Cryptology

Les bases de la cryptologie

Antoine Puissant

Enseignants : M. Filiol, M. Nicolas Bodin

2015 - 2016

Résumé

Ceci est le résumé

Table des matières

1 Introduction

Tous les états ont la copie des certificats des grands groupes (Google, Microsoft, etc...).
La cryptologie c'est la science du secret. Il faut savoir contre quoi et qui on se protège, et comment.
Un système d'information est un système fonctionnant à l'électricité, traitant de l'information depuis sa création jusqu'à sa destruction.
Un système d'information peut être soumis à deux types d'attaques :

- Les actions naturelles, non malveillantes, non ciblés. On parle alors de sûreté.
- La sécurité. On a ici un caractère ciblé, adaptatif, imprévisible.

Ici, on ne s'occupera que de la sécurité.On va parler de cryptographie dans la défense et de cryptanalyse dans l'attaque.
Pour contrer les attaques, il va falloir protéger l'information. Pour cela, on va passer d'un réseau rouge (non chiffré) à un réseau noir (chiffré). On va ainsi parler de COMSEC (Communication Security). Ici le canal n'est pas sécurisé. Pour sécuriser aussi le canal, on va parler de TRANSEC [1] comme avec la stéganographie.

1.1 Les acteurs de la cryptologie

Il y a un ou plusieurs émetteurs, un ou plusieurs destinataires. Cela se fait sur un canal ou un réseau public. Il peut y avoir un ou plusieurs intrus malveillants. Enfin, il peut y avoir un ou plusieurs messages.

2 Les menaces et solutions de la cryptologie

La cryptologie est partout.
On va pouvoir retrouver quatre menaces :

- Atteinte à la confidentialité des données. E va écouter A.
- E envoie un message à B en se faisant passer par A. On a lors un problème d'identification et d'authentification.
- E modifie un document. On a alors un problème d'intégrité.
- A envoie un message à B puis le répudie. On a alors un problème de signature.

Il y a aussi des problèmes annexes tels que :

- Les problèmes d'accusés de réception ; d'émission
- Les problèmes de falsification de la date de création d'un message, d'un problème d'horodatage.

Pour ces deux problèmes, on peut mettre en place des systèmes de vérification poussée. Cependant, au bout d'un moment, on doit se baser sur la confiance entre humains.
On a aussi les problèmes TRANSEC : brouillage de communication, coupure de canal, A qui veut cacher l'envoie d'un message.

1. Transmission Security

2.1 Les solutions

La confidentialité Les données doivent rester inintelligibles à toutes personnes non destinataires

L'intégrité

L'authentification

La signature

2.2 Le chiffrement

2.2.1 Chiffrement symétrique

A et B ont la même clé secrète pour chiffrer et déchiffrer.
Le problème reside dans la gestion de clés car A et B soivent avoir les clés au préalable de la communication.

Chiffrement par flot

Système qui opère individuellement sur chaque bit du clair en utilisant un transformation qui varie en fonction du bit d'entrée. C'est un système de chiffrement à la volé.
On va beaucoup retrouver ce système de chiffrement dans les systèmes de communications.

L'avantage de ce chiffrement est la rapidité. Son implémentation en terme de porte logique est beaucoup plus simple en électronique. Cela peut permettre un système de secret parfait.
Le gros inconvéniant réside dans la mise en place et la distribution de clés. Le secret parfait est aussi difficile à réaliser (clé assez longue pour être de la même taille que le clair).

2.2.2 Chiffrement asymétrique

On va ici couper l'information (qui est longue) en blocs.
Système qui divise le clair en blocs de taille fixe et chiffre un bloc à la fois. La taille des blocs est en général de 64 à 128 bits.
Dans un système de chiffrement à clés publiques, chaque utilisateur dispose d'un couple de clés : une clé privée et une clé publique.
La clé privée n'est connue que de l'utilisateur qui l'a créée.
D'un point de vue mathématique, on utilise les fonctions mathématiques à sens unique avec trappe.

Fonction à sens unique

Une fonction $f : M \to C$ est une FSU si :
— Connaissant $m \in M$, il est facile de calculer $f(m)$
— Connaissant $f(m)$, il est difficile de trouver m

Fonctions avec trappe

Une telle fonction f est dite avec trappe, si de plus la connaissance de $f(m)$ est d'une information supplémentaire (clé privée) permet de retrouver m facilement.
La construction repose de ces fonctions repose sur des problèmes mathématiques dits très difficile : factorisation, résidu quadratique, logarithme discret, problèmes d'empilement, etc...
La sécurité de ces systèmes repose sur la supposition (non démontrée) qu'il existe des problèmes véritablement difficiles.

