Examen Final de Cryptanalyse

Responsable : M. Bombar

Introduction

Le sujet est composé d'un quizz et de 3 problèmes indépendants. La partie 2 doit représenter à lui seul probablement la moitié de l'examen, les deux derniers sont plus courts. La partie 3 est plutôt orienté informatique, tandis que la partie 4 est plus orientée mathématiques (ça ne veut pas dire qu'une partie est forcément plus facile qu'une autre).

Le sujet est long. Il n'est pas nécessaire de le traiter entièrement pour avoir une bonne note, mais essayez d'en faire le plus possible tout de même.

1 Quizz

(Q1) À quoi sert une fonction de filtrage dans un chiffrement par flot ?

- (a) Augmenter l'aléa lors de la génération des clés
- (b) Augmenter la complexité linéaire de la suite chiffrante.
- (c) Diminuer la complexité linéaire de la suite chiffrante.
- (d) Assurer que la suite chiffrante n'est pas périodique.

(Q2) Laquelle de ces différentielles est **impossible** pour un chiffrement par blocs comme l'AES.

- (a) $\mathrm{0xFF}\cdots\mathrm{FF} \rightarrow \mathrm{0x00}\cdots\mathrm{00}$
- (b) $\mathrm{0xFF}\cdots\mathrm{F0} \rightarrow \mathrm{0xFF}\cdots\mathrm{F0}$
- (c) $\mathrm{0x00}\cdots\mathrm{00} \rightarrow \mathrm{0x00}\cdots\mathrm{00}$

(Q3) Que peut-on dire de la probabilité d'une différentielle particulière, par rapport à la même différentielle sur plusieurs rondes (c'est-à-dire la probabilité d'un chemin différentiel (ou trace différentielle) de mêmes extrémités) ?

- (a) Les probabilités sont forcément égales.
- (b) La probabilité du chemin différentiel peut-être plus faible.
- (c) La probabilité du chemin différentiel peut-être plus élevée.

(Q4) Quelle est la probabilité maximale la plus faible possible pour une boîte S sur n bits donnée ?

- (a) $\frac{1}{2}$
- (b) 2^{-n}
- (c) 2^{1-n}

(Q5) Quel est l'avantage de la cryptanalyse algébrique par rapport à la cryptanalyse différentielle ou linéaire ?

- (a) Elle est indépendante de la structure interne du chiffrement utilisé.
- (b) Elle est toujours plus rapide.
- (c) Elle demande moins de couples clair/chiffré.
- (d) Elle identifie l'existence de clés faibles.

(Q6) Soit H une fonction de hachage résistante aux collisions. La fonction $H \circ H$ est-elle résistante aux collisions ?

2 Cryptanalyse Linéaire et Différentielle

On considère le chiffrement de type SPN représenté en Figure 1.

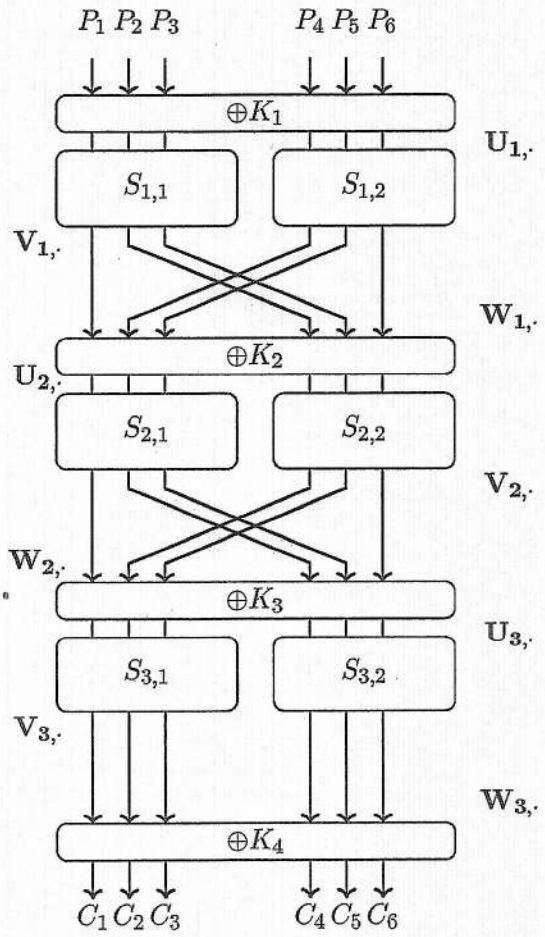

FIGURE 1 – Chiffrement par blocs de type SPN

— Les bits du texte clair sont notés $P_1, \ldots, P_6$ et les bits du chiffré sont notés $C_1, \ldots, C_6$.

— On note $U_{i,j}$ le j-ème bit d'entrée de la couche S_i. Par exemple, la boîte $S_{1,1}$ a en entrée les bits $U_{1,1}, U_{1,2}, U_{1,3}$ et la boîte $S_{1,2}$ a en entrée les bits $U_{1,4}, U_{1,5}, U_{1,6}$.

— On note $V_{i,j}$ le j-ème bit de sortie de la couche S_i.

