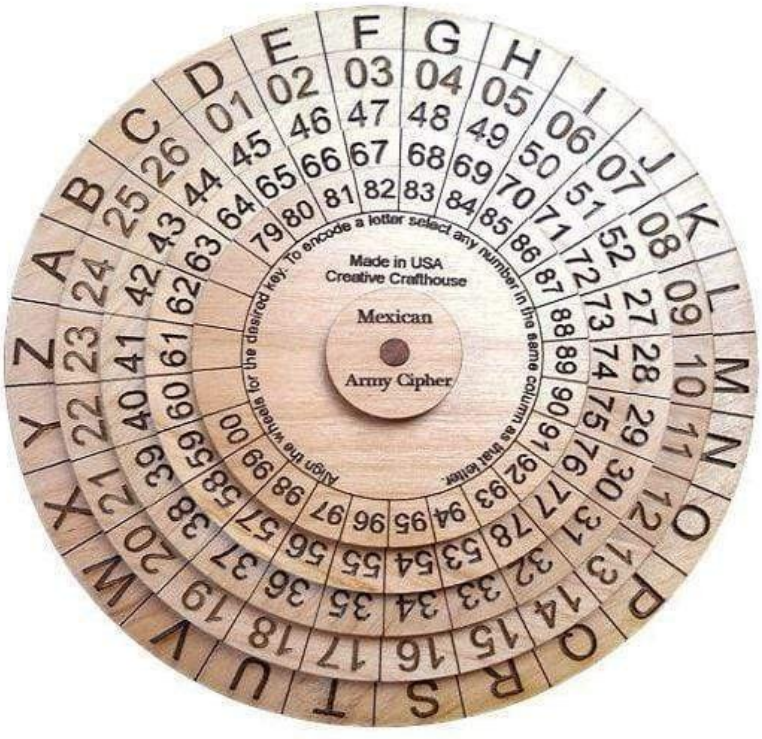
« TESTONSAPRESENTLESEQUATIONSSURDESEXEMPLESDECHIFFREMENTAFFINE »

## 3.2 Substitution homophonique : Le Disque de l'Armée Mexicaine

Le disque de l'armée mexicaine est un chiffre homophonique basé sur quatre alphabets chiffrants composés de nombres de 01 à (1)00. Ce chiffre a été en usage avant la première guerre mondiale, lors des querelles de frontières entre le Mexique et les États-Unis. Il a été cassé par le capitaine d'infanterie Parker Hitt, un des meilleurs cryptanalystes américains de l'époque. Il fut très peu utilisé pendant la première guerre mondiale, à cause de sa faible sécurité. Originellement, ce chiffre était disposé sur un tableau comme ci-dessous:

| 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1er alphabet chiffrant | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 |
| 2ème alphabet chiffrant | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 27 | 28 |
| 3ème alphabet chiffrant | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 |
| 4ème alphabet chiffrant | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 00 | | | | |

Plus tard, afin de le rendre plus portable et plus facile à utiliser, les alphabets furent écrits sur cinq disques rotatifs de différentes tailles, et empilés comme esquissé sur le dessin ci-dessous. L'alphabet clair était écrit sur le disque le plus grand.




L'utilisation de ce disque est très simple: on convient tout d'abord d'un nombre qui donnera l'orientation des disques intérieurs. Ce nombre est formé de 8 chiffres et a le format *aabbccdd*, où *aa* est un nombre compris entre 01 et 26, *bb* est compris entre 27 et 52, *cc* entre 53 et 78 et *dd* entre 79 et (1)00. Chacun de ces quatre nombres, appartenant tous à des disques différents, devront se

trouver sous la lettre claire A. Dans notre exemple, le nombre-clef est 12295379: sous la lettre A sont positionnés le 12, le 29, le 53 et le 79. Pour chiffrer, on remplace simplement la lettre claire par un des trois ou quatre nombres qui se trouvent sous elle.

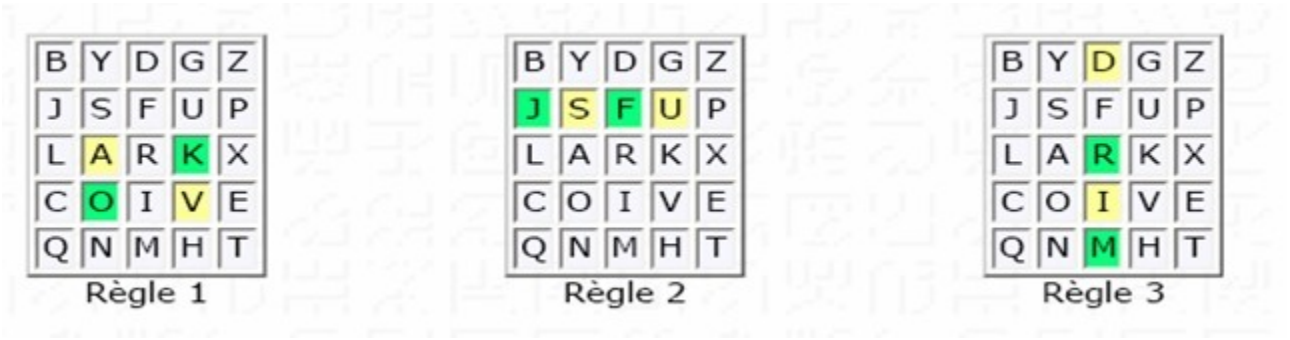
## 3.3. Chiffrement polygraphique

Il s'agit ici de chiffrer un groupe de n lettres par un autre groupe de n symboles. On citera notamment le chiffre de Playfair et le chiffre de Hill. Ce type de chiffrement porte également le nom de substitutions polygrammiques.

### 3.3.1. Chiffre de Playfair (1854)

https://www.apprendre-en-ligne.net/crypto/subst/playfair.html

Le chiffre Playfair, chiffre polygrammique, a été popularisé par Lyon Playfair, mais il a été inventé par Sir Charles Wheatstone (1854), un des pionniers du télégraphe électrique. On dispose les 25 lettres de l'alphabet (W exclu car inutile, on utilise V à la place) dans une grille 5x5, ce qui donne la clef. La variante anglaise consiste à garder le W et à fusionner I et J. **Ici, la clé secrète est représentée sous la forme d'une grille.**




**Méthode de chiffrement**

On chiffre le texte par groupes de deux lettres (des bigrammes) en appliquant les règles suivantes:

1. Si les deux lettres sont sur les coins d'un rectangle, alors les lettres chiffrées sont sur les deux autres coins. Exemple **OK** devient **VA**, **BI** devient **DC**, **GO** devient **YV**. La première des deux lettres chiffrées est sur la même ligne que la première lettre claire.
2. Si deux lettres sont sur la même ligne, on prend les deux lettres qui les suivent immédiatement à leur droite: **FJ** sera remplacé par **US**, **VE** par **EC**. **La notion de successeur est cyclique comme dans une anneau (ou un tore).**
3. Si deux lettres sont sur la même colonne, on prend les deux lettres qui les suivent immédiatement en dessous: **BJ** sera remplacé par **JL**, **RM** par **ID**.
4. Si le bigramme est composé de deux fois la même lettre, on insère une lettre représentant nul (usuellement le X) entre les deux pour éliminer ce doublon.

Pour déchiffrer, on applique les règles ci-dessus à l'envers. Pour former les grilles de chiffrement, on utilise un **mot-clef secret** pour créer un alphabet désordonné avec lequel on remplissait la grille ligne par ligne.

### 3.3.2. Chiffre de Hill (1929)

https://www.apprendre-en-ligne.net/crypto/hill/hill3.html

Le chiffre publié en 1929 par **Lester S. Hill** (1891-1961) est un **chiffre polygraphique** (comme le chiffre de Playfair), c'est-à-dire qu'on ne (dé)chiffre pas les lettres les unes après les autres, mais par paquets. Nous étudierons ici la **version bigraphique** du chiffre de Hill, puisque nous grouperons les lettres deux par deux, mais on peut imaginer des paquets plus grands, par exemple des paquets de trois lettres.

**Chiffrement**

Les lettres sont d'abord remplacées par leur rang dans l'alphabet. Les lettres $P_k$ et $P_{k+1}$ du texte clair seront chiffrées $C_k$ et $C_{k+1}$ avec la formule ci-dessous:

$$\begin{pmatrix} C_k \\ C_{k+1} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} P_k \\ P_{k+1} \end{pmatrix} \pmod{26}$$

Ce qui signifie, pour fixer les idées, que les deux premières lettres du message clair ($P_1$ et $P_2$) seront chiffrées ($C_1$ et $C_2$) selon les deux équations suivantes:

$$C_1 \equiv aP_1 + bP_2 \pmod{26}$$
$$C_2 \equiv cP_1 + dP_2 \pmod{26}$$

**Exemple de chiffrement**

**Alice** prend comme clef de cryptage la matrice $\begin{pmatrix} 9 & 4 \\ 5 & 7 \end{pmatrix}$ pour chiffrer le message "je vous aime" .

Après avoir remplacé les lettres par leur rang dans l'alphabet (a=1, b=2, etc.), elle obtiendra:

$$C_1 \equiv 9 \cdot 10 + 4 \cdot 5 \pmod{26} = 110 \pmod{26} = 6$$
$$C_2 \equiv 5 \cdot 10 + 7 \cdot 5 \pmod{26} = 85 \pmod{26} = 7$$

Elle fera de même avec les 3e et 4e lettres, 5e et 6e, etc. Elle obtiendra finalement:

| Lettres | j | e | v | o | u | s | a | i | m | e |
|---|---|---|---|---|---|---|---|---|---|---|
| Rangs ($P_k$) | 10 | 5 | 22 | 15 | 21 | 19 | 1 | 9 | 13 | 5 |
| Rangs chiffrés ($C_k$) | 6 | 7 | 24 | 7 | 5 | 4 | 19 | 16 | 7 | 22 |
| Lettres chiffrées | F | G | X | G | E | D | S | P | G | V |

**Remarque**

Certains auteurs posent "A"=1, "B"=2, ..., "Z"=0. On a utilisé ici cette convention. Cependant, d'autres auteurs posent "A"=0, "B"=1, ..., "Z"=25. Les deux conventions se défendent. L'essentiel est que les protagonistes se mettent d'accord avant d'échanger des messages. On pourrait même imaginer de prendre un alphabet désordonné, par exemple "A"=15, "B"=6, etc., ce qui constituerait un surchiffrement.

**Déchiffrement**

Pour déchiffrer, le principe est le même que pour le chiffrement: on prend les lettres deux par deux, puis on les multiplie par une matrice.

$$\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} \pmod{26}$$

**Cette matrice doit être l'inverse de matrice de chiffrement (modulo 26)**. Ordinairement l'inverse de la matrice $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ est:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Mais que cela signifie-t-il dans le contexte $Z_{26}$? Reprenons notre exemple.

