Initiation à la cryptographie : théorie et pratique

Houda FERRADI

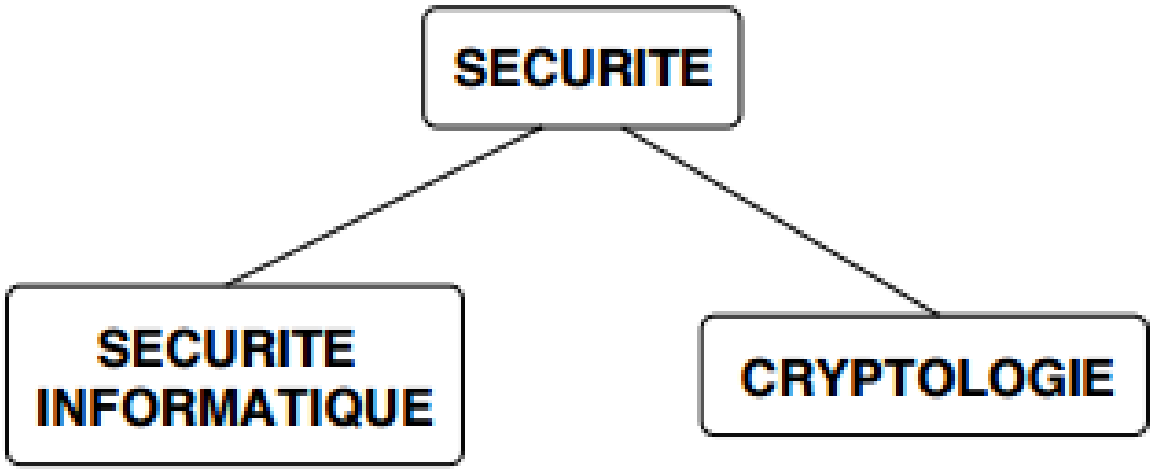
Université Paris 13 Villetaneuse

7 janvier 2016

Chapitre 1 : But du cours

- Comprendre les problématiques de cryptographie liées aux systèmes d'informations.
- Comprendre les principaux systèmes cryptographiques modernes utilisés pour la transmission et le stockage sécurisé de données.
- Comment analyser la sécurité des systèmes cryptographiques.
- Introduction aux infrastructures pour les systèmes à clé publique et clé secrète.

Notion de sécurité



La terminologie de la cryptographie

Cruptos : Caché.

Graphein : Ecrire

cryptologie==science du secret

- La **cryptologie** : *Mécanisme permettant de camoufler des messages i.e., de le rendre incompréhensible pour quiconque n'est pas autorisé*. Elle fait partie d'une ensemble de théories et de techniques liées à la transmission de l'information (théorie des ondes électromagnétiques, théorie du signal, théorie des codes correcteurs d'erreurs, théorie de l'information, théorie de la complexité,...).
  Cryptologie = Cryptographie + Cryptanalyse
- La **cryptographie** : est l'art de rendre inintelligible, de crypter, de coder, un message pour ceux qui ne sont pas habilités à en prendre connaissance.
- La **cryptanalyse** : art de "casser" des cryptosystèmes.
- **Protocole** : description de l'ensemble des données nécessaires pour mettre en place le mécanisme de cryptographie : ensemble des messages clairs, des messages cryptés, des clés possibles, des transformations.
- **Signature** *s* : Chaine de caractères associées à un message donné (et aussi possiblement à une entité) et le caractérisant.

Remarque

Cryptologie $\subset$ Sécurité

La terminologie de la cryptographie

- Le **chiffrement**, noté E_k, est l'action de chiffrer un **message en clair**, noté M, en un **message chiffré**, noté C, et cela de façon à ce qu'il soit impossible de retrouver le message en clair à partir du message chiffré sans la clé ;
- cette clé s'appelle la **clé secrète** et elle est unique ;
- l'action inverse du chiffrement est le **déchiffrement**. Cette action s'effectue uniquement en possession de la clé secrète ;
- le déchiffrement d'un message chiffré sans la clé secrète s'appelle le **décryptage** ;
- un système de chiffrement s'appelle un **cryptosystème** ;
- un **cryptographe** est une personne qui conçoit des cryptosystèmes ;
- un **cryptanaliste** est une personne qui tente de casser les cryptosystèmes.

Crypto : Bref Histoire de codes secrets

Les 3 ages de la crypto : jusqu'au 20e siecle (fin de la 1ere guerre mondiale)
L'age artisanal :

1. **Hiéroglyphes :** Pour étaler son savoir, pour cacher l'emplacement des trésors.
2. **Ancien Testament :** Code atbash : a=z (taw), b=y (shin)> k=l (lamed) et ainsi de suite, inversant l'alphabet...
3. Transposition Sparte ou scytale (5ème siècle av JC)
4. Substitution César (1er si'ecle av JC), Vigen'ere (XVI 'eme).
   Cryptanalyse des codes mono et poly alphabétiques :
5. El Kindi (IX ème siècle)
6. Babbage/Kasiski (XIX'eme siècle).
   Mécanisation de la cryptographie et de la cryptanalyse :
7. Enigma (1918)
8. Vers un chiffrement parfait : Vernam, théorie de l'information
9. **Seconde guerre mondiale :** Plusieurs livres marqués à l'encre invisible, déplacements des navires japonais dans le Pacifique.

Crypto : Historique

- **L'age technique : Le 20e siecle (1919-1975) :** Cryptosystèmes symétriques
  - I.e. conventionnels ou à clé secrète
  - Même clé pour chiffrement / déchiffrement ; doit être secrète
  - N personnes N (N-1)/2 clés
  - Clés courtes ; plus rapide
- **L'age scientifique à partir de 1976** : W. Diffie et M. Hellman introduisent crypto à clé publique (ou asymétrique) :
  - Clé de chiffrement est publique (i.e., connue de tous)
  - Seul clé déchiffrement reste secrète
  - N personnes 2 N clés
- **1978** : Premier système chiffrement à clé publique (RSA) : RSA est un des seuls à résister à la cryptanalyse.