L'avantage de ce chiffrement, il n'y a pas d'échange de clé nécessaire.
Cependant, ce système est très lent et n'est que supposé sûr [2].

2.2.3 Les certificats

2.2.4 Chiffrement hybride

On chiffre le message de manière symétrique. On chiffre la clé de manière asymétrique. On envoie les deux chiffrés (la clé chiffré en asymétrique et le message en symétrique).

2.3 L'intégrité : le hachage

Une fonction de hachage est une fonction qui transforme une chaîne de longueur quelconque en une chaîne fixée (128 ou 160 bits) appelée condensé ou haché (MD5, SHA-A, RIPEMD-160).
H est une fonction de hachage si :

- $H(M)$ se calcule facilement à partir de M
- H est sans collision : connaissant M et $H(M)$, il est difficile de trouver $M' \neq M$ tel que $H(M') = H(M)$

Le hachage permet d'assurer l'intégrité, grâce au MIC [3].
On peut aussi assurer l'authentification si on adjoint un paramètre secret grâce au MAC [4]. Il est préférable d'utiliser un MAC afin de savoir qui a produit le message.

2.4 La signature électronique

2.4.1 Authentification

A envoie un message à B. On dit que B authentifie A si et seulement si :

- A peut prouver à B qu'il est bien A
- E≠A ne peut prouver à B qu'il est A

Avec cela, on a que que de l'identification. Pour avoir de l'authentification, il faut être sûr que le message n'a pas été modifié. L'intégrité doit être prise en compte.

Authentification = identification + intégrité

2. Snowden dual Random Bit Generator
3. Message Integrity Code
4. Message authentification Code

Un message M est signé par A si et seulement si :
— A peut prouver à B qu'il est bien A
— E≠A ne peut prouver à B qu'il est A
— B peut prouver à un tiers D que A est l'auteur (et le seul) du message M
Si D=A, alors il y a non répudiation par A

Signature = Authentification + conviction de transfert

Dès que l'on a un message M assez volumineux, on va prendre un hache de M, $H(M)$. On va alors signer le hache, $H(M)$ et chiffrer M.

3 Cryptanalyse

3.0.2 Les clés

La sécurité doit reposer uniquement sur la clé, pas sur le système cryptographique.
Une clé est un paramètre secret d'un algorithme cryptographique.
Une clé est composée de deux facteurs :
— Entropie : C'est l'incertitude autour de la clé.
— Cryptopériode : C'est la durée de vie opérationnelle d'une clé.
Sur les appareils embarqués, on utilise de la Lightweight Cryptography (64 bits).

En terme d'attaque, il faut supposer que l'algorithme est connu.
On va avoir deux types d'attaques :
— Attaque à chiffré seul : l'attaquant connaît le chiffré et veut connaître le clair et si possible la clé.
On peut faire des attaques exhaustive de clés (attaques par dictionnaire) ou de l'utilisation de redondance dans le clair.
— Attaque en clair probable, connu ou choisi

4 Cryptographie symétrique

4.1 Introduction

On a un message M, K une clé et Γ un cryptogramme.
Equivocation de la clé, c'est l'incertitude de la clé. Elle doit toujours être supérieure à celle du clair (il doit être beaucoup plus complexe de trouver la clé que le clair).
L'AES et le DES ne sont pas des système à chiffrement parfait. Il y a un nombre fixé de clés. Ainsi, on doit réutiliser certaines clés. Les seuls systèmes parfaits sont les systèmes dont la clé est aussi longue que le texte. C'est une addition modulo 2 de la clé et du clair (bit à bit). C'est un système soit aléatoire, soit pseudo aléatoire en fonction de la génération de la suite chiffrante.
Sur les systèmes par flot, s'il y a une erreur lors de la transmission du chiffré, cette erreur ne ne propage pas.

4.2 Machines à pseudo-aléa

On a un algorithme dans lequel on met une clé K dont la taille est fixée. Cette clé, au travers d'un automate va être expansées en un cle $\Sigma K(C)$.

Exercice

Soit un DES à 3 tours

5 Modes de chiffrements

Un mode permet de traiter les chiffrer et clairs. Cela va permettre de passer de blocs en blocs. Dans la vie de tous les jours, ça ajoute une sécurité dans le gestion des blocs.

5.1 Electronic Code Book – ECB

Assez intuitif. On prend 64 bits, on chiffre et on le repose sur le disque. Problème de redondance.

5.2 CBC

On chiffre une fois et au tour d'après, on XOR le plaintext avec le ciphertext précédent. Ainsi, on a alors une propagation en cas d'erreur. La propagation permet de garder l'intégrité du message.

5.3 CTS

CBC + Stealing.