**Exemple de déchiffrement**

Pour déchiffrer le message d'**Alice**, **Bob** doit calculer

$$\begin{pmatrix}9 & 4\\ 5 & 7\end{pmatrix}^{-1} = \frac{1}{43}\begin{pmatrix}7 & -4\\ -5 & 9\end{pmatrix} \pmod{26} = (43)^{-1}\begin{pmatrix}7 & -4\\ -5 & 9\end{pmatrix} \pmod{26}$$

Comme pgdc(43,26)=1, $(43)^{-1}$ existe dans $Z_{26}$ et $(43)^{-1}$ égale 23 . **Bob** a donc maintenant la matrice de déchiffrement:

$$\begin{pmatrix}9 & 4\\ 5 & 7\end{pmatrix}^{-1} = 23\begin{pmatrix}7 & -4\\ -5 & 9\end{pmatrix} \pmod{26} = \begin{pmatrix}161 & -92\\ -115 & 207\end{pmatrix} \pmod{26} = \begin{pmatrix}5 & 12\\ 15 & 25\end{pmatrix} \pmod{26}$$

**Bob** prend donc la matrice $\begin{pmatrix}5 & 12\\ 15 & 25\end{pmatrix}$ pour déchiffrer le message "FGXGE DSPGV". Après avoir remplacé les lettres par leur rang dans l'alphabet (A=1, B=2, etc.), il obtiendra:

$$P_1 \equiv 5\cdot 6 + 12\cdot 7 \pmod{26} = 114 \pmod{26} = 10$$
$$P_2 \equiv 15\cdot 6 + 25\cdot 7 \pmod{26} = 265 \pmod{26} = 5$$

Il fera de même avec les 3e et 4e lettres, 5e et 6e, etc. Il obtiendra finalement:

| Lettres chiffrées | F | G | X | G | E | D | S | P | G | V |
|---|---|---|---|---|---|---|---|---|---|---|
| Rangs chiffrés ($C_k$) | 6 | 7 | 24 | 7 | 5 | 4 | 19 | 16 | 7 | 22 |
| Rangs ($P_k$) | 10 | 5 | 22 | 15 | 21 | 19 | 1 | 9 | 13 | 5 |
| Lettres | j | e | v | o | u | s | a | i | m | e |

---

## 3.4 Substitutions polyalphabétiques

### 3.4.1 Chiffre de Vigenère (1568)

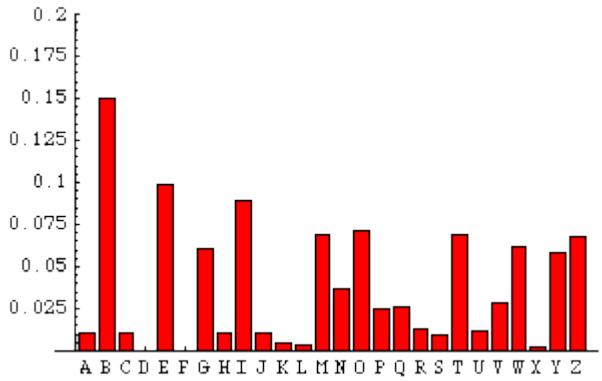
https://www.apprendre-en-ligne.net/crypto/vigenere/index.html

Le **chiffre de Vigenère** est une amélioration décisive du chiffre de César. Sa force réside dans l'utilisation non pas d'un, mais de 26 alphabets décalés pour chiffrer un message. On peut résumer ces décalages avec un carré de Vigenère. Ce chiffre utilise une **clef** qui définit le décalage pour chaque lettre du message (A: décalage de 0 cran, B: 1 cran, C: 2 crans, ..., Z: 25 crans).

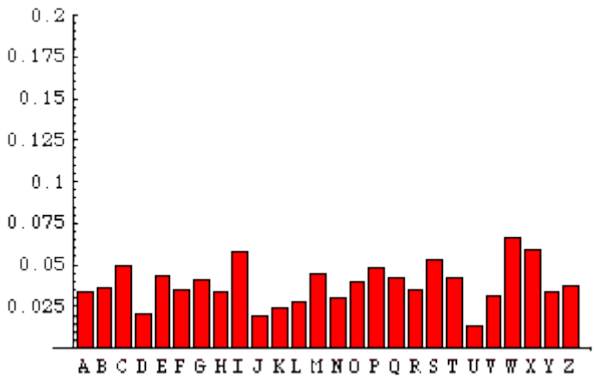
**Exemple:** chiffrons le texte "CHIFFRE DE VIGENERE" avec la clef "BACHELIER" (cette clef est éventuellement répétée plusieurs fois pour être aussi longue que le texte clair).

| | C | H | I | F | F | R | E | D | E | V | I | G | E | N | E | R | E |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Clef | B | A | C | H | E | L | I | E | R | B | A | C | H | E | L | I | E |
| Décalage | 1 | 0 | 2 | 7 | 4 | 11 | 8 | 4 | 17 | 1 | 0 | 2 | 7 | 4 | 11 | 8 | 4 |
| Chiffré | D | H | K | M | J | C | M | H | V | W | I | I | L | R | P | Z | I |

La grande force du chiffre de Vigenère est que la même lettre sera chiffrée de différentes manières. Par exemple le E du texte clair ci-dessus a été chiffré successivement M V L P I, **ce qui rend inutilisable l'analyse des fréquences classique**. Comparons les fréquences des lettres d'une fable de la Fontaine (Le chat, la belette et le petit lapin) chiffrée avec une substitution simple et celles de la même fable chiffrée avec le chiffre de Vigenère:




**Substitution simple**




**Chiffre de Vigenère**

On voit bien que l'histogramme n'a plus rien à voir avec celui d'une substitution simple: il est beaucoup plus "plat". Ce chiffre, qui a résisté trois siècles aux cryptanalystes, est pourtant relativement facile à casser, grâce à une méthode mise au point indépendamment par **Babagge** et **Kasiski**. Une autre méthode complètement différente a été encore mise au point plus tard par le commandant **Bazeries**.
Si la clef est aussi longue que le texte clair, et moyennant quelques précautions d'utilisation, le système est appelé masque jetable.

### 3.4.1.1 Cryptanalyse du Commandant Bazeries

Le commandant Bazeries est l'inventeur d'une méthode de décryptement plus simple et d'un emploi plus général que la méthode de Kasiski/Babbage. Elle se base sur l'existence d'un mot probable et préconise la recherche du mot-clef. Étant donné un cryptogramme chiffré au moyen du chiffre de Vigenère et renfermant un mot supposé connu, on "soustrait" le mot probable à une séquence du message chiffré de même longueur jusqu'à ce que la clef apparaisse.

**Exemple**
Soit le cryptogramme

> BILKO PFFGM LTWOE WJCFD SHKWO NKSEO VUSGR LWHGW FNVKW GGGFN
> RFHYJ VSGRF RIEKD CCGBH RYSXV KDIJA HCFFG YEFSG ZWG

qui est supposé renfermer le mot ATTAQUE.

En soustrayant ATTAQUE à la séquence débutant à la première position du cryptogramme, on obtient:

| Chiffré | B | I | L | K | O | P | F |
|---|---|---|---|---|---|---|---|
| Clair | A | T | T | A | Q | U | E |
| Décalage | -0 | -19 | -19 | -0 | -16 | -20 | -4 |
| Clef | B | P | S | K | Y | V | B |

ce qui semble ne rien donner. En commençant en position 2 on obtient:

| Chiffré | I | L | K | O | P | F | F |
|---|---|---|---|---|---|---|---|
| Clair | A | T | T | A | Q | U | E |
| Décalage | -0 | -19 | -19 | -0 | -16 | -20 | -4 |
| Clef | I | S | R | O | Z | L | B |

ce qui ne semble pas meilleur. On continue ainsi jusqu'à la position 25 où l'on voit apparaître:

| Chiffré | O | N | K | S | E | O | V |
|---|---|---|---|---|---|---|---|
| Clair | A | T | T | A | Q | U | E |
| Décalage | -0 | -19 | -19 | -0 | -16 | -20 | -4 |
| Clef | O | U | R | S | O | U | R |

Le mot OURS est apparu. C'est le mot-clef que l'on cherchait. En utilisant cette clef, le déchiffrement donne:

NOUS AVONS SUBI UNE VIOLENTE ATTAQUE CE MATIN. PERTES IMPORTANTES. DEMANDONS PILONNAGE DES POSITIONS ENNEMIES.

**3.4.1.2 Cryptanalyse de Kasiski**
https://www.apprendre-en-ligne.net/crypto/vigenere/decodevig.html

Cette technique consiste à chercher des séquences de lettres qui apparaissent plus d'une fois dans le texte. En effet, dans ce cas, il n'y aura que deux possibilités :

- soit la même séquence de lettres du texte clair a été chiffrée avec la même partie de la clef,
- soit deux suites de lettres différentes dans le texte clair auraient par pure coïncidence engendré la même suite dans le texte chiffré (probabilité faible).

Le 1er cas étant le plus probable, il en déduit le nombre de facteurs de la clef puis par une méthode de fréquence de distribution des lettres cryptées il en déduit les lettres du texte clair.

En prenant par exemple la clef KILO, la lettre E peut être chiffrée en O, M, P ou S selon que K, I, L ou O sont utilisés pour la chiffrer. Ainsi le mot *thé* peut être chiffré en DPP, BSS, EVO ou HRM.

**Exemple :**

| K | I | L | O | K | I | L | O | K | I | L | O | K | I | L | O | K | I | L | O | K | I | L | O | K |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| t | h | e | r | u | s | s | e | t | h | e | j | a | s | m | i | n | t | h | e | c | h | i | n | e |
| D | P | P | F | E | A | D | S | D | P | P | X | K | A | X | W | X | B | S | S | M | P | T | B | O |

Dans cet exemple THE est chiffré en DPP la première et la deuxième fois, et en BSS la troisième. C'est pourtant la faiblesse du chiffre de Vigenère: ces répétitions apparaissent parce que dans l'original, les mêmes séquences de lettres sont chiffrées avec la même partie de la clef.

**3.4.1.3 Cryptanalyse de Friedman**

Le **test de Friedman** (aussi appelé **test kappa**) s'appuie sur la métrique appelée Indice de Coïncidence (IC). Il a pour premier objectif de **déterminer si un texte a été chiffré avec un chiffre monoalphabétique ou polyalphabétique**. Comme second bénéfice, **il suggère la longueur du mot-clef si le chiffre est polyalphabétique**. L'Indice de Coïncidence (IC) est défini comme la probabilité que deux lettres choisies aléatoirement dans un texte soient identiques. Elle permet de déterminer la langue et la taille de la clef.

Voici quelques indices calculés sur des textes contemporains dans différentes langues.

| Langue | allemand | anglais | espagnol | esperanto | français | italien | norvégien | suédois |
|---|---|---|---|---|---|---|---|---|
| IC | 0.072 | 0.065 | 0.074 | 0.069 | 0.074 | 0.075 | 0.073 | 0.071 |

Soit le message suivant, chiffré avec Vigenère (369 lettres):
PERTQ UDCDJ XESCW MPNLV MIQDI ZTQFV XAKLR PICCP QSHZY DNCPW EAJWS ZGCLM QNRDE OHCGE ZTQZY HELEW AUQFR OICWH QMYRR UFGBY QSEPV NEQCS EEQWE EAGDS ZDCWE OHYDW QERLM FTCCQ UNCPP QSKPY FEQOI OHGPR EERWI EFSDM XSYGE UELEH USNLV GPMFV EIVXS USJPW HIEYS NLCDW MCRTZ MICYX MNMFZ QASLZ QCJPY DSTTK ZEPZR ECMYW OICYG UESIU GIRCE UTYTI ZTJPW HIEYI ETYYH USOFI XESCW HOGDM ZSNLV QSQPY JSCAV QSQLM QNRLP QSRLM XLCCG AMKPG QLYLY DAGEH GERCI RAGEI ZNMGI YBPP

On va considérer les sous-chaînes obtenues en prenant les lettres à intervalle donné:
Intervalle de 1: PERTQ UDCDJ XESCW MPNLV ... (texte original)
Intervalle de 2: PRQDD XSWPL ... et ETUCJ ECMNV ...
Intervalle de 3: PTDJS MLIIQ ... , EQCXC PVQZF... et RUDEW NMDTV
...
On calcule ensuite les IC pour toutes ces sous-chaînes.

| Intervalle | Indice de coïncidence |
|---|---|
| 1 | 0.0456107 |
| 2 | 0.0476954, 0.0443098 |
| 3 | 0.044249, 0.0494469, 0.0426771 |
| 4 | 0.0465839, 0.0453894, 0.0449116, 0.0425227 |
| 5 | 0.0799704, 0.0925583, 0.0836727, 0.0795282, 0.0684932 |
| 6 | 0.0512956, 0.0407192, 0.0371585, 0.0382514, 0.0661202, 0.0431694 |

On remarque que quand l'intervalle est de 5, **l'IC correspond plus ou moins avec l'IC caractéristique du français** (en tout cas, c'est cette ligne qui s'approche le plus de 0.074, les autres lignes étant plutôt proches de 0.038). **La longueur de la clef utilisée est donc probablement 5.** Pour découvrir la clef elle-même, on peut ensuite procéder comme le faisait Kasiski (ceci est laissé en exercice.

