CHAPITRE 1.

INTRODUCTION A LA CRYPTOGRAPHIE

Cruptos ($\chi\rho\upsilon\pi\tau o\sigma$) : caché, dissimulé

Graphein ($\gamma\rho\alpha\phi\epsilon\iota\nu$) : écrire

`http://math.univ-lyon1.fr/~roblot/masterpro.html`

Notions de base

Les quatre buts de la cryptographie

Confidentialité : mécanisme pour transmettre des données de telle sorte que seul le destinataire autorisé puisse les lire.

Intégrité : mécanisme pour s'assurer que les données reçues n'ont pas été modifiées durant la transmission.

Authentification : mécanisme pour permettre d'identifier des personnes ou des entités et de certifier cette identité.

Non-répudiation : mécanisme pour enregistrer un acte ou un engagement d'une personne ou d'une entité de telle sorte que celle-ci ne puisse pas nier avoir accompli cet acte ou pris cet engagement.

Terminologie

Alphabet $\mathcal{A}$: ensemble fini de symboles utilisés pour écrire les messages.

Message clair m : chaîne de caractères composée de lettres de l'alphabet $\mathcal{A}$ et dont on veut en général conserver la confidentialité. On note $\mathcal{M}$ l'ensemble de tous les messages clairs possibles.

Message crypté c : chaîne de caractères composée de lettres de l'alphabet $\mathcal{A}$, correspondant à un message clair, et dont la diffusion à des entités non autorisées ne doit pas dévoiler pas d'information sur ce message clair. On note $\mathcal{C}$ l'ensemble de tous les messages cryptés.

Cryptage : transformation d'un message clair en un message crypté.

Décryptage : transformation inverse du cryptage qui permet de retrouver à partir d'un message crypté, le message clair correspondant.

Terminologie

Signature s : chaîne de caractères associées à un message donné (et aussi possiblement à une entité) et le caractérisant.

Transformation T_k : fonction qui associe à un message clair ou crypté, une autre donnée qui peut être un message clair, crypté ou une signature. En général, ce sont des fonctions qui dépendent de clés.

Clé k : donnée supplémentaire permettant de construire les fonctions de cryptage et de décryptage. Sans connaissance de la clé de décryptage, le décryptage doit être *impossible*. On note $\mathcal{K}$ l'ensemble de toutes les clés.

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...

Fonctions cryptographiques

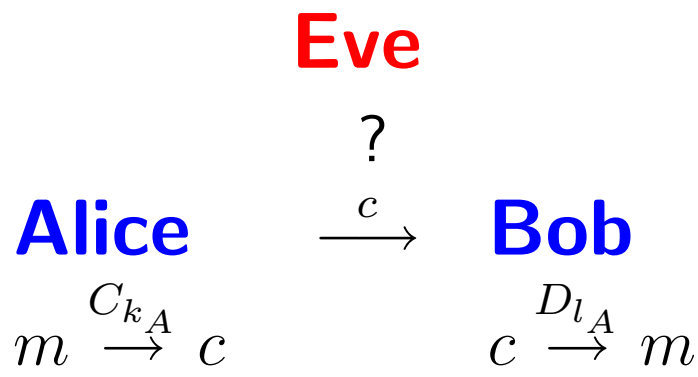
Notations. Pour une fonction $f : X \to Y$. on note $\mathrm{Im}(f)$ l'image de f et f^{-1} la fonction réciproque (si f est bijective).

Problème de l'inversion. Pour une fonction $f : X \to Y$ et pour $y \in \mathrm{Im}(f) \subset Y$. Le problème de l'inversion est de trouver $x \in X$ tel que $f(x) = y$.

Fonction à sens unique. C'est une fonction $f : X \to Y$ telle que, pour *presque tout* élément $y \in \mathrm{Im}(f)$, le problème de l'inversion est *impossible* à résoudre. On veut aussi que $f(x)$ soit facile à calculer pour tout $x \in X$.