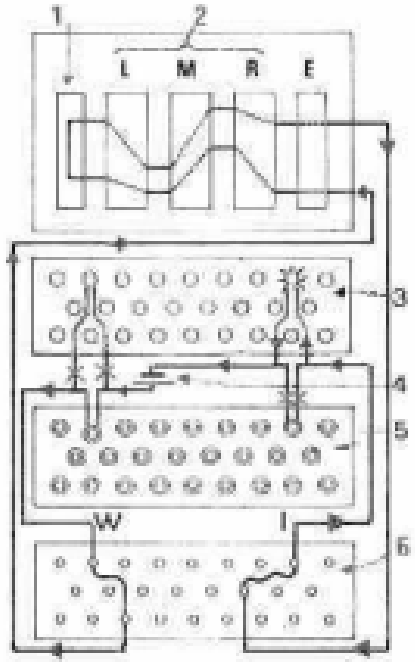
L'age artisanal

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. :

Après avoir enroulé la ceinture sur la scytale, le message était écrit en plaçant une lettre sur chaque circonvolution (axe). **Question : comment le destinataire déchiffrerait le message sur le scytale ?**

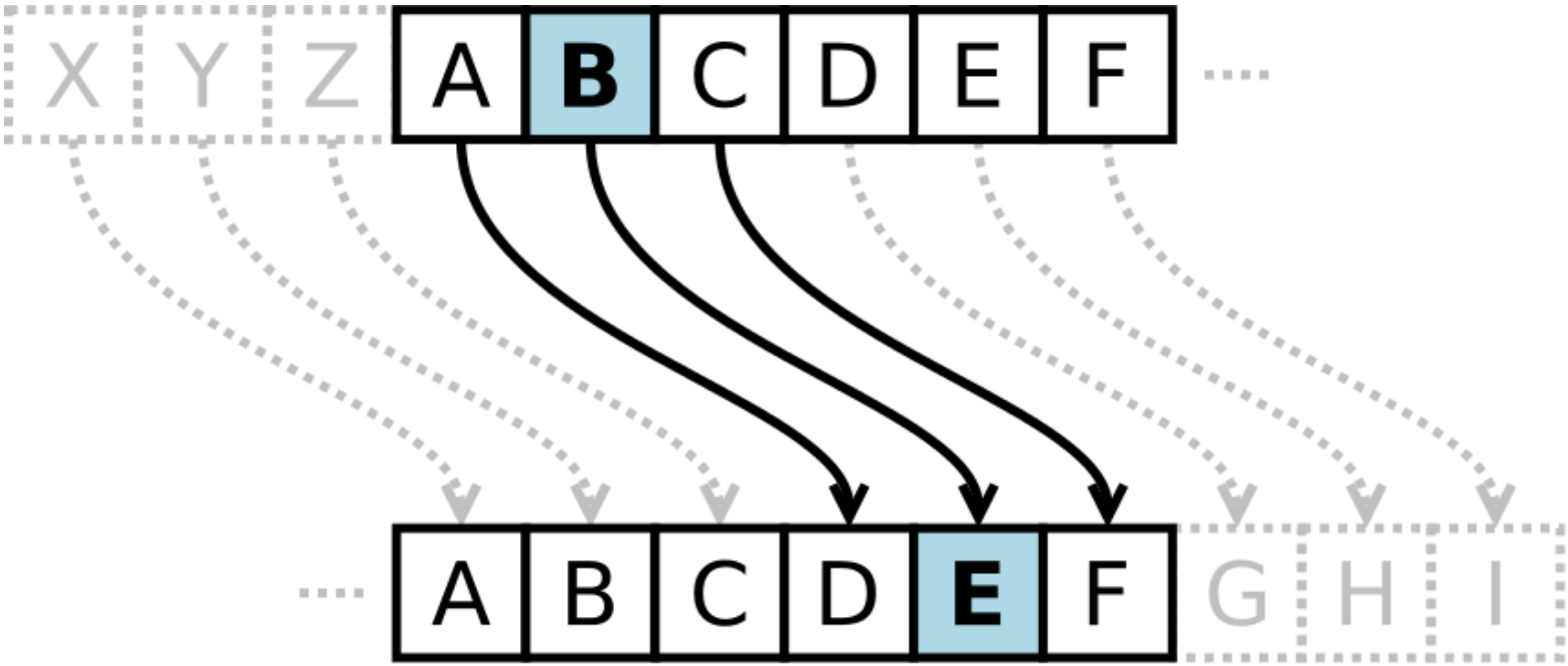
L'age technique

Alain Turing 2e guerre mondiale.
La technique utilisée : Substitutions et permutations automatiques Le crypto joue un rôle important



Le chiffrement de césar

Le chiffrement par décalage, aussi connu comme le chiffre de César ou le code de César (voir les différents noms), est une méthode de chiffrement très simple utilisée par Jules César dans ses correspondances secrètes (ce qui explique le nom « chiffre de César »).



Le chiffrement aujourd'hui

- Jusqu'au milieu des année 70, les seuls cryptosystèmes connus étaient symétriques (on dit aussi conventionnels ou à clé secrète) : la **clé de chiffrement** K_C était la même que **la clé de déchiffrement** K_D (ou du moins, la connaissance de la **clé de chiffrement** permettait d'en déduire la clé de déchiffrement) ce qui obligeait à garder secrète la clé K_C elle aussi.
- En 1976, W. Diffie et M. Hellman introduisirent le concept de **cryptographie à clé publique** (ou asymétrique). Dans ce type de système, la **clé de chiffrement** est publique, c'est à dire connue de tous. Seule la clé de déchiffrement reste secrète.
- En 1978, le premier système de chiffrement à clé publique fut introduit par R. Rivest, A. Shamir et L. Adleman : le système RSA. Ce système est un des seuls qui aient résisté à la cryptanalyse.