**L'indice de Coïncidence (IC) : Un peu de théorie**

L'**indice de Coïncidence** (IC) est la probabilité que deux lettres choisies aléatoirement dans un texte soient identiques. Il fut inventé par **William Friedman** et publié en 1920, dans l'article "The Index of Coincidence and its Applications in Cryptography", Riverbank Publications Number 22. Pour calculer cet indice, soient:

$n$ = nombre de lettres du texte
$n_1$ = nombre de A dans le texte
$n_2$ = nombre de B dans le texte
$n_3$ = nombre de C dans le texte
...
$n_{26}$ = nombre de Z dans le texte

La probabilité de tirer deux "A" parmi les n lettres d'un texte (clair ou crypté) est:

$$P(2\ fois\ A) = \frac{C_2^{n_1}}{C_2^n} = \frac{\frac{n_1(n_1-1)}{2}}{\frac{n(n-1)}{2}} = \frac{n_1(n_1-1)}{n(n-1)}$$

Pour calculer la probabilité de tirer deux lettres identiques, il faut faire la somme des 26 possibilités:

$$IC = \sum_{i=1}^{26} \frac{n_i(n_i - 1)}{n(n-1)}$$

**Remarques importantes**

1. Pour un langage de 26 lettres où chaque lettre a la même fréquence (1/26), IC = 0.038 (vérifiez ce résultat avec la formule ci-dessus).
2. Pour tout chiffre monoalphabétique, la distribution des fréquences est invariante, donc l'IC sera le même que pour le texte clair. Idem pour les chiffres de transposition.
3. Donc, si on calcule l'IC d'un texte chiffré avec un chiffre monoalphabétique, on devrait trouver un IC "proche" de 0.074 (en français). Si l'IC est beaucoup plus petit (p. ex. 0.050), le chiffre est probablement polyalphabétique.

### 3.4.2 Chiffre de Beaufort

Le chiffre de l'amiral anglais Sir Francis Beaufort (1774-1857) fut publié après sa mort par son frère. Il semblerait que ce chiffre ait en fait été inventé par Jean Sestri vers 1710.

Le chiffre de Beaufort est une variante du chiffre de Vigenère. Il utilise le carré de Vigenère d'une autre manière. Au lieu d'additionner la clef au message clair, Beaufort soustrait le message clair de la clef. Une propriété intéressante de ce chiffre est qu'il est réversible: si on chiffre deux fois de suite un message, on retrouve le message original.

**Exemple**

Chiffrons le texte "CHIFFRE DE BEAUFORT" avec la clef "BACHELIER" (les couleurs correspondent ici à celles utilisées dans le carré de Vigenère).

| Clef | B | A | C | H | E | L | I | E | R | B | A | C | H | E | L | I | E |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Clair | C | H | I | F | F | R | E | D | E | B | E | A | U | F | O | R | T |
| Décalage | -2 | -7 | -8 | -5 | -5 | -17 | -4 | -3 | -4 | -1 | -4 | 0 | -20 | -5 | -14 | -17 | -19 |
| Chiffré | Z | T | U | C | Z | U | E | B | N | A | W | C | N | Z | X | R | L |

---

### 3.4.2 Chiffre de Vernam

Le masque jetable est défini comme un chiffre de Vigenère avec la caractéristique que la clef de chiffrement a la même longueur que le message clair.

| Clair | M | A | S | Q | U | E | J | E | T | A | B | L | E |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Clef | X | C | A | A | T | E | L | P | R | V | G | Z | C |
| Décalage | 23 | 2 | 0 | 0 | 19 | 4 | 11 | 15 | 17 | 21 | 6 | 25 | 2 |
| Chiffré | J | C | S | Q | N | I | U | T | K | V | H | K | G |

Pour utiliser ce chiffrement, il faut respecter plusieurs propriétés :

- choisir une clef aussi longue que le texte à chiffrer,
- utiliser une clef formée d'une suite de caractères aléatoires,
- protéger votre clef,
- ne jamais réutiliser une clef.

Exemple illustrant l'inviolabilité :

Soit le texte chiffré : cuskqxwmfwituk
Soit le masque jetable possible : bgfbcdfbfdecdg
Résultat : BONJOURLATERRE
Soit un autre masque jetable : quauwtedbdisjg
Résultat : MASQUESJETABLE
Il est donc impossible de déterminer le bon masque !

Le système du masque jetable, avec les précautions indiquées ci-dessus, est absolument inviolable si l'on ne connaît pas la clef. Il est couramment utilisé de nos jours par les États. En effet, ceux-ci peuvent communiquer les clefs à leurs ambassades de manière sûre via la valise diplomatique.
Le problème de ce système est de communiquer les clefs de chiffrement ou de trouver un algorithme de génération de clef commun aux deux partenaires.
De plus, la création de grandes quantités des clefs aléatoires devient vite problématique. N'importe quel système couramment utilisé pourrait exiger des millions de caractères aléatoires de façon régulière.
La distribution des clés est également complexe. La longueur de la clé étant égale à celle du message,une bonne organisation est nécessaire.

**3.5 Transpositions**
Un chiffre de transposition consiste à changer l'ordre des lettres, donc à construire des anagrammes. Cette méthode est connue depuis l'Antiquité. Une analyse statistique sur les chiffrements par transposition n'est pas utile, puisque seul l'ordre des symboles est différent; les symboles restent les mêmes. Donc, les symboles les plus fréquents dans le message clair resteront évidemment les plus fréquents dans le message chiffré.

Pour de très brefs messages, comme un simple mot, cette méthode est peu sûre car il n'y a guère de variantes pour redistribuer une poignée de lettres. Par exemple un mot de trois lettres ne peut être tourné quand dans 6 (=3!) positions différentes. Ainsi "col" ne peut se transformer qu'en "col", "clo", "ocl", "olc", "lco" ou "loc". Bien entendu, lorsque le nombre de lettres croît, le nombre d'arrangements augmente rapidement et il devient quasiment impossible de retrouver le texte original sans connaître le procédé de brouillage.

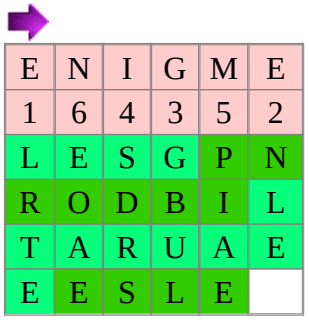
## 3.5.1. La scytale spartiate
Une forme de transposition utilise le premier dispositif de cryptographie militaire connu, la scytale spartiate, remontant au Ve siècle avant J.-C. La scytale consiste en un bâton de bois autour duquel est entourée une bande de cuir ou de parchemin, comme le montre la figure ci-dessous. L'expéditeur écrit son message sur toute la longueur de la scytale et déroule ensuite la bande qui apparaît alors couverte d'une suite de lettres sans signification. Le messager emportera la bande de cuir, l'utilisant comme ceinture, les lettres tournées vers l'intérieur. Le destinataire enroulera alors cette bande sur son bâton (de même diamètre) pour lire le message clair.

## 3.5.2. Chiffre de Ubchi
Le chiffre UBCHI était utilisé par les Allemands au tout début de la première guerre mondiale, mais John Falconer mentionne déjà les principes de ce chiffre en 1685 dans son ouvrage *Cryptomenysices Patefacta*.

**Chiffrement**
Chiffrons le message "LE LOUP EST DANS LA BERGERIE". Nous avons prévenu au préalable notre destinataire que nous allions utiliser le mot-clef ENIGME, qui correspond au **chiffre-clef 164352** (on numérote les lettres du mot-clef selon l'ordre alphabétique). Note: dans l'exemple ci-dessous, les deux nuances de vert sont simplement là pour aider à comprendre le mécanisme de chiffrement.
La première opération consiste à réécrire les colonnes horizontalement, en suivant leur

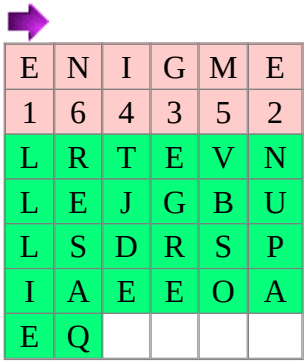
numérotation (on recopie la colonne 1, puis la 2, etc.). On peut ensuite ajouter *n* nulles (3 dans l'exemple ci-dessous: J, V et Q), puis on répète la première opération une seconde fois.

| E | N | I | G | M | E |
|---|---|---|---|---|---|
| 1 | 6 | 4 | 3 | 5 | 2 |
| L | E | L | O | U | P |
| E | S | T | D | A | N |
| S | L | A | B | E | R |
| G | E | R | I | E | |



| E | N | I | G | M | E |
|---|---|---|---|---|---|
| 1 | 6 | 4 | 3 | 5 | 2 |
| L | E | S | G | P | N |
| R | O | D | B | I | L |
| T | A | R | U | A | E |
| E | E | S | L | E | |

| E | N | I | G | M | E |
|---|---|---|---|---|---|
| 1 | 6 | 4 | 3 | 5 | 2 |
| L | E | S | G | P | N |
| R | O | D | B | I | L |
| T | A | R | U | A | E |
| E | E | S | L | E | J |
| V | Q | | | | |



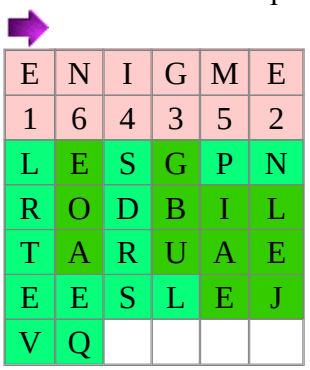
| E | N | I | G | M | E |
|---|---|---|---|---|---|
| 1 | 6 | 4 | 3 | 5 | 2 |
| L | R | T | E | V | N |
| L | E | J | G | B | U |
| L | S | D | R | S | P |
| I | A | E | E | O | A |
| E | Q | | | | |

Le message chiffré est donc: LRTEV NLEJG BULSD RSPIA EEOAE Q.