Fonction à sens unique à trappe. C'est une fonction $f : X \to Y$ à sens unique et telle que la connaissance d'une donnée supplémentaire, la trappe ou clé, permet de résoudre facilement le problème de l'inversion.

Confidentialité



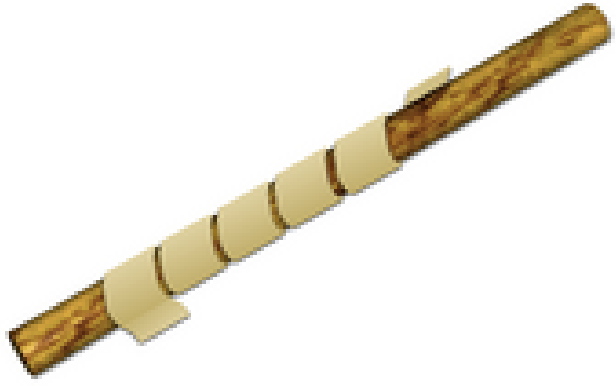
Protocole.

- $\mathcal{M}$ ensemble des messages clairs
- $\mathcal{C}$ ensemble des messages cryptés
- $\mathcal{K}$ ensemble des clés
- $C_k : \mathcal{M} \to \mathcal{C}$ fonction de cryptage
- $D_l : \mathcal{C} \to \mathcal{A}$ fonction de décryptage

Exemples historiques de protocoles de cryptographie

1. La scytale
2. Le cryptogramme de César
3. La permutation de lettres
4. Le chiffrement de Vigenère
5. Le chiffrement de Hill

Scytale



Message crypté

KTMIOILMDLONKRIIRGNOHGWT

Cryptogramme de César

$$
\begin{array}{lllllll}
A \to E & B \to F & C \to G & D \to H & E \to I & F \to J & G \to K \\
H \to L & I \to M & J \to N & K \to O & L \to P & M \to Q & N \to R \\
O \to S & P \to T & Q \to U & R \to V & S \to W & T \to X & U \to Y \\
V \to Z & W \to A & X \to B & Y \to C & Z \to D &
\end{array}
$$

Exemple

`ATTAQUE AU MATIN` $\longrightarrow$ `EXXEUYI EY QEXMR`

Clé : entier entre 1 et 26

Permutations de lettres

$$
\begin{array}{lllllll}
A \to D & B \to R & C \to K & D \to X & E \to V & F \to H & G \to L \\
H \to N & I \to S & J \to O & K \to P & L \to Q & M \to W & N \to I \\
O \to T & P \to J & Q \to E & R \to U & S \to Z & T \to A & U \to C \\
V \to F & W \to B & X \to Y & Y \to G & Z \to M &
\end{array}
$$

Exemple

`ATTAQUE AU MATIN` $\longrightarrow$ `DAADECV DC WDASI`

Clé : permutations sur 26 lettres

Nombre de clés : $26! = 403291461126605635584000000 \simeq 2^{88}$

Cryptanalyse de la permutation de lettres

Message à décrypter

CEGCL AM NMGAL LJC ZWIWJLL LH CYEWJ RMYCWLJ ZEHC
GHL LJC UMQWCLL RMY ALJ QLANLJ A MGCYL RMY ALJ MVGWCMWHJ
AM CYEWJWLPL RMY SLGF VGW ZMHJ ALGY AMHNGL JL HEPPLHC
SLACLJ LC ZMHJ AM HECYL NMGAEWJ

On attaque en utilisant les fréquences d'apparition des lettres en français.