Théorie de l'information

la théorie de l'information est due à Claude E. Shannon et exposee dans un article paru apres-guerre : *A Mathematical Theory of Communications, 1948* :

- C'est une theorie mathématique basée sur les probabilites et visant à decrire les systéemes de communication ‘
- elle a subi l'influence d'autres theoriciens de l'informatique : A. Turing, J. von Neumann, N. Wiener et presente des convergences avec les travaux de R.A. Fisher
- le probleme est celui de la communication entre une source et un recepteur : la source emet un message que le recepteur lit : on voudrait quantifier l'information que contient chaque message emis. ´ *il est clair que si l'émetteur dit toujours la meme chose, la quantité d'information apportée par une répétition supplémentaire est nulle*
- l'entropie existe en version combinatoire, en probabilites discrétes ou encore en probabilites continues .
- A noter que la quantite d'information n'est pas une propri été intrinseque d'un certain objet, mais une proprieté de cet objet en relation avec un ensemble de possibilites auquel il appartient. Dans le cas contraire, on a recours a la ‘ complexite de Kolmogorov ´ .

Les applications réelles de la cryptologie

- **Communications sécurisée :**
  - Web : SSL/TLS, ssh, gpg
  - Sans fil : GSM, Wifi, Bluetooth
- Chiffrement des fichiers : EFS, TrueCrypt
- Protection de données personnelles : cartes de crédit, cartes Navigo, passeports électroniques et bien plus encore !

La cryptologie moderne

Dans le dictionnaire d'Oxford (2006) définit la cryptographie comme "*l'art d'écrire et de résoudre des codes*".

Cette définition n'est plus d'actualité car elle n'inclut pas la cryptologie moderne. Au 20e siècle la cryptologie s'intéresse à des domaines plus large que la sécurité des communications, notamment : L'authentification des messages et les signatures, la monnaie anonyme, partage de fichiers ...

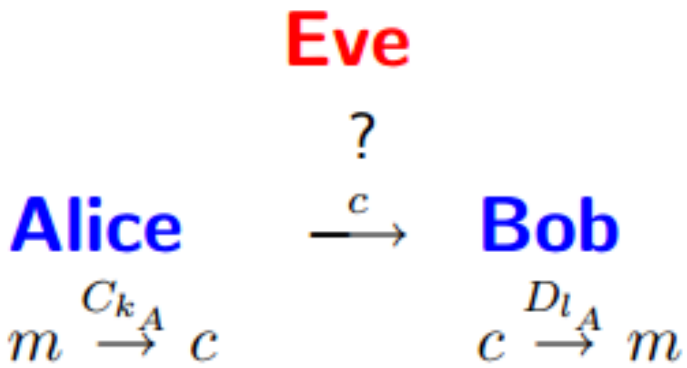
Introduction à la confidentialité des communications

- Confidentialité : garantir le secret de l'information transmise ou archivée.
- Domaine militaire : transmission de documents "secret défense", stratégies, plans
- Domaine médical : confidentialité des dossiers de patients
- Domaine commercial : pour les achats sur Internet, transmission sécurisée du numéro de carte bancaire
- Domaine industriel : transmission d'informations internes à l'entreprise à l'abris du regard des concurrents !

Confidentialité

Exemple

Un expéditeur Alice veut envoyer un message à un destinataire **Bob** en évitant les oreilles indiscrète d'**Eve** (adversaire passif), et les attaques malveillantes de **Martin**(adversaire actif).



Ou **assurer un stockage sécurisé : localement ou dans un serveur distant.**

Confidentialié

Dans un cryptosystème on distingue 5 choses :

- L'espace des messages clairs $\mathcal{M}$ sur un alphabet $\mathcal{A}$ (qui peut être l'alphabet latin, mais qui sera dans la pratique $\{0,1\}$ car tout message sera codé en binaire pour pouvoir être traité par l'ordinateur).
- L'espace des messages chiffrés $\mathcal{C}$ sur un alphabet $\mathcal{B}$ (en général égal à $\mathcal{A}$).
- L'espace des clés $\mathcal{K}$.
- Un ensemble de transformations de chiffrement (chaque transformation étant indexée par une clé) : $E_k : \mathcal{M} \to \mathcal{C}$.
- Un ensemble de transformations de déchiffrement (chaque transformation étant indexée par une clé) : $D_K : \mathcal{C} \to \mathcal{M}$