**Déchiffrement**

Pour déchiffrer, on réécrit au préalable le message ligne par ligne dans une grille qui a autant de colonnes qu'il y a de lettres dans le mot-clef. La première opération consiste à lire **les lignes** du message de haut en bas et **recopier les lettres colonne par colonne, en suivant leur numérotation, et en ne remplissant que les cases vertes** (on remplit d'abord la colonne 1, puis la 2, etc.). On supprime dans un deuxième temps les nulles (les *n* dernières lettres du tableau), puis on répète la première opération une seconde fois pour retrouver le message clair.

| E | N | I | G | M | E |
|---|---|---|---|---|---|
| 1 | 6 | 4 | 3 | 5 | 2 |
| L | R | T | E | V | N |
| L | E | J | G | B | U |
| L | S | D | R | S | P |
| I | A | E | E | O | A |
| E | Q | | | | |



| E | N | I | G | M | E |
|---|---|---|---|---|---|
| 1 | 6 | 4 | 3 | 5 | 2 |
| L | E | S | G | P | N |
| R | O | D | B | I | L |
| T | A | R | U | A | E |
| E | E | S | L | E | J |
| V | Q | | | | |

| E | N | I | G | M | E |
|---|---|---|---|---|---|
| 1 | 6 | 4 | 3 | 5 | 2 |
| L | E | S | G | P | N |
| R | O | D | B | I | L |
| T | A | R | U | A | E |
| E | E | S | L | E | |

| E | N | I | G | M | E |
|---|---|---|---|---|---|
| 1 | 6 | 4 | 3 | 5 | 2 |
| L | E | L | O | U | P |
| E | S | T | D | A | N |
| S | L | A | B | E | R |
| G | E | R | I | E | |

Un cas particulier d'algorithmes de chiffrement par blocs avec itération est la famille des **chiffres de Feistel**. Dans ce système de chiffrement, un bloc de texte en clair est découpé en deux ; la transformation de ronde est appliquée à une des deux moitiés, et le résultat est combiné avec l'autre moitié par ou exclusif. Les deux moitiés sont alors inversées pour l'application de la ronde suivante. **Un avantage de ce type d'algorithmes est que chiffrement et déchiffrement sont structurellement identiques**.

À titre d'exemple, nous allons chiffrer par un réseau de Feistel à deux rondes un message constitué de quatre bits ($2^4$ = 16 possibilités de messages), ce qui revient à construire une bijection de quatre bits vers quatre bits à partir de deux fonctions $f_1$ et $f_2$ de deux bits vers deux bits.

Nous considérerons que pour une certaine clef entrée, ces fonctions sont les suivantes:

## 4.1. Chiffrement Symétrique

Deux grandes catégories de chiffrement symétrique :

1. **Chiffrement par blocs :** Le message en clair M est une suite de bits (i.e. suite de 0 et 1). Il est découpé en blocs de mème taille (ex: 64 bits ou 128 bits). Exemple d'algorithmes : DES, AES, IDEA, RC6, BLOWFISH, ...
2. **Chiffrement par flots :** Le message en clair M est une suite de bits. Il est traité bit par bit. Exemple d'algorithmes : RC4, Bluetooth E0/1, GSM A5/1,

### 4.1.1. Chiffrement par blocs

#### 4.1.1.1 Approche de la notion de chiffrement par blocs

**De la cryptographie classique aux cryptosystèmes symétriques modernes :**

1. Des chiffrements classiques (par transposition et par substitution) vus au chapitre 3 sont des chiffrements par blocs .
2. **La substitution ajoute de la confusion** au procédé de chiffrement et **la transposition ajoute de la diffusion en éparpillant l'influence moyenne (selon les différentes clés) de chaque bit du clair**, sur les bits du chiffré.
3. Mais **aucun de ces deux procédés ne produit à la fois de la confusion et de la diffusion** ==> pas une réelle sécurité.
4. **Les systèmes modernes, pour assurer une véritable sécurité, doivent produire à la fois de la confusion et de la diffusion,** faute de quoi ils ne résistent pas aux attaques que nous décrirons plus loin.

**Blocs, clés, taille des blocs et taille des clés :**

1. Dans un système par blocs, chaque texte clair est découpé en blocs de même longueur et chiffré bloc par bloc.
2. La taille du texte clair est fixe. Ceci donne plus de maîtrise sur les propriétés du chiffrement
3. Le système de chiffrement par blocs le plus utilisé jusqu'à l'an 2000 est le DES
4. La taille de bloc a un impact sur la sécurité (n = 64 ou 128 bits). **Les modes opératoires permettent des attaques quand plus de $2^{n/2}$ blocs sont chiffrés avec une même clé.**
5. **La taille de la clé soit être suffisamment grande (k >128 bits).** Dans un bon algorithme, la meilleure attaque doit coûter au moins $2^k$ opérations ( la technique généralement utilisée est l'attaque exhaustive , encore appelée attaque par force brute).
6. Exemple:
   - AES: n= 128 bits , k=128, 192, 256 bits
   - 3 DES: n= 64bits k=… , 168 bits

**Quelques notations, notion de permutation et notion de chiffrement:**

1. $E_k(m)$ désigne le chiffrement du message m par la clé k
2. $D_k(m)$ désigne le chiffrement du message m par la clé k
3. Le message m est découpé en blocs de k bits: $m = m_1 m_2 m_3 ... m_p]$ où $mi \in \{0,1\}^k$.
4. La fonction de chiffrement est une permutation aléatoire. Une permutation est une bijection particulière: $E : \{0,1\}^k \rightarrow \{0,1\}^k$.
5. Exemple de permutation sur deux bits :
   00 → 11
   01 → 10
   10 → 11
   11 → 00
   En suivant cette permutation, le chiffrement 00 est 11, celui de 01 est 10, celui de 10 est 11, celui de 11 est 00.
6. Le message m est chiffré bloc par bloc :

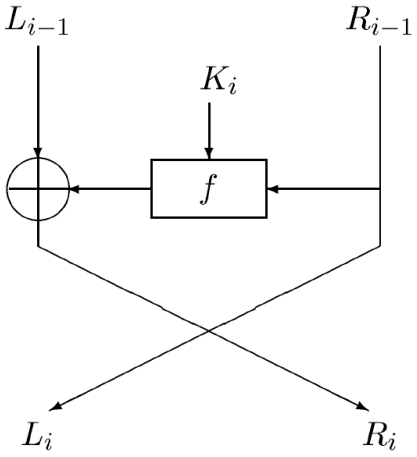
$E_k(m) = E_k(m_1 m^2 m_3 ... m_p)$
$= E_k(m_1)E_k(m_2)... E_k(m_p).$
$= c_1 c_2 c_3 \ldots c_p$

où $c_i$ est le chiffrement du bloc $m_i$ par la fonction de chiffrement E.

**Principes fondamentaux pour le chiffrement par blocs :**

1. Le message est découpé en blocs de taille fixe et chiffré bloc par bloc. Le chiffrement est basé sur la confusion et la diffusion.
2. **La confusion** vise à cacher n'importe quelle structure algébrique dans le système pour la rendre intelligible (table de substitution S-box).
3. **La diffusion** doit permettre à chaque bit de texte clair d'avoir une influence sur une grande partie du texte chiffré. Ce qui signifie que la modification d'un bit du bloc d'entrée doit entraîner la modification de nombreux bits du bloc de sortie correspondant.
4. Les 2 notions ont été introduites par Shannon :
   - La confusion est assurée par une substitution non-linéaire (S-Box)
   - La diffusion est assurée par une permutation linéaire

### 4.1.1.2. Réseau de Feistel




Le système de chiffrement par blocs le plus utilisé jusqu'à l'an 2000 est le DES. Il fait partie de la classe plus générale des chiffrements de Feistel.
Feistel est un un cas particulier d'algorithmes de chiffrement par blocs avec itération est la famille des chiffres de Dans ce système de chiffrement, **un bloc de texte en clair est découpé en deux ; la transformation de ronde est appliquée à une des deux moitiés, et le résultat est combiné avec l'autre moitié par ou exclusif. Les deux moitiés sont alors inversées pour l'application de la ronde suivante :**

$$L_i = R_{i-1} \text{ et } R_i = L_{i-1} + f(R_{i-1}, K_i)$$

**où $K_i$ est une partie de la clé K.**.
Un avantage de ce type d'algorithmes est que chiffrement et déchiffrement sont structurellement identiques.

**Rappel de Logique :**

1. **OU Inclusif :** Aussi appelé disjonction inclusive :A + B est vrai si l'un des deux argument est vrai et faux dans tous les autres cas.
2. **OU Exclusif :** Aussi appelé XOR (eXclusive OR) ou disjonction exclusive :A ⊕ B est vrai si l'un des deux argument est vrai et faux dans tous les autres cas.
3. **ET Logique** :AB est vrai si l'un les deux arguments sont est vrais et faux dans tous les autres cas.
4. **Négation Logique :** La valeur de la négation de A est l'opposé de A.

| A | B | OU Inclusif A + B | OU Exclusif A ⊕ B | ET Logique A B | NON Logique Négation deA |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 1 | 1 | 1 | 0 | |
| 1 | 0 | 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 0 | 1 | |

**Exemple de réseau de Feistel à deux rondes :**

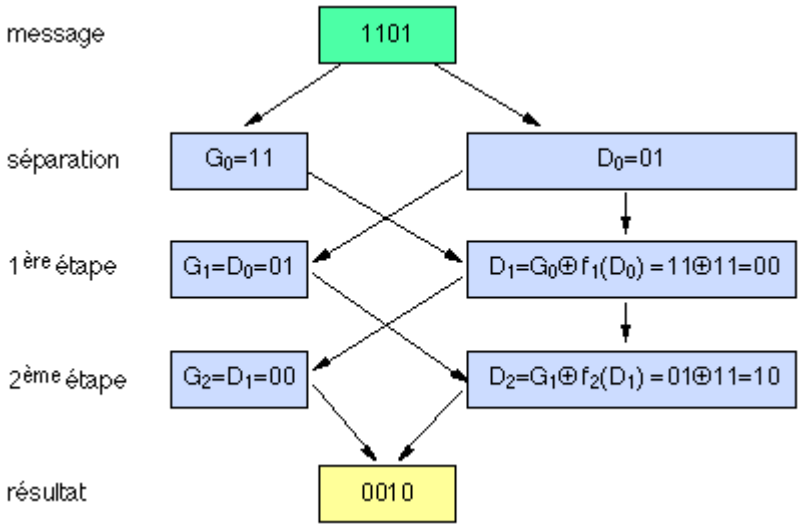
À titre d'exemple, nous allons chiffrer par un réseau de Feistel à deux rondes un message constitué de quatre bits ($2^4$ = 16 possibilités de messages), ce qui revient à construire une bijection de quatre bits vers quatre bits à partir de deux fonctions f1 et f2 de deux bits vers deux bits. Nous considérerons que pour une certaine clef entrée, ces fonctions sont les suivantes:

| **f1 :** Entrée | | Sortie |
|---|---|---|
| 00 | → | 01 |
| 01 | → | 11 |
| 10 | → | 10 |
| 11 | → | 01 |

| **f2 :** Entrée | | Sortie |
|---|---|---|
| 00 | → | 11 |
| 01 | → | 00 |
| 10 | → | 00 |
| 11 | → | 01 |

Notons que ni f1 ni f2 ne sont des bijections. À titre d'exemple, chiffrons le message 1101. G désigne la moitié gauche du message à chiffrer, D la moitié droite. Le schéma du réseau de Feistel (à deux rondes) obtenu à partir de f1 et f2 est donné à droite en y illustrant le chiffrement de 1101. Les chiffrements obtenus des 16 messages possibles sont dans le tableau (à gauche)..

| **Message en clair** | **Chiffrement du message** |
|---|---|
| 0000 | 0100 |
| 0001 | 1100 |
| 0010 | 1010 |
| 0011 | 0111 |
| 0100 | 0011 |
| 0101 | 1001 |
| 0110 | 1111 |
| 0111 | 0000 |
| 1000 | 1101 |
| 1001 | 0101 |
| 1010 | 0001 |
| 1011 | 1110 |
| 1100 | 1000 |
| 1101 | 0010 |
| 1110 | 0110 |
| 1111 | 1011 |




**Théorème de Sécurité des Réseaux Feistel (Luby-Rackoff, 1985):** Si une fonction aléatoire sûre est utilisée pour trois tours de Feistel avec trois clés indépendantes, on obtient alors une fonction pseudo aléatoire avec des permutations pseudo aléatoire .