E	A	S	I	N	T	R	L
17,3%	8,4%	8,1%	7,4%	7,1%	7,0%	6,6%	6,0%

L	M	C	J	A	W	H	G
27	16	15	15	12	11	10	10

Essai de décodage : (L, M) → (E, A)

CEGCE AA NAGAE EJC ZWIWJEE EH CYEWJ RAYCWEJ ZEHC GHE EJC
UAQWCEE RAY AEJ QEANEJ A AGCYE RAY AEJ AVGWCAWHJ AA
CYEWJWEPE RAY SEGF VGW ZAHJ AEGY AAHNGE JE HEPPEHC SEACEJ
EC ZAHJ AA HECYE NAGAEWJ

Essai de décodage : (L, M) → (E, A)

CEGCE AA NAGAE EJC ZWIWJEE EH CYEWJ RAYCWEJ ZEHC GHE EJC
UAQWCEE RAY AEJ QEANEJ A AGCYE RAY AEJ AVGWCAWHJ AA
CYEWJWEPE RAY SEGF VGW ZAHJ AEGY AAHNGE JE HEPPEHC SEACEJ
EC ZAHJ AA HECYE NAGAEWJ

Essai de décodage : (L, M, A) → (E, A, L)

CEGCE LA NAGLE EJC ZWIWJEE EH CYEWJ RAYCWEJ ZEHC GHE EJC
UAQWCEE RAY LEJ QELNEJ L AGCYE RAY LEJ AVGWCAWHJ LA
CYEWJWEPE RAY SEGF VGW ZAHJ LEGY LAHNGE JE HEPPEHC SELCEJ
EC ZAHJ LA HECYE NAGLEWJ

Essai de décodage : (L, M, A, J) → (E, A, L, S)

CEGCE LA NAGLE ESC ZWIWSEE EH CYEWS RAYCWES ZEHC GHE ESC
UAQWCEE RAY LES QELNES L AGCYE RAY LES AVGWCAWHS LA
CYEWSWEPE RAY SEGF VGW ZAHS LEGY LAHNGE SE HEPPEHC SELCES
EC ZAHS LA HECYE NAGLEWS

Essai de décodage : (L, M, A, J, C) → (E, A, L, S, T)

TEGTE LA NAGLE EST ZWIWSEE EH TYEWS RAYTWES ZEHTGHE EST
UAQWTEE RAY LES QELNES L AGTYE RAY LES AVGWTAWHS LA
TYEWSWEPE RAY SEGF VGW ZAHS LEGY LAHNGE SE HEPPEHT SELTES
ET ZAHS LA HETYE NAGLEWS

E	A	S	I	N	T	R	L
17,3%	8,4%	8,1%	7,4%	7,1%	7,0%	6,6%	6,0%

L	M	C	J	A	W	H	G
27	16	15	15	12	11	10	10

Essai de décodage : (L, M, A, J, C, W, H) → (E, A, L, S, T, I, N)

TEGTE LA NAGLE EST ZIIISEE EN TYEIS RAYTIES ZENT GNE EST
UAQITEE RAY LES QELNES L AGTYE RAY LES AVGITAINS LA
TYEISIEPE RAY SEGF VGI ZANS LEGY LANNGE SE NEPPENT SELTES
ET ZANS LA NETYE NAGLEIS

. . . . . .

http://math.univ-lyon1.fr/~roblot/masterpro.html

Chiffrement de Vigenère

Correspondance lettre $\leftrightarrow$ nombre. $A = 0,\ B = 1,\ \ldots,\ Z = 25$

Addition sur les lettres. J + W = F $(9 + 22 \mod 26 = 5)$

Exemple

```
NOUS ATTAQUERONS AU MATIN PAR LE NORD
VIGE NEREVIGENER EV IGENE REV IG ENER
KYCY PZMGNEMXDTL GR WIZXT IGO VM TDXW
```

Clé : mot de l lettres (l fixé)

Nombre de clés : 26^l (Note : $26^{19} \simeq 26!$)

Cryptanalyse du chiffrement de Vigenère (Friedman-Babbage-Kasiski)

Indice de coïncidence.
Probabilité que 2 lettres prises au hasard dans un texte soient égales

$$IC = \frac{1}{n^2} \sum_{l=A}^{Z} n_l^2$$

Fréquence et IC. Notons p_l la fréquence de la lettre $l = A, \ldots, Z$.
Alors, on a

$$IC = \sum_{l=A}^{Z} p_l^2$$

Il suit que l'IC d'un texte (suffisamment long) en français est 0.078.