Cryptographie conventionnelle

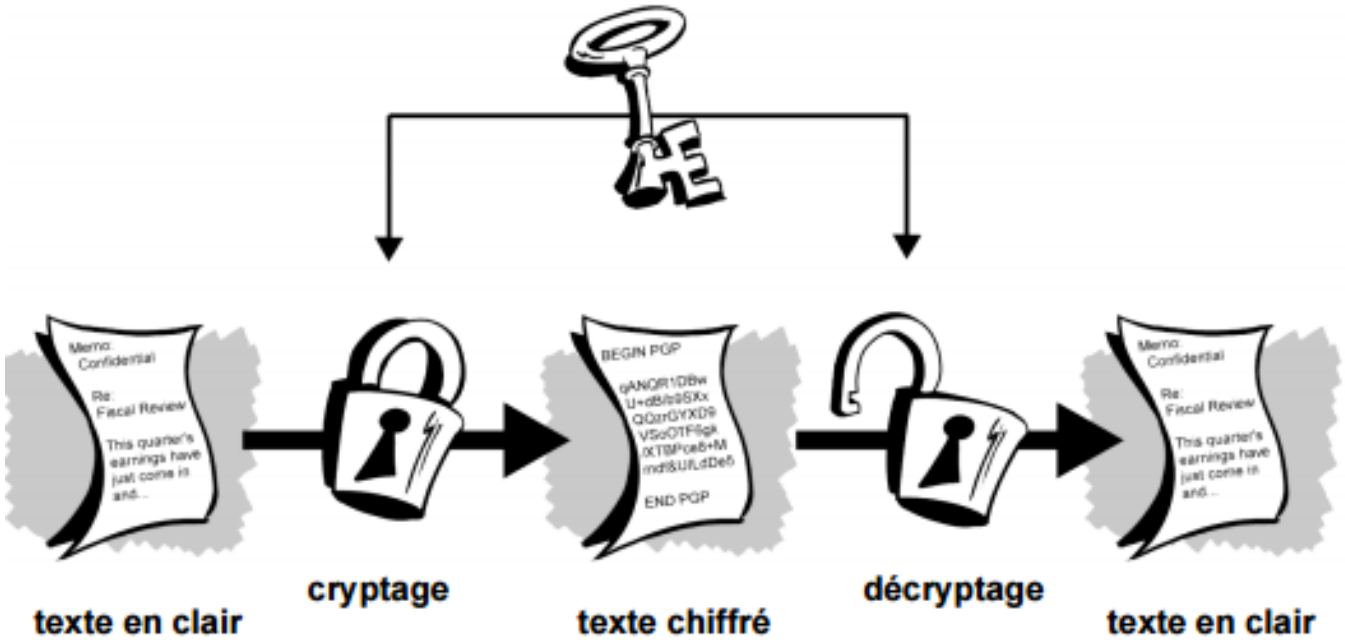


FIGURE: Cryptage et décryptage conventionnels

Cryptographie à clé publique (ou asymetrique)

Les systèmes à clé publiques ou asymétriques comme : RSA, El-Gamal, cryptosystème elliptiques, Diffie-Hellman...

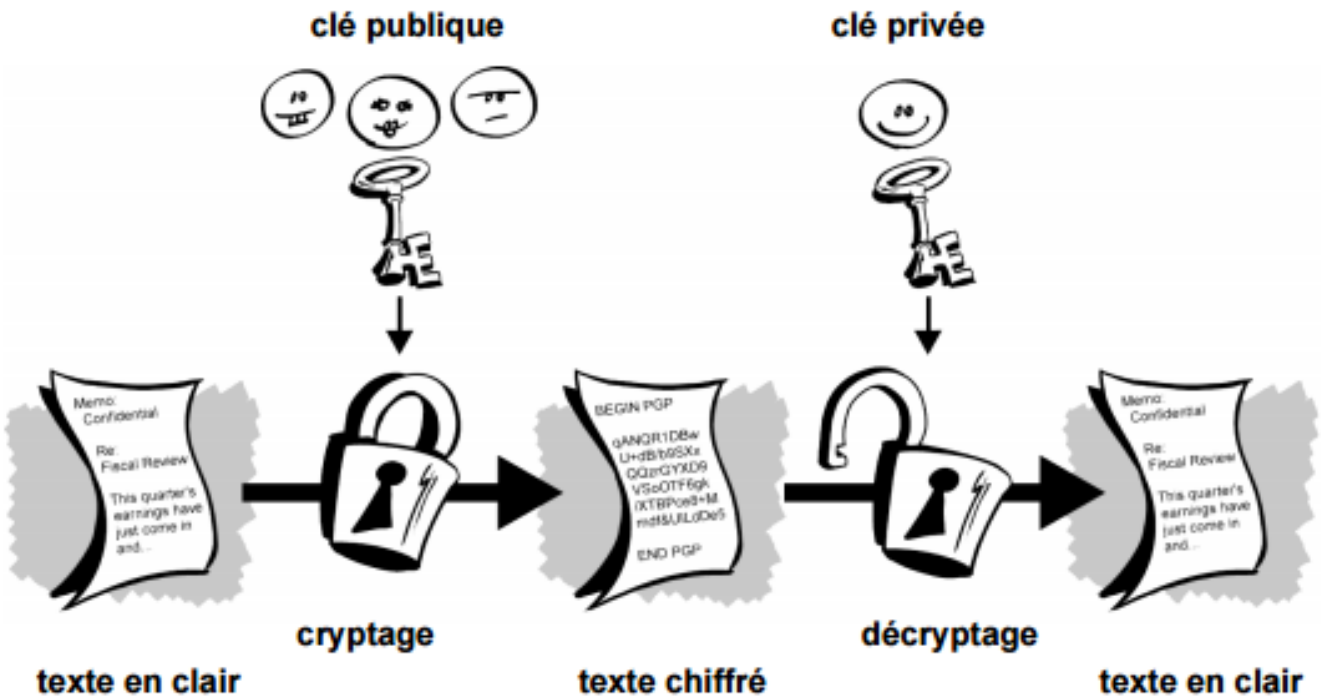


FIGURE: Cryptage et décryptage à clé publique

Cryptographie à clé secrète (ou symetrique)

Les systèmes à clé secrètes ou symétriques comme : (DES, AES, IDEA,...)