**Exemple:** DES utilise 16 rondes (ou cycles) du réseau Feistel

### 4.1.1.3. D.E.S. - Data Encryption Standard

**Présentation de DES :**

**Histoire du DES :**

1. **1970:** Chiffrement par bloc « Lucifer » développé par IBM : k=128 et m=128
2. **1973:** DES (Data Encryption Standard) Adopté comme standard US par le Bureau National des Standards Américains NBS (FIPS 46-2) pour le chiffrement par blocs
3. **1975:** DES a été choisi comme norme au Etats-Unis est devenu le système cryptographique le plus utilisé dans le monde.
4. **1976:** Le DES, un variant de Lucifer, est adopté comme ce standard : Taille de bloc = 64 bits, k = 56 bits, m = 64
5. **1997:** Le DES commence à être critiqué à cause de la taille trop faible de sa taille de clés k = 56
6. **1997:** DES cassé par la recherche exhaustive (AES en 2000.)
7. **1998 :** Le défi « DES Challenge » a été lancé pour casser DES: la machine "Deep Crack" (spécialement conçu pour attaquer DES) a réussi en quelques jours à retrouver la clé par une attaque exhaustive.

#### 4.1.1.4. A.E.S. - Advanced Encryption Standard

### 4.1.2. Chiffrement par flots

## 4.2. Chiffrement par clé publique

### 4.2.1. Chiffre de Merkle-Hellman

Le premier cryptosystème à clef publique, qui fut proposé par Ralph Merkle (photo de gauche) et Martin Hellman en 1978, est basé sur le problème du sac à dos (*Knapsack problem* en anglais). Il n'est plus utilisé actuellement puisque ce chiffre, ainsi que de nombreuses variantes, a été cassé au début des années 80 par Adi Shamir.

**Le problème du sac à dos** consiste à empiler des objets dans un sac, de manière à atteindre (si possible) un poids total fixé. Plus formellement, étant donnés des poids entiers $P_1$, ..., $P_n$ et un but T, il s'agit de trouver $b_1$, ..., $b_n$, valant 0 ou 1, tels que

$$\mathbf{T = b_1P_1 + b_2P_2 + ... + b_nP_n}$$

Si la suite des poids $P_k$ est supercroissante (chaque poids est strictement supérieur à la somme de tous les précédents), alors il existe une méthode de résolution simple (algorithme glouton):

On peut vérifier qu'avec la suite de poids supercroissante $P_1$=2, $P_2$=3, $P_3$=6, $P_4$=12 et T=15 on obtient la solution $b_1$=0, $b_2$=1, $b_3$=0, $b_4$=1.

Au contraire, si la suite des poids n'est pas supercroissante, **le seul algorithme connu consiste à essayer successivement toutes les solutions** ($b_1$, $b_2$, ..., $b_n$) possibles. Si la suite est suffisamment longue, c'est un algorithme impraticable.

**Suite supercroissante**

Une **suite supercroissante** est une suite croissante de *n* nombres tels que $s_j > \sum_{i=1}^{j-1} s_i$ pour *j* compris entre 2 et *n*. Cela signifie que le $j^{ème}$ terme doit être plus grand que la somme des *j*-1 termes qui le précèdent dans la suite.


Les pièces de monnaie en Euro forment une suite supercroissante. En effet:

- 1 ct + 2 ct < 5 ct
- 1 ct + 2 ct + 5 ct < 10 ct
- 1 ct + 2 ct + 5 ct + 10 ct < 20 ct
- etc.

Il en va de même pour toutes les monnaies.

**Méthodes de chiffrement et de déchiffrement**
C'est un chiffre à clef publique, basé sur la difficulté de résoudre le problème du sac à dos avec une suite de poids non supercroissante. On part de l'observation suivante:

> Le problème du sac à dos est **homogène** : on ne change pas les solutions en multipliant (ou en divisant) tous les poids $P_i$ et le but T par un même entier p.

Les multiplications peuvent être effectuées modulo un entier m. Alors, même si la suite de poids initiale est supercroissante, la nouvelle suite ainsi obtenue ne le sera en général pas. En pratique, **on choisit m supérieur à la somme des poids initiaux**.

**Exemple.** On part de la suite supercroissante **1, 3, 6**. Tous les calculs sont faits modulo **12**. En multipliant tous les poids par 7 on obtient la suite non supercroissante **7, 9, 6**. Grâce à la propriété d'homogénéité, on voit que la même solution ($b_1$, $b_2$, $b_3$) permet de réaliser le but T avec les poids 1, 3, 6 et le but Tx7 avec les poids 7, 9, 6. Le tableau ci-dessous donne la correspondance entre les buts et les solutions:

| T | Tx7 | $b_1$, $b_2$, $b_3$ |
|---|---|---|
| 0 | 0 | 0, 0, 0 |
| 1 | 7 | 1, 0, 0 |
| 3 | 9 | 0, 1, 0 |
| 4 | 4 | 1, 1, 0 |
| 6 | 6 | 0, 0, 1 |
| 7 | 1 | 1, 0, 1 |
| 9 | 3 | 0, 1, 1 |
| 10 | 10 | 1, 1, 1 |

Pour chiffrer un bloc de k bits, on calcule simplement le poids T résultant du sac. Pour garantir que le déchiffrement avec la même clef est impossible, **on chiffre avec la suite non supercroissante**. Pour déchiffrer, on doit déterminer les poids dont la somme réalise le but T. L'idée consiste à revenir à la suite supercroissante initiale, sans changer la solution. Il suffit pour cela de diviser tous les poids et le but T par p. La suite non supercroissante constitue la **clef publique**. La suite supercroissante initiale, avec p et m, forment la **clef privée**.

**Premier exemple de chiffrement/déchiffrement**
La clef privée de Bob est [**1**, **3**, **6**], avec p = **7** et m = **12**.
Avec la clef publique [**7, 9, 6**] fournie par Bob, Alice chiffre ainsi la chaîne 101: T = **7**x1 + **9**x0 + **6**x1 = 13. Elle envoie donc le message "13" à Bob.
Pour déchiffrer le message, Bob utilise sa clef privée (notez que 7 est son propre inverse modulo 12), et il applique l'algorithme glouton. Il obtient:

> $13x7^{-1} \pmod{12} = 91 \pmod{12} = 7 \pmod{12}$ = **1**x1 + **3**x0 + **6**x1. Le message clair est donc 101.

**Plus généralement, pour que p soit inversible modulo m, il suffit de le choisir premier avec m.**
Il est important de noter qu'un texte chiffré avec la clef privée ne pourra pas être déchiffré avec la clef publique, puisqu'elle n'est pas supercroissante: les deux clefs du chiffre de Merkle-Hellman ne peuvent pas être permutées.

**Deuxième exemple de chiffrement/déchiffrement**

En utilisant 5 bits, on peut coder $2^5 = 32$ caractères. Supposons que l'on utilise la table des 30 caractères suivants (ce tableau vous rappelle peut-être l'alphabet bilitère de Bacon):

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 00000 | | g | 00110 | | m | 01100 | | s | 10010 | | y | 11000 |
| b | 00001 | | h | 00111 | | n | 01101 | | t | 10011 | | z | 11001 |
| c | 00010 | | i | 01000 | | o | 01110 | | u | 10100 | | , | 11010 |
| d | 00011 | | j | 01001 | | p | 01111 | | v | 10101 | | . | 11011 |
| e | 00100 | | k | 01010 | | q | 10000 | | w | 10110 | | ? | 11100 |
| f | 00101 | | l | 01011 | | r | 10001 | | x | 10111 | | | 11101 |

La clef privée de Bob est [**3**, **4**, **9**, **19**, **38**, **77**], avec p = **27** et m = **155**. Il peut maintenant calculer la clef publique avec laquelle Alice chiffrera son message: [3x27 (mod 155), 4x27 (mod 155), 9x27 (mod 155), 19x27 (mod 155), 38x27 (mod 155), 77x27 (mod 155)], ce qui donne [**81**, **108**, **88**, **48**, **96**, **64**].

Alice veut transmettre le mot "baby" à Bob. Elle devra donc, en se référant à la table ci-dessus, chiffrer la chaîne 00001 00000 00001 11000. Comme la clef publique de Bob comporte 6 nombres, elle devra ensuite regrouper ces bits par paquets de 6, et au besoin ajouter des bits aléatoires pour obtenir un nombre de bits multiple le 6: 000010 000000 001110 001101

Le premier bloc de 6 bits sera chiffré : **81**x0 + **108**x0 + **88**x0 + **48**x0 + **96**x1 + **64**x0 = 96.
Le deuxième bloc de 6 bits sera chiffré : **81**x0 + **108**x0 + **88**x0 + **48**x0 + **96**x0 + **64**x0 = 0.
Le troisième bloc de 6 bits sera chiffré : **81**x0 + **108**x0 + **88**x1 + **48**x1 + **96**x1 + **64**x0 = 232.
Le quatrième bloc de 6 bits sera chiffré : **81**x0 + **108**x0 + **88**x1 + **48**x1 + **96**x0 + **64**x1 = 200.
Le message chiffré est donc: [96, 0, 232, 200]

Pour déchiffrer, Bob calcule d'abord l'inverse modulo **155** de **27**, qui vaut 23. Il utilise ensuite sa clef privée et applique l'algorithme glouton. Il obtient alors:

96x23 (modulo 155) = 2208 (modulo 155) = 38 = 000010.
0x23 (modulo 155) = 0 (modulo 155) = 0 = 000000.
232x23 (modulo 155) = 5336 (modulo 155) = 66 = 9 + 19 + 38 = 001110.
200x23 (modulo 155) = 4600 (modulo 155) = 105 = 9 + 19 + 77 = 001101.

Le message déchiffré est donc [000010, 000000, 001110, 001101], ce qui est bien le message qu'avait envoyé Alice. Pour retrouver le mot, Bob n'a plus qu'à regrouper les bits par paquets de 5 et consulter le tableau de conversion.

**Il est important que le nombre de bits par paquet soit différent de la longueur de la clef. En effet, si ce n'était pas le cas, on pourrait attaquer le texte chiffré par une analyse des fréquences, car chaque lettre serait chiffrée par le même nombre.**

Il est clair que plus la clef est longue, plus le message sera difficile à décrypter. Dans la pratique, on utilise au moins 250 nombres dans la clef et le module m est choisi pour avoir une longueur comprise entre 100 et 200 bits.