Exercice. Démontrer les diverses formules de l'IC. Puis, calculer le IC d'un texte totalement aléatoire.

Exercice sur les propriétés de l'indice de coïncidence

On travaille avec l'alphabet $\mathcal{A} = \{A, B, C\}$. On suppose que ces lettres apparaissent dans un texte avec les probabilités suivantes

$$\text{A} = 68\%, \quad \text{B} = 18\%, \quad \text{C} = 14\%$$

1. Calculer l'indice de coïncidence du texte.
2. On applique la transformation (A, B, C) $\rightarrow$ (B, C, A) au texte. Calculer l'indice de coïncidence du texte obtenu.
3. On applique la transformation (A, B, C) $\rightarrow$ (B, C, C) au texte. Calculer l'indice de coïncidence du texte obtenu.
4. Soit $k \geq 2$. On considère le texte obtenu en sélectionnant une lettre sur k. Estimer son indice de coïncidence ?
5. On applique un chiffrement de Vigenère de longueur k au texte, puis on sélectionne une lettre sur k dans le texte chiffré. Calculer l'indice de coïncidence du nouveau texte.

Message à décrypter

KIIFSIRV A E NEEF HJYKR SPFCVI GI KRVMMSYI XSLC HZ ZVAX YI EBVY IJG UPM JR
HZGYNMIE RH QDPZRY YI C RUPMEBBZ HV PIOXV NRIIV RX KIEQEIX CRUPIC YI WEIBQZXIR
XJQSN E NIGG GZRK QMS QZYPDQVGVZW TR JPX LA SPVRTEI WRAW DRKRVHMKGIIGV DYD HLEE YY
UVB CYZG EP ZZAKO GZAU HEIF PZW INZVKVF UP MC CVJHLVWDX WHVZRK VQHIEFIN IE
NQZVZDYZ IE RYMSGR II EJVI NYI HRZ DFAI GEITI YI UVB CYZG GZRKF QDPCRW LYZ FI
YIJFMIEZG SWPZDYZQVAX V P VDYVXVHV YIGHMN PV GVZRKR GDRHHMZQV CEMECYIGI EBVY NLFUP
EL DYVVRAXDIDR TVVRYPZPV FYY

IC des sous-séquences :

Intervalle 1 :	KIIFSIRVAENEEFHJYKRSPFCVIGI...	0,05257
Intervalle 2 :	KISRANEHYRPCI.. IFIVEEFJKSFV...	0,05184
Intervalle 3 :	KFREEJRFI... ISVNFYSCG... IIAEHKPVI...	0,05245
Intervalle 4 :	KSAEYPI... IIEFKFG... IRNHRCI... FVEJSVK...	0,05053
Intervalle 5 :	KINJ... IREY... IVEK... FAFR... SEHS...	0,07981

$\Rightarrow$ Clé $C_1C_2C_3C_4C_5$ de longueur 5

Message extrait 1 : 1 lettre sur 5

KINJPGMXZYYPZIDYPZOIKIPWZJNZSDZPPIDHIDYCPOHZVPJDZHNZZMIN-
ZGYCZDLYIWZVVYNZDZMGYPVDVZY

Première lettre : 16 Z $(19,3\%)$ donc C_1 = Z - E = V

Message extrait 2 : 1 lettre sur 5

IREYFIMSZIIMGEPIMHXIIXIEXQIRQQWXVWRMGHYYZGEWKMHXRIIVISEY-
DEIYRPYIEPQPXIPRRQEINEVIVP

Deuxième lettre : 17 I $(20,7\%)$ donc C_2 = I - E = E

• • • • • •

http://math.univ-lyon1.fr/~roblot/masterpro.html

Théorème

Le protocole de chiffrement de Vigenère avec une clé de longueur égale ou supérieure à la longueur du message, et utilisée une et une seule fois, est totalement sûr.