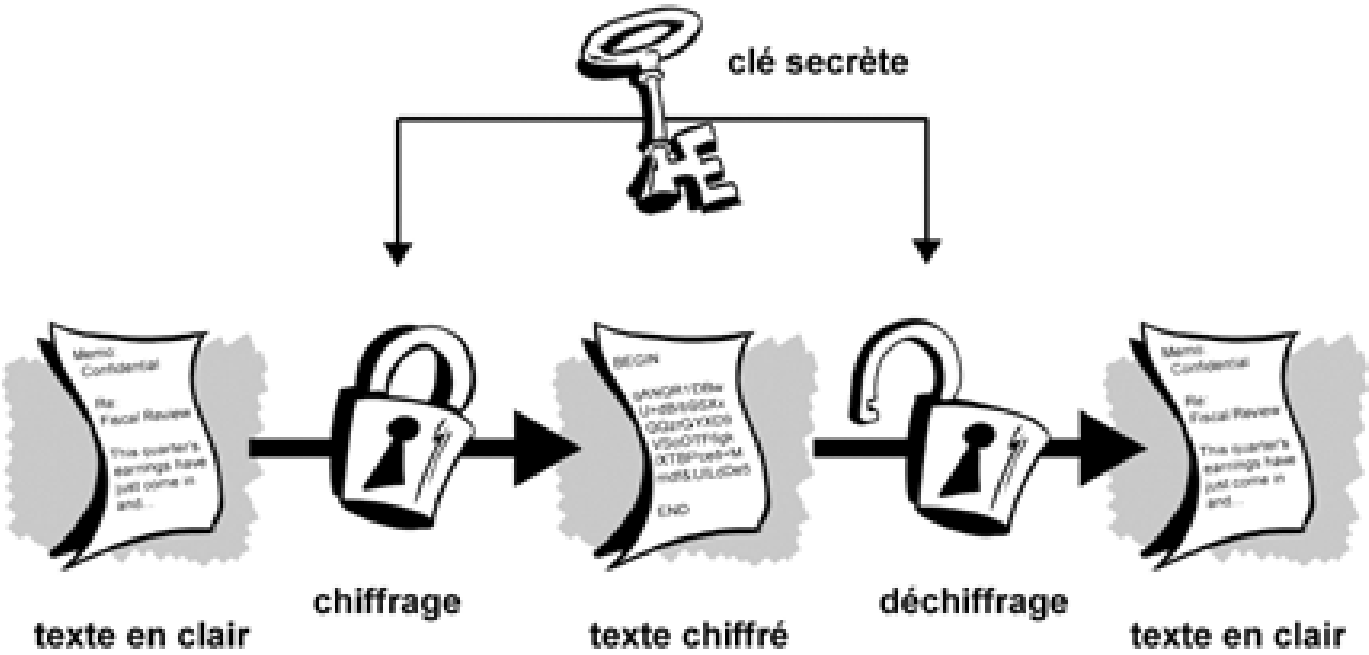


FIGURE: Cryptage et décryptage à clé secrète

Chiffrement hybride (Key wrapping)

Comment garantir la confidentialité des échanges tout en garantissant la confidentialité de la clé secrète k_s ?

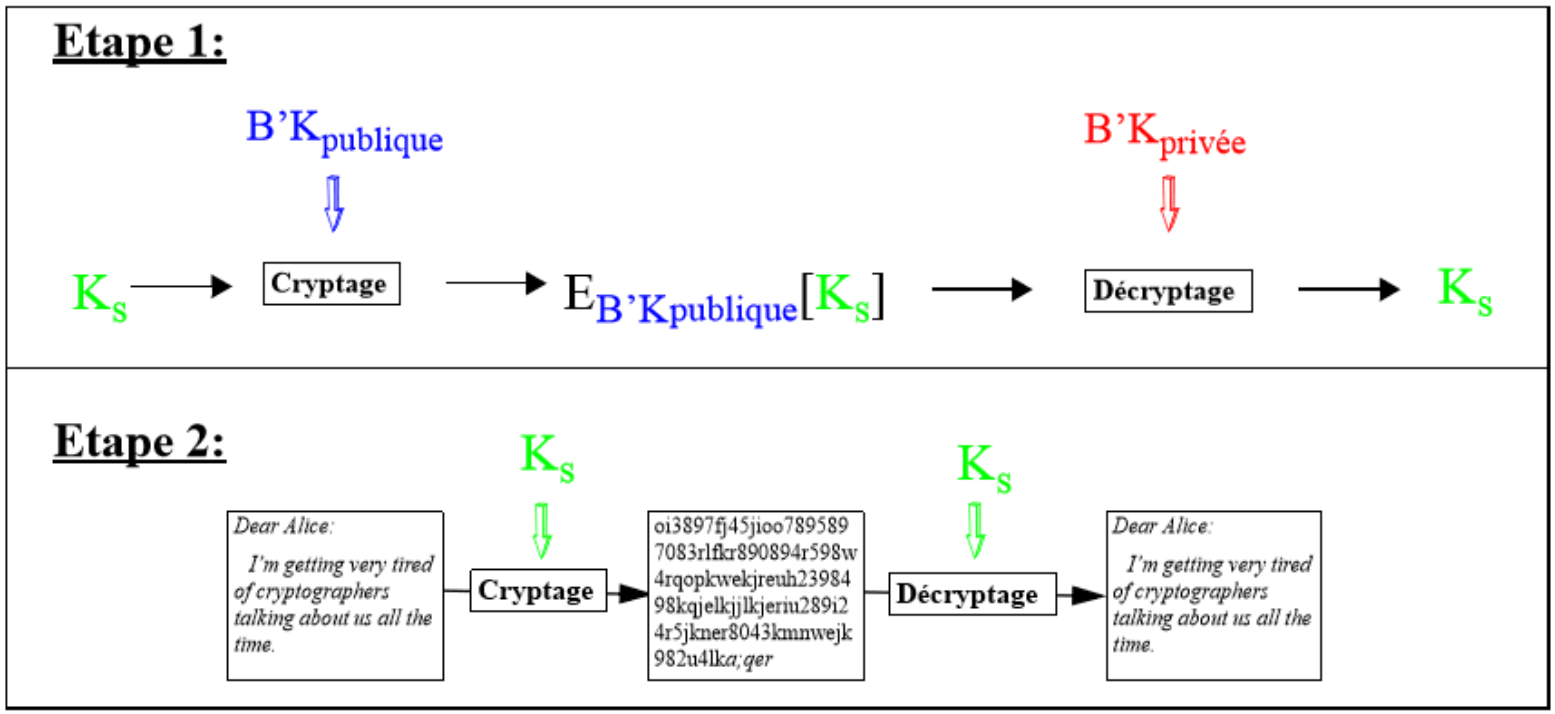


FIGURE: Crypto symétrique + Crypto asymétrique

Exemple de chiffrement hybride (Key wrapping)

Dans la pratique (voir par exemple PGP, Pretty Good Privacy), à moins qu'on ait à communiquer des messages de petite taille (de quelques k-octets), on agit comme suit : Supposons qu'Alice veuille envoyer un message secret à Bob :

- Alice choisit un cryptosystème à clé publique et utilise ce système pour envoyer secrètement à Bob un mot binaire K, qu'on appelle clé de session ; pour ce faire, elle chiffre K au moyen de la clé publique de Bob (une alternative possible est l'utilisation d'un protocole d'échange de clés tel que celui de Diffe-Hellman).
- Bob récupère K à l'aide de sa clé secrète.
- Alice et Bob détiennent maintenant un mot secret K et Alice l'utilise pour chiffrer son message long dans un cryptosystème symétrique.
- Bob peut déchiffrer le message à l'aide de K.