**Algorithme glouton**
**Pour** i = n **à** 1 **faire**
  **Si** $T \geq P_i$ **alors**
    $T = T - P_i$
    $b_i = 1$
  **sinon**

$b_i = 0$

**Si** T = 0 alors $\{b_1, ..., b_n\}$ est solution **sinon** il n'y a pas de solution.

### 4.2.2. Chiffre de Rivest - Shamir - Adleman (RSA)

L'implémentation fut achevée en 1978 par Rivest, Shamir et Adleman. Depuis, ce système de chiffrement est appelé RSA, qui sont les initiales de ces trois chercheurs.

Il est basé sur le calcul exponentiel. Sa sécurité repose sur la fonction unidirectionnelle suivante : le calcul du produit de 2 nombres premiers est aisé. La factorisation d'un nombre en ses deux facteurs premiers est beaucoup plus complexe.

Il s'agit du système le plus connu et le plus largement répandu, basé sur l'élévation à une puissance dans un champ fini sur des nombres entiers modulo un nombre premier. Le nombre d'exponentiation prend environ $O((\log n)^3)$ opérations ce qui est rapide et facile. Il emploie de grands nombres entiers (par exemple représentés sur 1024 bits).

Ce cryptosystème utilise deux clés d et e, interchangeables Le chiffrement se fait selon

$$\mathbf{C = M^e \bmod n}$$

et le déchiffrement par

$$\mathbf{M = C^d \bmod n.}$$

La sécurité repose sur le coût nécessaire pour factoriser de grands nombres. Le nombre de factorisation prend environ $O(e^{\log(n)\ \log(\log\ (n))})$ opérations ce qui demande un temps de calcul trop important pour les machines actuelles, dans un cadre privé. On l'utilise pour la confidentialité, l'authentification, ou encore une combinaison des 2.

On appellera **Bob** la personne qui désire recevoir un message chiffré, et **Alice** la personne qui envoie le message.

**Choix de la clef**

**Bob** choisit deux grands entiers naturels premiers **p** et **q** (d'environ 100 chiffres chacun ou plus) et fait leur produit **n = p·q**. Puis il choisit un entier **e** premier avec **(p-1)·(q-1)**. Enfin, il publie dans un annuaire, par exemple sur le web, **sa clef publique**: **(RSA, n, e)**.

**Chiffrement**

**Alice** veut donc envoyer un message à **Bob**. Elle cherche dans l'annuaire la clef de chiffrement qu'il a publiée. Elle sait maintenant qu'elle doit utiliser le système RSA avec les deux entiers **n** et **e** (prenons par exemple **n**=5141=53*97 et **e**=7, premier avec 52*96=4992). Elle transforme en nombres son message en remplaçant par exemple chaque lettre par son rang dans l'alphabet.

"JEVOUSAIME" devient : "10 05 22 15 21 19 01 09 13 05".

Puis elle découpe son message chiffré en blocs de même longueur représentant chacun un nombre plus petit que **n**. Cette opération est essentielle, **car si on ne faisait pas des blocs assez longs (par exemple si on laissait des blocs de 2 dans notre exemple), on retomberait sur un simple chiffre de substitution que l'on pourrait attaquer par l'analyse des fréquences.**
Son message devient : "010 052 215 211 901 091 305"

Un bloc **B** est chiffré par la formule $\mathbf{C = B^e \bmod n}$, où **C** est un bloc du message chiffré qu'**Alice** enverra à **Bob**. Après avoir chiffré chaque bloc, le message chiffré s'écrit :

"0755 1324 2823 3550 3763 2237 2052".

**Déchiffrement**
**Bob** calcule à partir de **p** et **q**, **qu'il a gardés secrets**, la clef **d** de déchiffrage (c'est sa **clef privée**). Celle-ci doit satisfaire l'équation **e·d mod ((p-1)(q-1)) = 1**. Ici, **d**=4279.
Chacun des blocs **C** du message chiffré sera déchiffré par la formule $\mathbf{B = C^d \bmod n}$.
Il retrouve : "010 052 215 211 901 091 305"

**Sécurité de la méthode**
Tout l'intérêt du système RSA repose sur le fait qu'à l'heure actuelle il est pratiquement impossible de retrouver dans un temps raisonnable **p** et **q** à partir de **n** si celui-ci est très grand (ou alors, si c'est possible, les cryptanalystes qui ont trouvé la méthode la gardent secrète). **Bob** est donc le seulà pouvoir calculer **d** dans un temps court. De plus, il n'a jamais à transmettre les entiers **p** et **q**, ce qui empêche leur piratage.

**Résumé sur RSA : Clés, Chiffrement et Déchiffrement**
1. Génération de 2 nombres premiers très grands p et q
2. Calcul de n = p*q et $\Phi(n)$ =(p-1)*(q-1)
3. Déterminer e tel que $3 < e < \Phi(n)$ et $(e, \Phi(n)) = 1$
4. Calculer d tel que $e * d \equiv 1 \bmod \Phi(n)$
5. Clé publique : (e,n)
6. Clé privée : (d,n)
7. p et q doivent rester secrets, voire supprimés
8. Chiffrement du message en clair M : $C = M^e \bmod n$
9. Déchiffrement : $M = C^d \bmod n$

**Conseils d'utilisation du RSA**
Pour garantir une bonne sécurité, il faut respecter certains règles telles que :
1. Ne jamais utiliser de valeur n trop petite,
2. N'utiliser que des clés fortes (p-1 et q-1 ont un grand facteur premier),
3. Ne pas chiffrer de blocs trop courts,
4. Ne pas utiliser de n communs à plusieurs clés,
5. Si (d,n) est compromise ne plus utiliser n.

**4.2.3. Chiffre de El Gamal**
C'est un algorithme à clef publique présent à la base de la norme U.S. de signature électronique. Il fut inventé par Taher ElGamal en 1984. Il est basé sur la difficulté de calculer des logarithmes discrets. Le problème du logarithme discret consiste à retrouver un entier $\lambda$ tel que

$$h = g^{\lambda} \bmod p.$$

**Principe du chiffrement**
Soit un entier premier p très grand et $p - 1$ doit avoir un grand facteur premier. On produit :
1. une clé secrète s, telle que $s \in (1...p - 2)$,
2. une clé publique reposant sur l'entier p, un entier a premier avec p, et l'entier P tel que

$$P = a^s \bmod p$$

Le nombre a est pris tel que $a \in (0...p - 1)$ et $\forall k \in (1...p - 2)$ :

$$a^k <> 1 \bmod p$$

Soit un message M , avec M < p. On détermine un nombre aléatoire k qui n'est connu que de celui qui chiffre et différent à chaque message. On calcule alors

$$C1 = a^k \bmod p$$
$$C2 = M.P^k \bmod p$$

On obtient alors le message chiffré C = (C1 , C2). Le message chiffré est alors deux fois plus long que le message original

**Principe du déchiffrement**
A la réception, on calcule

$$\begin{aligned} R1 &= (C1)^s \bmod p \\ &= a^{sk} \bmod p \\ &= P^k \bmod p \end{aligned}$$

Le destinataire possède la clé privée (s). Ayant $P^k$, on divise C2 par cette valeur :

$$\begin{aligned} D_K(C) &= C2 /(R1) \\ &= (M.P^k \bmod p)/(P^k \bmod p) \\ &= M \end{aligned}$$

$P^k$ est donc considéré comme un masque appliqué sous forme multiplicative à M. Pour décrypter le message, il faudra soit trouver directement un masque jetable, soit trouver la clé privée s, solution de $P = a^s \bmod p$ (et donc trouver le logarithme discret).

**Exemple :** Soient p = 2579, a = 2, s = 765. Il vient
– Clé privée S k = (765)
– Clé publique P k = (2579, 2, 949) car $2^{765} \bmod 2579 = 949$
Pour chiffrer M = 1299, on choisit k = 853. Il vient
C 1 = $2^{853} \bmod 2579 = 435$
C 2 = $1299 * 949^{853} \bmod 2579 = 2396$
On peut effectivement vérifier que $2396/(435^{765}) \bmod 2579 = 1299$.

**Efficacité et sécurité**
**El Gamal est 2 fois plus lent que le RSA. L'inconvénient majeur reste la taille des données chiffrées qui représente 2 fois celle des données en clair.**
La recherche de la clé privée (s) à partir de la clé publique est équivalente au problème du logarithme discret (NP). MAIS il n'est pas prouvé que la cryptanalyse d'un message chiffré avec El Gamal est équivalente au logarithme discret. En d'autres termes, si le problème du logarithme est résolu polynomialement, alors El Gamal sera cassé. Cependant, rien ne prouve qu'il n'est pas cassable par un autre moyen.

### 4.2.4. Chiffre de Rabin

Prérequis: Arithmétique modulo m,
algorithme d'Euclide étendu,
théorème des restes chinois

Le **chiffre de Rabin** offre une sécurité calculatoire équivalente au problème de la factorisation de **n=pq.** C'est un cryptosystème à clefs publiques.

**Soit n le produit de deux nombres premiers distincts p et q tels que p et q soient congrus à 3 modulo 4** (nous verrons plus loin le pourquoi de cette condition). Soit le nombre **B** compris entre 0 et **n**-1. Le nombre chiffré **y** (compris entre 0 et **n**-1) s'obtient à partir du nombre clair **x** (compris entre 0 et **n**-1) par la formule:

$$y = x(x + B) \mod n$$

Inversement, le nombre clair **x** s'obtient à partir du nombre chiffré **y** par la formule:

$$x = \sqrt{\frac{B^2}{4} + y} - \frac{B}{2}$$

On notera que les opérations arithmétiques sont effectuées dans $\mathbf{Z}_n$, et que la division par 2 ou 4 est en fait la multiplication par $2^{-1}$ et $4^{-1}$ modulo **n**, respectivement.

Cette fonction de chiffrement n'est pas injective. Il existe en effet quatre nombres clairs qui peuvent se chiffrer en le même nombre chiffré. Le destinataire du message n'a aucun moyen pour distinguer le bon nombre parmi les quatre qu'il trouvera. On verra dans un exemple complet comment remédier à ce problème.

## Exemple

Supposons que **p** = 7, **q** = 11 **(clefs privées)**, **n** = **pq** = 77 et **B** = 9 **(clefs publiques)**.
La formule de chiffrement est:

$$y = x^2 + 9x \pmod{77}$$

et la formule de déchiffrement est:

$$x = \sqrt{1+y} - 43 \pmod{77}$$

(en effet, d'après l'algorithme d'Euclide étendu, $4^{-1} \bmod 77 = 58$ et $9^2 \cdot 58 \bmod 77 = \mathbf{1}$. De même, $2^{-1} \bmod 77 = 39$ et $9 \cdot 39 \bmod 77 = \mathbf{43}$).