5.4 XTX

XEX + Stealing. Potentiellement créé pour l'AES.

5.5 GCM

Deux parties : une de chiffrement (Counter mode) et une d'authentification.

5.6 OFB

Pas de propagation d'erreurs.

6 Le Data Encryption Standard (1975)

Le texte de clair subit en premier lieu une permutation initiale. On a ensuite 16 tours. Enfin, une permutation initiale inverse afin de fournir le chiffré.

Exercice

Etude des protocoles :
— **TLS/SSL/HTTPS**
— IPSEC
— kerberos
— zero knoledge
— signature en aveugle (blind signature)
— Secure multi party computation

Quelle est la partie chiffrement, authentification, gestion de clé, certificat, etc...
Un dizaine de slides par présentations (le premier puis un dans la liste).

7 Introduction

La stéganographie amène plus de sécurité à la cryptographie. Le souci avec la cryptographie, c'est qu'on sait que le communication a lieu (incompréhensible). Avec la stéganographie, on va chercher à cacher le message dans un support anodin (invisible).
La stéganalyse c'est l'attaque. Son but est de détecter s'il y a eu une communication secrète. On ne cherche pas à récupérer un message.

7.1 Watermarking

C'est une façon d'insérer une marque dans un support. Cette marque peut être visible ou invisible, robuste ou faible.
Insérer des informations de l'acheteur dans une image (de manière invisible), c'est du fingerprinting.
Une marque robuste ne va pas être sensible au crop, à la compression, au changement de format. Un watermark fragile va casser à chaque modification d'une image.

8 Stéganographie

8.1 Problème du prisonnier

A et B son en prison. Ils veulent s'évader mais ne sont psa dans la même cellule. Il peuvent communiquer via un gardien. Il est alors impossible d'écrire des messages en clair, ni de la cryptographie sans que le gardien ai des doutes. Il vont alors utiliser de la stéganographie pour cacher les messages.
Maintenant pour qu'Alice envoie un message à Bob, elle va devoir suivre ses différentes étapes :

1. Encryption du texte qui nous donne un cyphertext. On ne met jamais un texte clair dans un fichier. On insère toujours notre flux dans une zone proche de l'aléa. Ainsi le cyphertext sera invisible.
2. On va insérer le texte. Pour cela, il nous faut 3 éléments : le cypherText, un cover medium et une clé stégano. On va alors avoir le stego medium. La clé stégano est, de manière générale, symétrique. Elle va permettre de savoir où chercher dans l'image. Elle va créer une permutation et va donner les positions d'insertion du message.

8.2 Définitions

Message : cypherText de N_M bits

Cover file : support de communication avec N_C élément modifiables pour ne pas « pourrir » mon support.

Taux d'insertion : $\frac{N_M}{N_C}$

8.3 Types d'insertion

8.3.1 Cover modification

On insère des bits dans les bits de données de l'image.

8.3.2 Cover selection

On cherche une source qui contient directement l'information dans son image (algorithme de Hopper). On a alors une image non modifiée.

8.3.3 Cover synthesis

A partir du message, on à la possibilité de générer un support anodin contenant l'image. Pour le moment, ce n'est pas encore faisable.

8.4 Types de données

8.4.1 Compressées

Images JPEG

Sons MP3, WMA, Ogg

Vidéos M-JPEG, MPEG, DivX, etc...

8.4.2 Non Compressées

Images BMP, RAW, PNG, GIF

Sons WAV, PCM

Vidéos impossible (trop gros)

8.4.3 Autre

Dans les documents (textes, pdf), pages web, programmes (codes sources), protocoles réseaux (champs non utilisés, VoIP, etc...).

8.5 Acquisition de l'image

Dans une image, le header est composé des informations de l'image et d'une palette. La paletter fait l'association entre une valeur en 3 octets (RGB) à une valeur d'un octets. On fait alors le lien entre les valeur des pixels et la palette. Pour cacher de l'information, on peut mettre des commentaires. Ce n'est pas caché de manière importante.
On peut modifier les pixels. On peut aussi faire des permutation sur la palette. C'est la permutation de la palette qui va mettre le message. Ça ne modifiera pas l'image mais en analysant la palette, on se rend compte que la palette n'est pas ordonnancée de manière logique.

8.6 Pixel modification : LSB replacing

LSB : Least Significant Bits. On fait un LSB des valeurs de pixels qui ont l'air random.
Le LSB replacing fonctionne bien pour des petits fichiers.