Principe Kerckhoffs

Le premier à avoir formalisé ce principe est le hollandais **Auguste Kerckhoffs**, qui écrit en 1883 dans le Journal des sciences militaires un article intitulé La cryptographie militaire.
On peut résumer ces conditions sous les éléments suivant :
Il faut qu'il n'exige pas le secret, et qu'il puisse sans inconvénient tomber entre les mains de l'ennemi :

1. La clé doit pouvoir en être communiquée et retenue sans le secours de notes écrites, et être changée ou modifiée au gré des correspondants ;
2. Le système doit être matériellement, sinon mathématiquement indéchiffrable ;
3. Il faut qu'il n'exige pas le secret, et qu'il puisse sans inconvénient tomber entre les mains de l'ennemi ;
4. La clé doit pouvoir en être communiquée et retenue sans le secours de notes écrites, et être changée ou modifiée au gré des correspondants ;
5. Il faut qu'il soit applicable à la correspondance télégraphique ;
6. Il faut qu'il soit portatif, et que son maniement ou son fonctionnement n'exige pas le concours de plusieurs personnes ;
7. Enfin, il est nécessaire, vu les circonstances qui en commandent l'application, que le système soit d'un usage facile, ne demandant ni tension d'esprit, ni la connaissance d'une longue série de règles à observer.

Principe de Kerckhoffs

Les points 1 et 3 sont les axiomes fondamentaux de la cryptographie moderne :

- La sécurité d'un cryptosystème ne doit pas reposer sur le secret de l'algorithme de codage mais qu'elle doit uniquement reposer sur la clé secrète du cryptosystème. Ce principe repose sur les arguments suivants :
- **La transparence :** Un cryptosystème sera d'autant plus résistant et sûr qu'il aura été conçu, choisi et implémenté avec la plus grande transparence et soumis ainsi à l'analyse de l'ensemble de la communauté cryptographique.
- **La portativité :** Si un algorithme est supposé être secret, il se trouvera toujours quelqu'un soit pour vendre l'algorithme, soit pour le percer à jour, soit pour en découvrir une faiblesse ignorée de ses concepteurs. A ce moment là c'est tout le cryptosystème qui est à changer et pas seulement la clé.

La crypto, pouruqoi faire ?

Les 4 buts de la cryptolographie afin d'assurer les services de sécurité :

- Confidentialité : Protection de la divulgation d'une information non-autorisée.
- Intégrité : Protection contre la modification non autorisée de l'information.
- Authentification d'entités : (entity authentication) procédé permettant à une entité d'être sure de l'identité d'une seconde entité à l'appui d'une évidence corroborante (ex. : présence physique, cryptographique, biométrique, etc.). Le terme identification est parfois utilisé pour désigner également ce service.

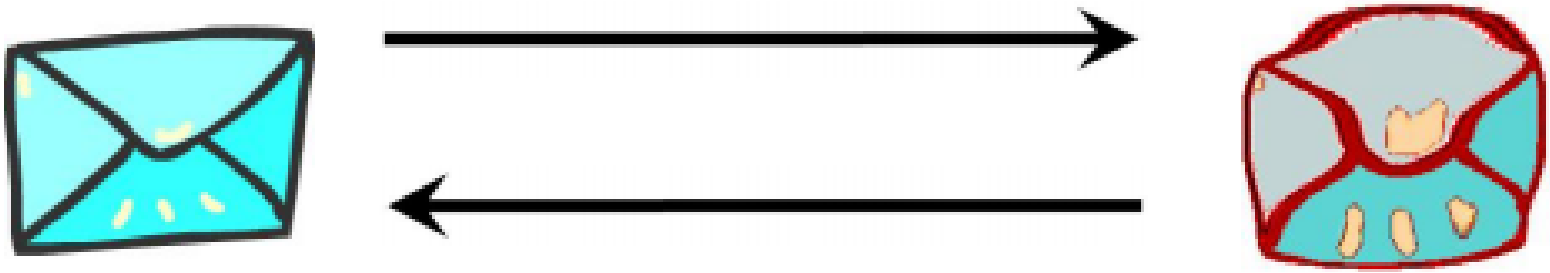
Services de sécurité

- Non-répudiation qui se décompose en trois :

1. non-répudiation d'origine l'émetteur ne peut nier avoir écrit le message et il peut prouver qu'il ne l'a pas fait si c'est effectivement le cas.
2. non-répudiation de réception le receveur ne peut nier avoir reçu le message et il peut prouver qu'il ne l'a pas reutiliser si c'est effectivement le cas.
3. non-répudiation de transmission l'émetteur du message ne peut nier avoir envoyé le message et il peut prouver qu'il ne l'a pas fait si c'est effectivement le cas.