### Calcul des quatre racines

Supposons que Bob (le destinataire) souhaite déchiffrer le message **y** = 22. Il faut donc trouver la racine carrée de 23 modulo 77, ce qui revient, comme le dit le théorème des restes chinois, au même que trouver la racine carrée de 23 modulo 7 et la racine carrée de 23 modulo 11. C'est ici qu'intervient le fait que **p** et **q** doivent être congrus à 3 modulo 4. En effet, dans ce cas, il existe une formule simple:

> Si **p** mod 4 = 3, alors la solution de
> $\mathbf{C}^{(1/2)} \bmod \mathbf{p}$
> est
> $\pm\mathbf{C}^{(p+1)/4} \bmod \mathbf{p}$

On va donc calculer:

$23^{(1/2)} \bmod 7 = 23^{(7+1)/4} \bmod 7 = 23^2 \bmod 7 = (23 \bmod 7)^2 \bmod 7 = 2^2 \bmod 7 = 4$
$23^{(1/2)} \bmod 11 = 23^{(11+1)/4} \bmod 11 = 23^3 \bmod 11 = (23 \bmod 11)^3 \bmod 11 = 1^3 \bmod 11 = 1$

Il ne reste qu'à résoudre le système d'équations ci-dessous avec le théorème des restes chinois:

$\mathbf{x} \bmod 7 = 4$
$\mathbf{x} \bmod 11 = 1$ (*)

On calcule successivement:

$\mathbf{M} = 77$, $\mathbf{M}_1 = 11$, $\mathbf{M}_2 = 7$,

$\mathbf{y}_1 = 11^{-1} \bmod 7 = 2$ et $\mathbf{y}_2 = 7^{-1} \bmod 11 = 8$ (valeurs trouvées avec l'algorithme d'Euclide étendu)

donc $\mathbf{x} = 4 \cdot 11 \cdot 2 + 1 \cdot 7 \cdot 8 = 144 \bmod 77 = 67$

**67** est donc la première des 4 racines carrées de 23 modulo 77. Les trois autres racines carrées se trouvent en résolvant le système (*) avec respectivement $\mathbf{a}_1 = -4$ et $\mathbf{a}_2 = 1$, $\mathbf{a}_1 = 4$ et $\mathbf{a}_2 = -1$, et enfin $\mathbf{a}_1 = -4$ et $\mathbf{a}_2 = -1$. Vous pourrez vérifier en exercice qu'elles valent respectivement **45**, **32** et **10**.

### Calcul des quatre nombres clairs

Finalement, les quatre nombres clairs possibles sont :

67 - 43 mod 77 = **24**
45 - 43 mod 77 = **2**

32 - 43 mod 77 = **66**
10 - 43 mod 77 = **44**
Vérifiez en exercice que ces quatre nombres se chiffrent en 22.
Vous trouverez dans la suite un exemple complet de chiffrement et déchiffrement.

**Un exemple complet**
Pour voir comment le chiffre de Rabin fonctionne, nous allons utiliser l'alphabet ordinaire + le symbole "*" que nous chiffrerons ainsi:

| Alphabet | 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 | * |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Codes | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |

Rappelons que 4 textes clairs correspondront au texte chiffré que recevra le destinataire. Afin qu'il puisse facilement retrouver le bon, le symbole * débutera chaque bloc de texte.

Nous utiliserons des blocs de quatre caractères et pour simplifier, nous avons choisi **B** = 0. Avec ces choix, le plus grand nombre possible est *ZZZ=26252525. Nous devons choisir **n** plus grand que cela, et, de plus, **n** doit être le produit de deux nombres premiers congrus à 3 modulo 4. Soient **p** = 6911 et **q** = 6947 (on peut vérifier que **p** et **q** sont deux nombres premiers de la forme 4 **k** + 3). Ces deux valeurs sont la clef privée et ne doivent pas être rendues publiques. On calcule ensuite que **n** = 6911·6947=48010717.

Les valeurs **n** = **48010717** et **B** = **0** constitue la **clef publique** que l'envoyeur utilisera pour chiffrer son message.

**Chiffrement**
La fonction de chiffrement est:

$$y = x^2 \mod n$$

Alice désire envoyer un message secret à Bob. Ce dernier lui a fait connaître ses clefs publiques et elle va les utiliser pour chiffrer le texte suivant:
JE T'AIME
Elle le décompose d'abord en blocs de trois caractères chacun,
JET AIM EKK
puis elle marque chacun des blocs en ajoutant le symbole * au début de chaque bloc:
*JET *AIM *EKK
puis les caractères sont convertis en leurs équivalents numériques (les zéros sont importants):
26090419 26000812 26041010
Elle chiffre d'abord le premier bloc: $C_1 = 26090419^2 \mod 48010717 = 46850914$
Le deuxième bloc est chiffré comme suit: $C_2 = 26000812^2 \mod 48010717 = 5842871$
Et le troisième ainsi: $C_3 = 26041010^2 \mod 48010717 = 12031786$
Le message chiffré transmis par Alice est la suite de nombres:
46850914 5842871 12031786

**Déchiffrement**
La formule de déchiffrement est:

$$x = \sqrt{y} \mod n$$

Bob, qui n'a communiqué à personne sa **clef privée**, à savoir les deux nombres **p** = **6911** et **q** = **6947**, sera le seul à pouvoir déchiffrer le message (bien sûr, dans notre exemple, **n** = 48010717 est facilement factorisable en n=6911·6947; en réalité, nous utiliserions des tailles de blocs beaucoup plus grandes et des nombres premiers beaucoup plus grands).
Pour déchiffrer le premier bloc, Bob doit résoudre la congruence

$$P_1 = 46850914^{(1/2)} \pmod{48010717}$$

D'après le théorème des restes chinois, résoudre l'équation ci-dessus revient à résoudre le système d'équations:

$$\begin{aligned} \mathbf{x} &= 46850914^{(1/2)} \pmod{6911} \\ \mathbf{x} &= 46850914^{(1/2)} \pmod{6947} \end{aligned} \qquad (*)$$

Il va donc calculer tout d'abord les racines carrées (par exemple avec la fonction *Mathematica* **PowerMod**):

$$46850914^{(1/2)} \bmod 6911 = 46850914^{((6911+1)/4)} \bmod 6911 = 1394$$
$$46850914^{(1/2)} \bmod 6947 = 46850914^{((6947+1)/4)} \bmod 6947 = 2513$$

Il ne lui reste plus qu'à résoudre, avec le théorème des restes chinois:

$$\begin{aligned} \mathbf{x} &\bmod 6911 = 1394 \\ \mathbf{x} &\bmod 6947 = 2513 \end{aligned} \qquad (*)$$

Pour cela, il calcule successivement:

$\mathbf{M} = 48010717$, $\mathbf{M}_1 = 6947$, $\mathbf{M}_2 = 6911$,

$\mathbf{y}_1 = 6947^{-1} \bmod 6911 = 192$ et

$\mathbf{y}_2 = 6911^{-1} \bmod 6947 = 6754$ (valeurs trouvées avec l'algorithme d'Euclide étendu).

Il obtient donc les quatre nombres clairs:

$\mathbf{x}_1 = 1394 \cdot 6947 \cdot 192 + 2513 \cdot 6911 \cdot 6754 \bmod 48010717 = 119158385278 \bmod 48010717 = 43796401$

$\mathbf{x}_2 = -1394 \cdot 6947 \cdot 192 + 2513 \cdot 6911 \cdot 6754 \bmod 48010717 = 115439683966 \bmod 48010717 = 21920298$

$\mathbf{x}_3 = 1394 \cdot 6947 \cdot 192 - 2513 \cdot 6911 \cdot 6754 \bmod 48010717 = -115439683966 \bmod 48010717 =$ **26090419**

$\mathbf{x}_4 = -1394 \cdot 6947 \cdot 192 - 2513 \cdot 6911 \cdot 6754 \bmod 48010717 = -119158385278 \bmod 48010717 = 4214316$

La bonne valeur est celle qui commence par 26, donc 26090419. Une fois décodé, ce nombre donne les lettres *JET.

En procédant de la même façon avec les deux derniers blocs, il retrouvera le message complet.

**Théorème des restes chinois**

Si $m_1, ..., m_n$ sont des entiers supérieurs à 2 deux-à-deux premiers entre eux, alors, pour des entiers $a_1, ..., a_n$, le système d'équations:

$$x \bmod m_1 = a_1$$
$$...$$
$$x \bmod m_n = a_n$$

admet une unique solution modulo $M = m_1 \cdot ... \cdot m_n$ donnée par la formule:

$$x = (a_1 M_1 y_1 + ... + a_n M_n y_n) \bmod M$$

où $M_i = M/m_i$ et $y_i = M_i^{-1} \bmod m_i$ pour $1 \leq i \leq n$.

**Exemple**

Pour n = 3, supposons que $m_1 = 7$, $m_2 = 11$ et $m_3 = 13$. On a donc M = 1001.

On calcule $M_1 = 143$, $M_2 = 91$, $M_3 = 77$.

D'après l'algorithme d'Euclide étendu, on a $y_1 = 5$, $y_2 = 4$, $y_3 = 12$.

Par exemple, pour le système d'équations:

x mod 7 = 5
x mod 11 = 3
x mod 13 = 10

la formule donne pour solution:

x = (5·143·5 + 3·91·4 + 10·77·12) mod 1001
= 13'907 mod 1001
= 894

On peut facilement vérifier cette solution en introduisant x dans le système d'équations.

## 4.3. Cryptographie à courbe elliptique

Les spécialistes en cryptographie font de plus en plus appel à une technique qui s'appuie sur des courbes elliptiques, proposée indépendamment par Victor Miller et Neal Koblitz en 1985. La théorie des courbes elliptiques en général est un domaine riche et a donné de nombreux résultats, dont la preuve du dernier théorème de Fermat par Andrew Wiles. Une courbe elliptique est un objet très simple mais qui a des propriétés tout à fait surprenantes. Elles ont normalement la forme suivante:

$$y^2+a_1xy+a_3y=x^3+a_2x^2+a_4x+a_6$$

Pour leur usage en cryptographie, $a_1$, $a_2$ et $a_3$ doivent être égaux à 0. Comme les cryptographes ont l'habitude de renommer $a_4$=a et $a_6$=b, on obtient :

$$y^2=x^3+ax+b$$

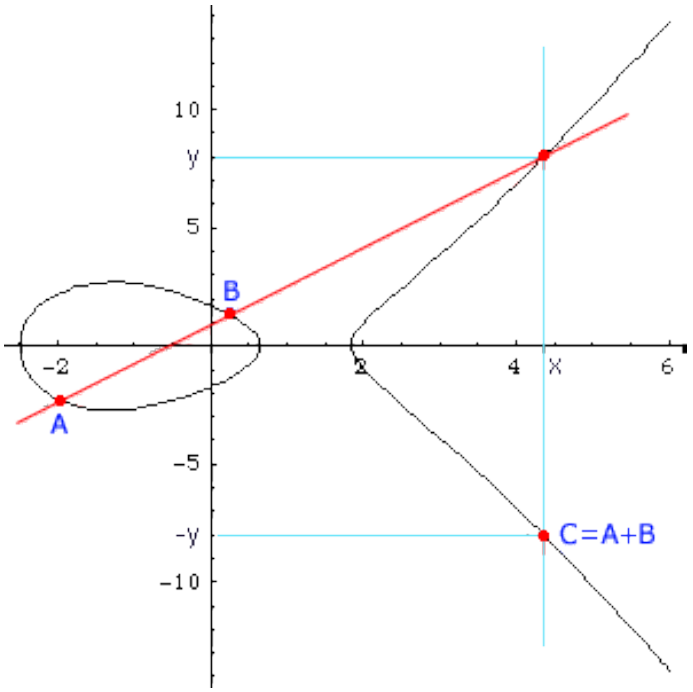
dont le discriminant

$$-(4a^3 + 27b^2)$$

est non nul. Pour la dessiner, pour a et b fixés, on calcule y tel que

$$y = (x^3 + ax + b)^{1/2}$$

Un exemple typique de courbe elliptique est donné sur la figure ci-dessous. Son équation est $y^2=x^3-5x^2+3$.