Chiffrement de Hill

Correspondance lettre $\leftrightarrow$ nombre. $A = 0,\ B = 1,\ \ldots,\ Z = 25$

Action des matrices. Un matrice 2×2 agit sur les vecteurs de longueur 2 de la manière suivante

$$(x, y) \begin{pmatrix} a & b \\ c & d \end{pmatrix} = (ax + cy, bx + dy)$$

En réduisant les nombres modulo 26, on en déduit une action sur les couples de lettres, par exemple pour la matrice $M = \begin{pmatrix} 3 & 21 \\ 5 & 8 \end{pmatrix}$

$$\texttt{RE} \to (17, 4) \xrightarrow{\times M} (71, 389) \quad \xrightarrow{\text{mod } 26} (19, 25) \to \texttt{TZ}$$

Exercice. A quelles conditions existe-t-il une matrice N permettant de renverser la transformation ? Calculer cette matrice pour l'exemple ci-dessus.

Exemple

Clé : $M = \begin{pmatrix} 3 & 21 \\ 5 & 8 \end{pmatrix}$ inversible

Message : RENDEZ VOUS CE SOIR

RE	ND	EZ	VO	US	CE	SO	IR
$(17, 4)$	$(13, 3)$	$(4, 25)$	$(21, 14)$	$(20, 18)$	$(2, 4)$	$(18, 14)$	$(8, 17)$
$\times M$	$\times M$	$\times M$	$\times M$	$\times M$	$\times M$	$\times M$	$\times M$
$(19, 25)$	$(2, 11)$	$(7, 24)$	$(3, 7)$	$(20, 18)$	$(0, 22)$	$(20, 22)$	$(5, 18)$
TZ	CL	HY	DH	US	AW	UW	FS

Nombre de clés : $(2^2 - 1)(2^2 - 2)(13^2 - 1)(13^2 - 13) = 157\,248$

Généralisation. On peut utiliser des matrices de taille plus grande, pour des matrices de taille 5, on obtient un nombre de clé de l'ordre de 2^{115}.

Cas particulier. Le digramme AA correspond au vecteur $(0, 0)$ et donc est toujours codé sur lui-même pour toute matrice M. Plusieurs méthodes permettent de pallier à ce problème.

Cryptanalyse du chiffrement de Hill

Attaque par les digrammes. On considère les digrammes qui apparaissent le plus souvent

ES	DE	LE	EN	RE	NT	ON	ER	TE
3,3%	2,4%	2,3%	2,1%	1,9%	1,7%	1,6%	1,5%	1,5%

Attaque par texte clair connu. Si on connaît une partie du texte clair et sa version cryptée, on peut en déduire des informations sur la clé, voire la clé entière.

Attaque par information partielle. En particulier, si on peut deviner une partie du message clair, on peut en déduire des informations sur la clé.

Exercice. Sachant que la version cryptée du message clair `CRYPTO` est le message `MPODXM`, en déduire le maximum d'information sur la clé de cryptage.

Protocoles de confidentialité

Choix des clés. Bob dispose d'un couple de clés (k_B, l_B) telles que $D_{l_B}(C_{k_B}(m)) = m$ pour tout $m \in \mathcal{M}$. k_B est la clé de cryptage et l_B la clé de décryptage.