8.7 Format JPEG

On part d'une image RGB, on la compresse.
On va changer l'espace colorimétrique. On va passer du RGB au YCbCr car l'œil humain est plus sensible au vert qu'au bleu et qu'au rouge.
On fait ensuite un sous-échantillonnage (perte de données, irréversible).
On fait ensuite une DCT [5]. Cela transforme une image en une somme de fréquences.
Une fois que l'on a la matrice des fréquences, on réalise une quantification. Plus on va choisir une quantification élevée, plus on va perdre de l'information.
A ce niveau la, on va retrouver beaucoup de 0 dans notre fichier. On va vouloir utiliser l'algorithme RLE pour la compression. On fait un parcours en zigzag afin de lire les coefficients supérieurs à 0 en premier. Un fois que nous avons notre suite, nous pouvons appliquer RLC. On va ensuite produire un arbre de Huffman pour chaque bloc BCT.
Afin d'insérer un message, il n'est seulement possible d'insérer notre message après la quantification.

9 Stéganalyse

Cela sert à détecter une communication secrète. Parmi un échantillon récupéré, on cherche s'il y a des cover et des stégo.

5. Discrete Cosine Transform

10 Cryptanalyse asymétrique

10.1 RSA

Le système à clé publique le plus utilisé.
On va prendre deux nombres premiers très grands p et q.

10.1.1 Signature

A émet un message m et veut le signer.

11 Detection and Operational Cryptanalis of weakly implementations

11.1 Cryptographie théorique VS cryptographie pratique

Quel est l'impact d'une mauvaise utilisation de clés ou une mauvaise implémentation d'un algorithme ?
$a = 0,3$, overlapping, souple (entre 5 et 10).

12 Principe des attaques par corrélation

for *init* $I \in R_1$ **do**
  Initialiser R_1 avec I
  Produire autant de bits (N) qu'il y a de cryptogrammes
  Calculer $Z_I = \sum_{i=1}^{k+1} \left(x_i^{\perp} \oplus c_T \oplus 1\right)$
  Stocker I et Z_I
end
Trier selon les Z_I
On a ensuite deux cas possibles :

I n'est pas le bon init
$Z_I \to \mathcal{N}\left(\frac{N}{2}, \frac{\sqrt{N}}{2}\right)$

I est le bon init
$Z_I \to \mathcal{N}\left(P_T.N, \sqrt{NP_T(1-P_T)}\right)$

13 Terminologie basique

Code Un code est une convention destinée à être diffusé le plus largement possible. Ce sont des conventions publiques, il n'y a pas de secret.

Chiffre Un chiffre est une convention qui est destinée à être diffusé le moins possible. La convention secrète est composé de une ou plusieurs clés.

Chiffrement C'est l'opération consistant à transformer un texte clair en une suite inintelligible d'apparence aléatoire (texte chiffré) en fonction d'un ou plusieurs éléments secrets (clés).

Déchiffrement C'est l'opération permettant d'accéder de façon légitime au texte clair en fonction de clés eventuelllement différentes de celles utilisées lors du chiffrement. On va trouver deux types de systèmes :
- Systèmes symétriques. La même clé est utilisée lors du chiffrement et du déchiffrement.
- Systèmes asymétriques.

Texte chiffré

Clé Paramètre secret fondamental qui intervient dans le processus de chiffrement.

Substitution Ce principe consiste au remplacement des caractères du texte en clair d'autres caractères.

Transposition Les caractères du texte en clair demeurent inchangés mais les positions respectives sont modifiées.

Cryptosystème C'est uns système qui comprend un algorithme de chiffrement, un algorithme de déchiffrement, un clair, un chiffré et une clé.

Cryptanalyse Opération consistant de manière illégitime à retrouver oar des méthodes mathématiques la ou les clés et le texte claur à partir du texte chiffré avec ou sans l'algorithme (encore appelé décryptement).

Cryptanalyse appliquée C'est l'opération identique que la cryptanalyse ; les méthodes utilisées sont de natures diverses et visent le système au niveau de son implémentation ou de sa gestion. Il y a plusieurs points à vérifier :
- L'implémentation (attaques par canaux cachés). On va retrouver différent système DPA, injection de fautes, etc... (side channel attacks). On a ici des attaques logicielles et matérielle.
- La gestion. La façon dont laquelle les clés sont gérées (taille, injection, etc...).
- Le risque humain.

Références

[1] AUTHOR. *REF TITLE*. ORG. 2015. URL : `http://www.url.com/`.

[2] BLUEKRYPT. *Cryptographic Key Length Recommendation*. BlueKrypt. 2015. URL : `http://www.keylength.com`.

[3] David KAHN. *The Codebreakers – The Story of Secret Writing*. Sphere, 1967. ISBN : 0-684-83130-9.

[4] Simon SINGH. *The code Book*. Livre de poche, 1999.