**Addition de deux points**

Prenons deux points A et B sur cette courbe. En général, la courbe passant par A et B recoupe la courbe en un troisième point de coordonnées (x, y). Son symétrique (x, -y) est lui aussi sur la courbe et on le désigne par A+B pour signifier qu'il est construit à l'aide de A et B. **La chose surprenante est que cette opération "+" possède toutes les propriétés de l'addition des nombres.** C'est-à-dire que l'on peut faire tous les calculs de type addition, soustraction et division avec un reste entier que nous faisons sur la droite des nombres réels sur cet objet tordu que constitue une courbe elliptique.

**Il est possible, comme l'ont démontré Miller et Koblitz, de coder avec cette opération bizarre au lieu de travailler avec l'addition usuelle.** Il en résulte une plus grande complexité des calculs qui fait dire aux spécialistes que le chiffrement par la méthode des courbes elliptiques avec **une clef de 192 bits assure le même niveau de sécurité qu'une clef de 1024 bits pour la méthode RSA.** Il est donc probable que cette méthode sera de plus en plus utilisée pour transmettre les clefs.

**Règles de l'addition**

Soit la courbe elliptique cryptographique $y^2 \bmod p = (x^3 + ax + b) \bmod p$. On reconnaît l'équation déjà vue ci-dessus, sauf que l'on travaille modulo p. Le nombre p doit être un nombre premier. Lors des calculs, il arrive parfois que l'on doit faire une division par 0. Quand cela arrive, le point résultat sera appelé "infini".

1. infini + infini = infini
2. $(x_1,y_1)$ + infini = $(x_1,y_1)$
3. $(x_1,y_1) + (x_1,-y_1)$ = infini
4. Si $x_1 \neq x_2$, $(x_1,y_1) + (x_2,y_2) = (x_3,y_3)$, avec

   $x_3 = (k^2-x_1-x_2) \bmod p$

   $y_3 = (k(x_1-x_3)-y_1) \bmod p$

   où $k = (y_2-y_1) \cdot (x_2-x_1)^{-1} \bmod p$
5. Si $y_1 \neq 0$, $(x_1,y_1) + (x_1,y_1) = 2(x_1,y_1) = (x_3,y_3)$, avec

   $x_3 = (k^2-2x_1) \bmod p$

   $y_3 = (k(x_1-x_3)-y_1) \bmod p$

   où $k = (3x_1^2+a) \cdot (2y_1)^{-1} \bmod p$

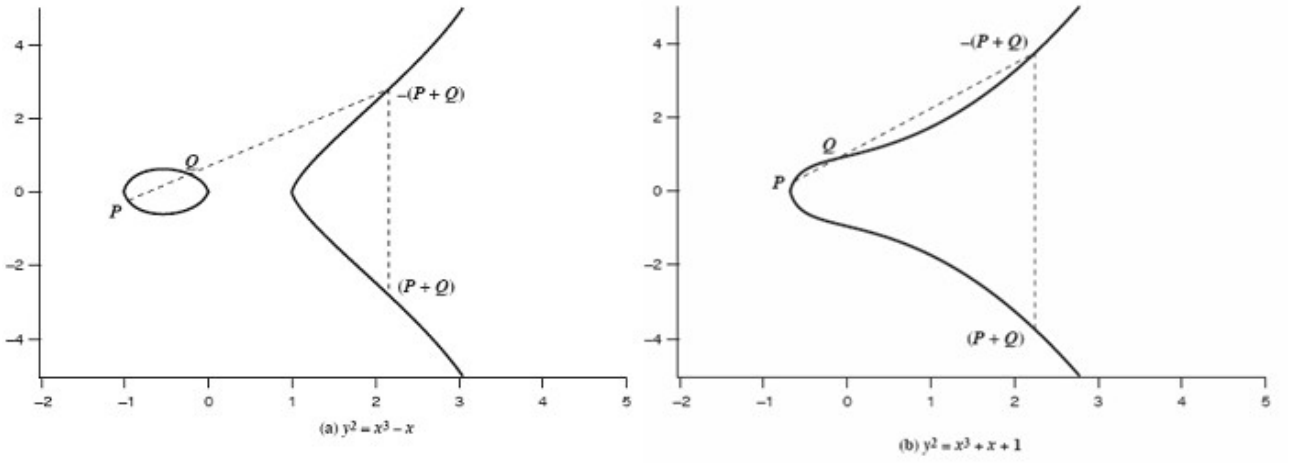
On remarque que pour calculer k, on doit trouver l'inverse d'un nombre modulo p. Pour trouver cet inverse, on utilise l'algorithme d'Euclide étendu.

**Multiplication d'un point par un nombre entier**

On remplace la multiplication par une série d'additions. Prenons un exemple. Soit la courbe $y^2$ mod 11 = ($x^3$ + x + 2) mod 11.
Calculons d·P, avec d=3 et P=(4, 2). On peut vérifier que le point P appartient bien à la courbe elliptique.
On peut remplacer 3·P par P + P + P. Calculons d'abord P + P.
D'après la règle 5, P + P = (4, 2) + (4, 2) = (8, 4) = 2·P.
D'après la règle 4, 2·P + P = (8, 4) + (4, 2) = (2, 10) = 3·P.

**Exercice :** vérifiez les calculs de 2·P et 3·P! Ces points appartiennent-ils à la courbe elliptique donnée?

Voici un autre exemple.




**Remarque :** Contrairement à ce que l'on peut croire à première vue, il ne s'agit pas de travailler à partir d'ellipses. La raison en est que les équations utilisées (équations cubiques) sont similaires à celles permettant de déterminer la circonférence d'une ellipse.

**Définition géométrique**

Soit une opération (l'addition, +) pour l'ensemble E(a, b) tel que a et b répondent à la condition du discriminant. Si 3 points sur une EC sont alignés, leur somme vaut O (point à l'infini).

1. O est l'identité pour l'addition : O = −O.
2. Pour n'importe quel point P + O = P .
3. L'opposé d'un point P (x, y) est P (x, −y)
4. Pour additionner 2 points P et Q, on trace la droite les reliant. Cela nous donne un point d'intersection R. On définit l'addition telle que P + Q = −R. En conséquence, on définit P + Q comme étant l'opposé de ce point R.

**Elliptic Curve Cryptography (ECC) sur $Z_p$**

Les variables et coefficients prennent des valeurs dans l'ensemble [0, p − 1] pour un certain nombre premier p, et où toutes les opérations sont calculées modulo p. L'équation devient
$y^2$ mod p = ($x^3$ + ax + b) mod p

Cette équation est par exemple satisfaite pour a = 1, b = 1, x = 9, y = 7 et p = 23.

$$7^2 \bmod 23 = (9^3 + 9 + 1) \bmod 23$$
$$49 \bmod 23 = 739 \bmod 23$$
$$3=3$$

On note $E_p$ (a, b) l'ensemble des couples d'entiers (x, y) qui satisfont cette équation. On parle de groupe elliptique.

**Exemple :** Soient p = 23 et la courbe elliptique $y^2 = x^3 + x + 1$. On est donc dans $E_{23}$ (1, 1). Comme nous travaillons dans $Z_p$ , les couples (x, y) répondant à l'équation sont donnés dans le suivant :

| | | |
|---|---|---|
| (0, 1) | (6, 4) | (12, 19) |
| (0, 22) | (6, 19) | (13, 7) |
| (1, 7) | (7, 11) | (13, 16) |
| (1, 16) | (7, 12) | (17, 3) |
| (3, 10) | (9, 7) | (17, 20) |
| (3, 13) | (9, 16) | (18, 3) |
| (4, 0) | (11, 3) | (18, 20) |
| (5, 4) | (11, 20) | (19, 5) |
| (5, 19) | (12, 4) | (19, 18) |

Pour tous points P, Q ∈ E p (a, b), on a :

1. **P + O = P**

2. **Si $P = (x_P , y_P )$, alors $P + (x_P , -y_P ) = 0$ et $(x_P , -y_P ) = -P$ .**
   Retour à l'exemple : Dans $E_{23}$ (1, 1), pour P = (13, 7), −P = (13, −7) = (13, 16)

3. **Si $P = (x_P , y_P)$ et $Q = (x_Q , y_Q)$ avec $P <> -Q$, alors on détermine $R = P + Q = (x_R , y_R)$ comme suit :**
   $$x_R = (\lambda^2 - x_P - x_Q) \bmod p$$
   $$y_R = (\lambda(x_P - x_R ) - y_P) \bmod p$$
   où
   $$\lambda = (y_Q - y_P)/(x_Q - x_P) \bmod p \text{ si } P <> Q$$
   $$\lambda = (3x^2_P + a)/2y_P \bmod p \text{ si } P <> Q$$

4. **La multiplication est définie comme une répétition d'additions (ex : 3P = P + P + P )**
   Retour à l'exemple : Soient P = (3, 10) et Q = (9, 7) dans E 23 (1, 1). Il vient
   $$\lambda = (7-10)/(9-3) \bmod 23 = (-3)/6 \bmod 23 = (-1)/2 \bmod 23 = 11$$
   $$x_R = 11^2 - 3 - 9 \bmod 23 = 109 \bmod 23 = 17$$
   $$y_R = (11(3 - 17) - 10) \bmod 23 = -164 \bmod 23 = 20$$

   Et ainsi, P + Q = (17, 20). Pour trouver 2P , on a
   $$(3(3^2) + 1)/(2 * 10) \bmod 23 = 5/20 = \bmod 23 = 1/4 \bmod 23$$
   $$x_R = 6^2 - 3 - 3 \bmod 23 = 30 \bmod 23 = 17$$
   $$y_R = (6(3 - 7) - 10) \bmod 23 = -34 \bmod 23 = 12$$
   et donc 2P = (7, 12).

**Échange de clefs**

Alice et Bob se mettent d'accord (publiquement) sur une courbe elliptique E(a,b,p), c'est-à-dire qu'ils choisissent une courbe elliptique $y^2 \bmod p = (x^3+ax^2+b) \bmod p$. Ils se mettent aussi d'accord, publiquement, sur un point P situé sur la courbe.

Secrètement, Alice choisit un entier $d_A$, et Bob un entier $d_B$. Alice envoie à Bob le point $d_AP$, et Bob envoie à Alice $d_BP$. Chacun de leur côté, ils sont capables de calculer $d_A(d_BP)=d_B(d_AP)=(d_Ad_B)P$, qui est un point de la courbe, et constitue leur clef secrète commune.

**Sécurité**

Si Eve a espionné leurs échanges, elle connaît E(a,b,p), P, $d_AP$ et $d_BP$. Pour pouvoir calculer $d_Ad_BP$, il faut pouvoir calculer $d_A$ connaissant P et $d_AP$. C'est ce que l'on appelle résoudre **le logarithme discret sur une courbe elliptique**. Or, actuellement, **si les nombres sont suffisamment grands, on ne connaît pas de méthode efficace pour résoudre ce problème en un temps raisonnable.**

**Inconvénients**

- La théorie des fonctions elliptiques est complexe, et encore relativement récente. Il n'est pas exclu que des trappes permettent de contourner le problème du logarithme discret.
- La technologie de cryptographie par courbe elliptique a fait l'objet du dépôt de nombreux brevets à travers le monde. Cela peut rendre son utilisation très coûteuse!