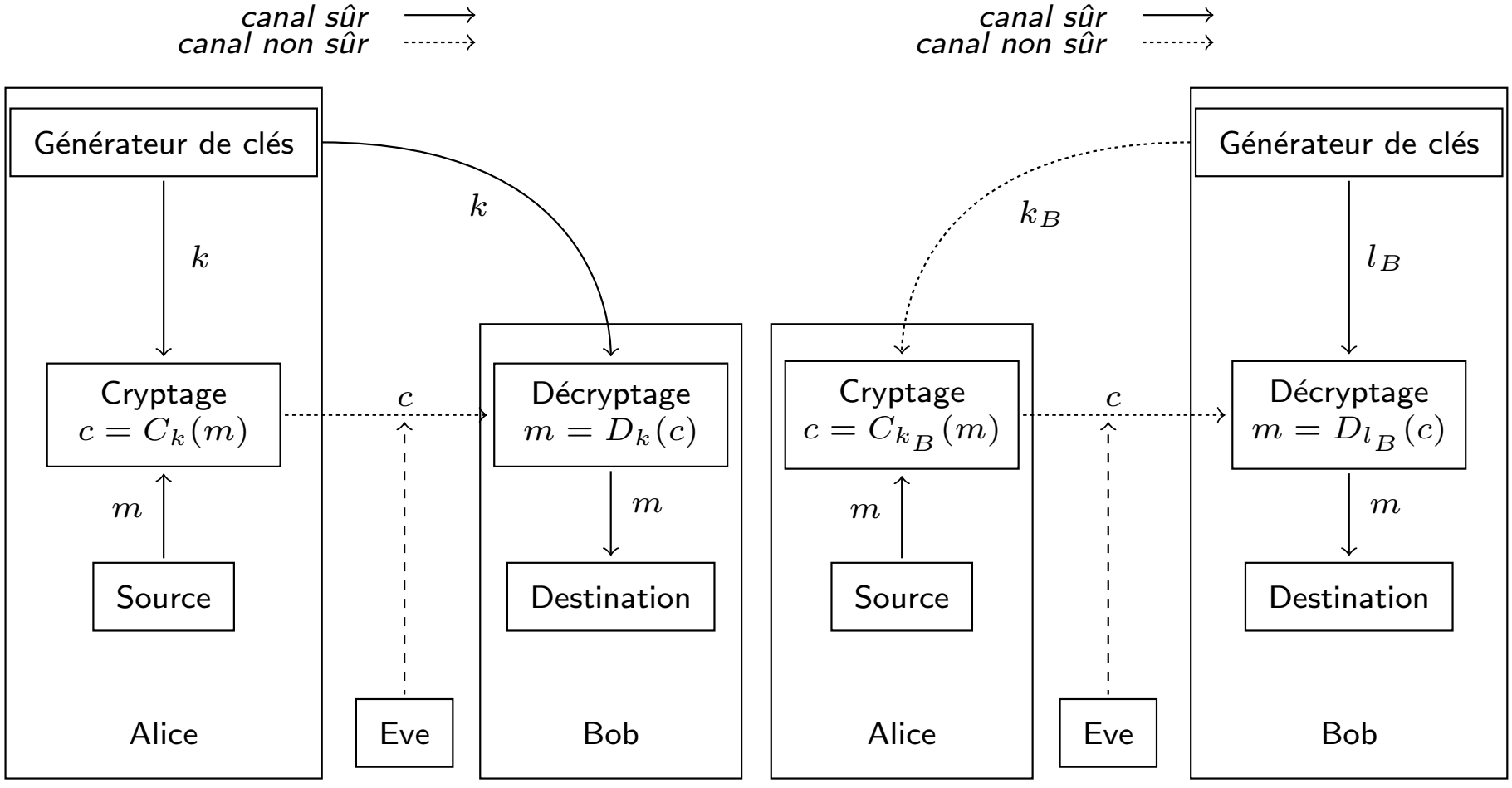
Transmission d'un message.

1. Alice veut transmettre le message m à Bob
2. Elle calcule $c = C_{k_B}(m)$ et transmet c à Bob
3. Bob retrouve le message clair à partir de c :
   $D_{l_B}(c) = D_{l_B}(C_{k_B}(m)) = m$

Types de protocole : Si $k_B = l_B$, alors c'est un protocole à clé secrète (ou symétrique), et k_B doit être secrète (l_B est toujours secrète). Sinon, k_B peut être rendue publique et c'est un protocole à clé publique.