L'integrité des messages

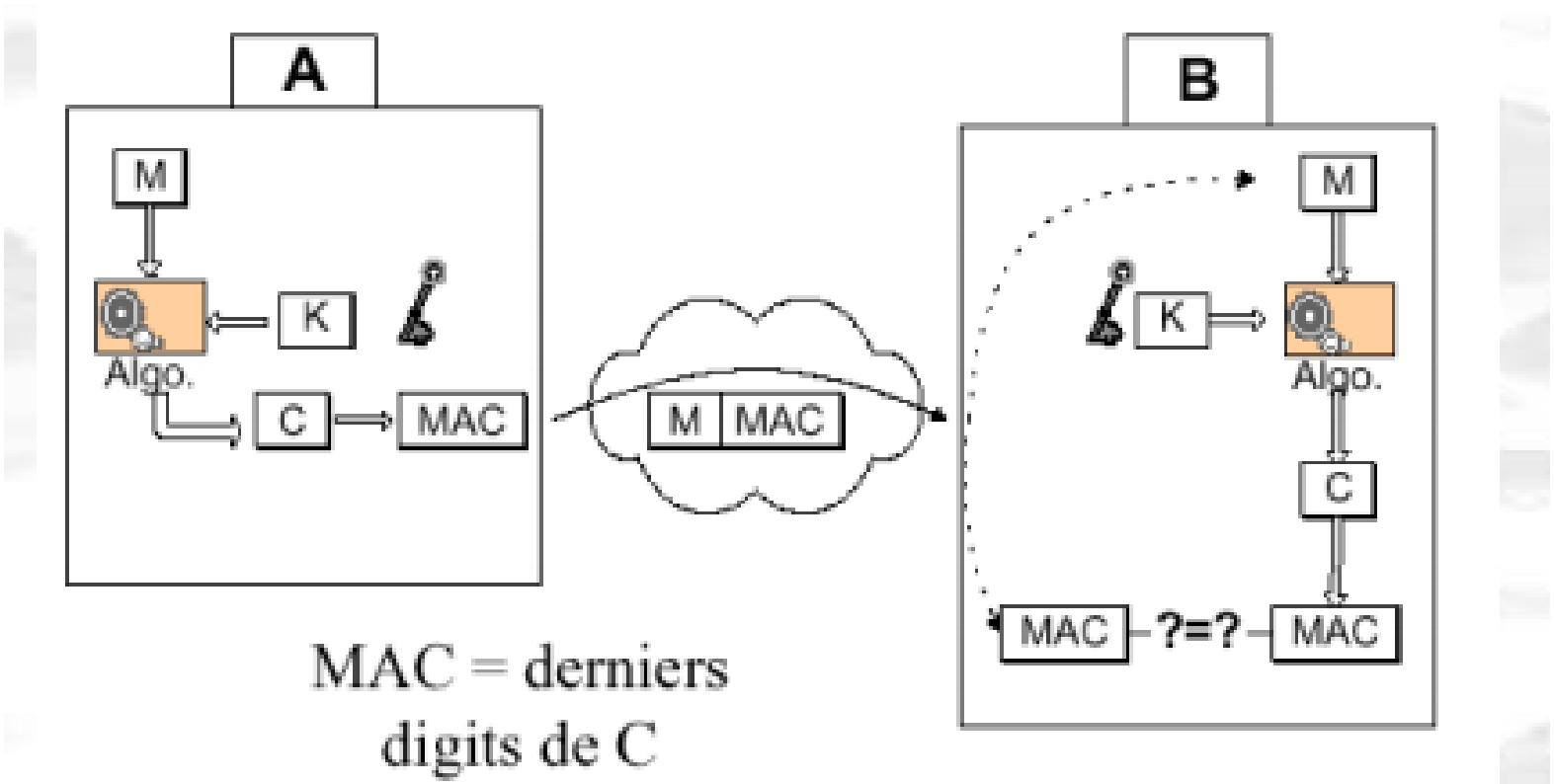
Comment s'assurer de l'intégrité des messages :



Le contenu de l'enveloppe arrive t'il "intact" ?

L'authenticité

S'assurer de la provenance des messages et de l'authenticité de l'émetteur :
Authentification grâce à K Par ex le triple DES (à l'aide d'un cryptosysteme symetrique)



La sécurité prouvée : méthode générale

Modèle de sécurité : Un cryptosystème n'est jamais absolument parfait. L'essentiel est d'obtenir le niveau de sécurité souhaité.
En général, pour prouver la sécurité d'un schéma cryptographique ou protocole, on aura besoin de 3 étapes :

1. **Préciser un modèle de sécurité formel :** (ou notions de sécurité) à garantir (voir le slide sur la cryptanalyse).
2. **Préciser les hypothèses algorithmiques :** les hypothèses calculatoires "acceptables" ou problème difficile.
3. **Présenter une réduction du problème :** Si quelqu'un arrive à casser la notion de sécurité de 1) alors quelqu'un peut casser le problème sous-jacent i.e le problème difficile) 2).

La sécurité prouvée : Notions de sécurité

Essentiellement, on dispose de deux notions :

- la sécurité sémantique : Il faut que l'adversaire (avec une puissance calculatoire limitée) soit incapable d'obtenir des informations significatives sur le message en clair à partir du texte chiffré et de la clé publique $==>$ Indistinguabilité, Non-reconnaissance de chiffré.
- la sécurité calculatoire : (voir slide sur hypothèses calculatoires).

En fonction des besoins, pour prouver la sécurité on définit :

- les objectifs de l'attaquant
- les moyens, soit les informations mises à sa disposition.

Cryptanalyse - Types d'attaques

On doit distinguer entre les **moyens d'attaques** (ou types d'attaques) d'un adversaire et les **buts d'attaques** d'un adversaire. L'attaquant connaît tout les details de l'algorithme de chiffrement/déchiffrement et qu'il ne lui manque que la clef spécifique pour le chiffrement.(axiome fondamental de Kerckhoffs). Les 4 types d'attaques :

- **Attaque à texte crypté uniquement :** L'attaquant ne dispose que d'un ou plusieurs messages chiffrés qu'il souhaite déchiffré. C'est le type d'attaque le plus difficile.
- **Attaque à texte chiffré connu :** Le cryptanalyste a non seulement accès aux textes chiffrés de plusieurs messages, mais aussi aux textes clairs correspondants. La tâche est de retrouver la ou les clés qui ont été utilisées pour chiffrer ces messages ou un algorithme qui permet de déchiffrer d'autres messages chiffrés avec ces mêmes clés.
- **Attaque à texte clair choisi :** ($IND - CPA$) L'opposant a accès à une machine chiffrante : Le cryptanalyste a non seulement accès aux textes chiffrés et aux textes clairs correspondants, mais de plus il peut choisir les textes en clair. Cette attaque est plus efficace que l'attaque à texte clair connu, car le cryptanalyste peut choisir des textes en clair spécifiques qui donneront plus d'informations sur la clé.
- **Attaque à texte chiffré choisi :** ($IND - CCA$) : L'opposant a accès à une machine déchiffrable : Le cryptanalyste peut choisir différents textes chiffrés à déchiffrer. Les textes déchiffrés lui sont alors fournis. Par exemple, le cryptanalyste a un dispositif qui ne peut être désassemblé et qui fait du déchiffrement automatique. Sa tâche est de retrouver la clé.

Hypothèses calculatoires :

- La **sécurité inconditionnelle** qui ne préjuge pas de la puissance de calcul du cryptanalyste qui peut être illimitée.
- La **sécurité calculatoire** qui repose sur l'impossibilité de faire en un temps raisonnable, compte tenu de la puissance de calcul disponible, les calculs nécessaires pour décrypter un message. Cette notion dépend de l'état de la technique à un instant donné.

Exemple

Même avec des ordinateurs faisant 10^9 opérations élémentaires par seconde un calcul qui nécessite 2^{100} opérations élémentaires est hors de portée actuellement car pour l'effectuer il faut environ 4.10^{13} années !

- La **sécurité prouvée** qui réduit la sécurité du cryptosystème à un problème bien connu réputé difficile, par exemple on pourrait prouver un théorème disant qu'un système cryptographique est sûr si un entier donné n ne peut pas être factorisé.
- La **confidentialité parfaite** qualité des codes pour lesquels un couple (message clair, message chiffré) ne donne aucune information sur la clé.

Le modèle de Dolev-Yao

On suppose que l'attaquant dispose est très intelligent et dispose de beaucoup de moyens pour modifier les communications du réseau. On suppose que l'attaquant :

- peut obtenir tous les messages circulant sur le réseau ; est un utilisateur légitime du réseau ;
- peut initier une communication avec tous les membres du réseau ;
- peut envoyer un message à tous les membres du réseau en se faisant passer pour un autre personne. Cependant, l'attaquant n'est pas tout puissant. On suppose, entre autres, que l'attaquant :
- ne peut pas deviner un entier choisi au hasard ;
- ne peut *deviner* la clé privée correspondant à une clé publique.

Evaluation d'un algo

Echelle de succés :

- Cassage complet : l'attaquant découvre la clé.
- Déduction globale : l'attaquant découvre des fonctions équivalentes aux fonctions de chiffrement et de décriffrement sans pour autant connaître la clé.
- Déduction locale : l'attaquant peut déchiffrer un ou plusieurs nouveaux messages chiffrés.
- déduction d'information : L'attaquant obtient de l'information sur la clé ou sur des messages chiffrés.

Critères d'évaluation :

- Temps : le nombre d'opérations de bases nécessaires
- Espace : la quantité de mémoire maximale nécessaire
- Données : le nombre de messages clairs/chiffrés nécessaires

Echelle des coûts en temps

Configurations :

- Un ordinateur de bureau avec 4 coeurs à 2,5 GHz : 2^{36} FLOPS
- Cluster du laboratoire de maths : 2^{40} FLOPS
- Supercomputer (à 133 million de dollars) : 2^{50} FLOPS

Nbre opérations		Config. A	Config. B	Config. C
Clé W.E.P. :	2^{40}	16 sec.	1 sec.	1 μsec.
Clé D.E.S. :	2^{56}	12 jrs	18 h	1 mn.
Collision MD5 :	2^{64}	$8\,{}^{1}/_{2}$ ans	194 jrs	4 h.
perm. de lettres :	2^{88}	142000 mill.	9000 mill.	8 mill.
Hill de dim. 5 :	2^{115}	10^6 univers	10^5 univers	83 univers
Clé 128 bits :	2^{128}	10^{10} univers	10^8 univers	10^5 univers
Clé 192 bits :	2^{192}	10^{11} univers^2	10^{10} univers^2	10^7 univers^2

exercice : La force brute

Le facteur de travail d'un algorithme est le nombre d'instructions élémentaires nécessaire à son exécution. La puissance d'une machine est le nombre d'instructions qu'elle exécute par unité de temps. Nous allons approximer la puissance d'un PC actuel à environ 2000 Mips (millions d'instructions par seconde). Le facteur de travail d'un algorithme optimisé pour tester une clé de 128 bits de l'algorithme AES est d'environ 1200 instructions élémentaires. On dispose d'un couple clair/chiffré connu et on désire retrouver la clé utilisée par force brute, c'est-à-dire en testant toutes les clés les unes après les autres. Une clé est constituée d'un mot de 128 bits. On suppose que toutes les clés sont équiprobables.

1. En combien de temps une machine de 2000 Mips teste-t-elle une clé ?
2. Combien y a-t-il de clés possibles ? Quel est le nombre moyen de clés 'a tester avant de trouver la bonne ?
3. A quel temps moyen de calcul cela correspond-il si on suppose qu'un seul PC effectue la recherche ? Si les 1 milliard de PC de l'Internet sont mobilisés à cette tâche ?

Diffie Hellman : Echange de clé

- **Algorithme à clé publique inventé en 1976 Objectif :**
- Permet l'échange d'une clé secrète sur un domaine non sécurisé, sans disposer au préalable de secret **Utilisation :**
- entre autres, dans SSL/TLS (Netscape).

Repose sur :

- connaissant g*a mod p et g*b mod p, il est très difficile d'en déduire a et b.

Diffie Hellman

- Privé : a pour Alice, b pour Bob Public :
- p : nbre premier
- G : appelé **générateur**
- clé publique d'Alice , clé publique de Bob