Principe de Kerckhoff : la sécurité du protocole repose dans le secret des clés et non celui des algorithmes

Protocoles à clé secréte et à clé publique



Protocole à clé secrète

Protocole à clé publique

Quelques principes de base en cryptographie

Performance. Un protocole sûr vaut mieux qu'un protocole efficace. Exemple : attaque par chronomètre sur les cartes à puce.

Simplicité. Un protocole ne doit jamais essayer de faire plus que ce qu'il est sensé faire. Exemple : extension des protocoles d'identification.

Le Maillon Faible. Un protocole n'est jamais aussi sûr que sa composante la plus faible. Exemple : le WiFi de l'université Lyon 1.

Raisonnement paranoïaque. Un protocole avec une faiblesse, aussi petite qu'elle soit, est un protocole qui n'est plus sûr. Exemple : le protocole WEP

Modèle de sécurité. Un protocole n'est jamais parfait. L'essentiel est d'obtenir le niveau de sécurité souhaité. Exemple : le WiFi de l'université Lyon 1 ?

Cryptanalyse – Attaques classiques

L'attaquant connaît les algorithmes de cryptage et décryptage.

Attaque à texte crypté uniquement : l'attaquant ne dispose que d'un ou plusieurs messages cryptés qu'il souhaite décrypter. C'est le type d'attaque le plus difficile.

Attaque à texte clair connu : l'attaquant dispose d'exemples de messages clairs avec les messages cryptés correspondants, ou d'une partie clair d'un message crypté. Le but est d'obtenir de l'information sur la clé.

Attaque à texte clair choisi : l'attaquant peut obtenir la version cryptée d'un certain nombre de messages clairs choisis, soit avant l'attaque (attaque hors ligne), soit au fur et à mesure (attaque en ligne). Le but est encore d'obtenir de l'information sur la clé.

Attaque à texte crypté choisi : l'attaquant peut obtenir la version cryptée d'un certain nombre de messages clairs choisis, et aussi la version claire d'un certain nombre de messages cryptés choisis. On distingue encore entre attaques hors ligne et en ligne.

Cryptanalyse – Autres types d'attaques

Attaque par le paradoxe des anniversaires : Il s'agit d'obtenir des collisions (utilisation deux fois d'une même valeur) pour obtenir de l'information. Si on utilise 2^n valeurs possibles, on peut espérer la première collision avec environ $2^{n/2}$ valeurs.

Attaque par précalcul : Il s'agit pour l'attaquant de précalculer des informations et de s'en servir pour identifier des messages ou des clés. Cela nécessite plus de travail mais permet aussi plus de flexibilité. Un cas extrême est la recherche exhaustive.

Attaque de différentiation : Il s'agit d'une attaque qui permet de différencier le protocole de cryptage utilisé d'un protocole de cryptage parfait. Cela couvre les attaques citées précédemment et toutes les attaques à venir !

Le modèle de Dolev-Yao

Environnement vulnérable. 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.

Cryptanalyse

Echelle de succés

- Cassage complet : l'attaquant découvre la clé.
- Déduction globale : l'attaquant découvre des fonctions équivalentes aux fonctions de cryptage et de décryptage sans pour autant connaître la clé.
- Déduction locale : l'attaquant peut décrypter un ou plusieurs nouveaux messages cryptés.
- Déduction d'information : l'attaquant obtient de l'information sur la clé ou sur des messages cryptés.

Critères d'évaluation

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

Echelle des coûts en temps

Configurations :

A. Un ordinateur de bureau avec 4 coeurs à 2,5 GHz : 2^{36} FLOPS

B. Cluster du laboratoire de maths : 2^{40} FLOPS

C. 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