— On note $W_{i,j}$ le j-ème bit en entrée de la clé K_{i+1}, c'est à dire que W_i est l'image de V_i par la permutation linéaire au tour i.

On suppose que les boîtes S sur 3 bits sont toutes identiques, et qu'elles sont décrites par la Table 1.

x	000	001	010	011	100	101	110	111
$S(x)$	110	101	001	000	011	010	111	100

TABLE 1 – Table définissant la boîte S vue comme une fonction Booléenne vectorielle sur 3 bits $S : \mathbb{F}_2^3 \to \mathbb{F}_2^3$

La couche linéaire de diffusion pour chaque tour est une simple permutation linéaire des bits. Elle est représentée graphiquement en Figure 1 ainsi que dans la Table 2.

input	1	2	3	4	5	6
output	1	4	5	2	3	6

TABLE 2 – Permutation

(Q7) Calculer le chiffré du texte clair 011010 en utilisant les clés

$$(K_1, K_2, K_3, K_4) = (010101, 001011, 111000, 111110).$$

Pour simplifier la correction, merci d'indiquer toutes les valeurs intermédiaires (c'est-à-dire les lignes U, V, W pour chaque étape). Vous pouvez représenter tout ça dans un tableau à 6 colonnes par exemple.

2.1 Cryptanalyse Linéaire

On rappelle que la table d'approximation linéaire d'une boîte S représente le biais des différentes approximations linéaires :

$$\mathrm{LAT}[\alpha][\beta] = 2^{n-1}\varepsilon_{\alpha,\beta}.$$

où

$$\mathbb{P}_{\mathbf{x}}\Big(\langle \alpha, \mathbf{x}\rangle + \langle \beta, S(\mathbf{x})\rangle = 0\Big) = \frac{1}{2}\big(1 + \varepsilon_{\alpha,\beta}\big).$$

La LAT de la boîte S est représentée dans la Table 3.

$\alpha \setminus \beta$	0	1	2	3	4	5	6	7
0	4	-	-	-	-	-	-	-
1	-	-2	-2	-	-	-2	2	-
2	-	-	-2	-2	-	-	-2	2
3	-	2	-	2	-	-2	-	2
4	-	-	2	-2	-	-	2	2
5	-	2	-	-2	-	-2	-	-2
6	-	-	-	-	-4	-	-	-
7	-	2	-2	-	-	2	2	-

TABLE 3 – Table des Approximations Linéaires

(Q8) Quel est le masque offrant le biais le plus important ?

(Q9) Déterminez une approximation linéaire sur plusieurs tours reliant les bits du clair P_1, P_2, P_4, P_5 avec certains bits d'entrée des boîtes S du dernier tour (donc de la forme $U_{3,j}$).

(Q10) Quelle est sa probabilité d'apparition ?

On se donne un certain nombre de couples clairs/chiffrés, représentés dans la Table 4.

Clair	Chiffré
100111	100100
000111	110010
001100	111001
011000	011101
001000	001101
011010	101001

TABLE 4 – Couples clairs/chiffrés I

(Q11) En utilisant l'approximation linéaire déterminée dans les questions précédentes, déterminer le premier et le troisième bit de la clé K_4.

Remarque

Ce chiffrement a été spécialement conçu pour que la cryptanalyse puisse se faire à la main, et ne nécessite pas un grand nombre de couples clairs/chiffrés.

2.2 Cryptanalyse Différentielle

On rappelle que l'entrée DDT$[\alpha][\beta]$ de la table de distribution des différentielles (DDT) d'une boîte S représente le nombre d'occurences de la différentielle $\alpha \to \beta$.

La DDT de la boîte S est représentée dans la Table 5.

$\alpha \setminus \beta$	0	1	2	3	4	5	6	7
0	8	-	-	-	-	-	-	-
1	-	4	-	4	-	-	-	-
2	-	-	-	-	2	2	2	2
3	-	-	-	-	2	2	2	2
4	-	-	-	-	2	2	2	2
5	-	-	-	-	2	2	2	2
6	-	4	4	-	-	-	-	-
7	-	-	4	4	-	-	-	-

TABLE 5 – Table de Distribution des Différences

On considère deux clairs P et P' de différence $\Delta P = P \oplus P' = 000001$.

(Q12) Déterminer les 6 valeurs possibles pour ΔU_3.

On se donne des couples clairs chiffrés supplémentaires représentés dans la Table 6.

Clair	Chiffré
100110	111110
000110	110110
001101	100000
011001	011111
001001	000011
011011	101000

TABLE 6 – Couples clairs/chiffrés II

(Q13) Déterminer les trois derniers bits restants pour la sous clé K_4.

3 DumbHash

Mise en situation : Vous êtes le responsable du pôle sécurité de votre entreprise. En particulier, vous êtes le chef de projet d'une application interne qui met en place plusieurs systèmes cryptographiques, dont des fonctions de hachage. Cependant, celui-ci commence à se faire vieux, et vous utilisez toujours SHA-1 même si vous la savez cassée.

Vous décidez alors de changer ça et réunissez vos ingénieurs juniors en pensant que c'est une tâche qui ne devrait pas prendre trop de temps. Un matin, un de vos ingénieurs vous amène sa nouvelle idée pour construire une fonction de hachage très rapide, sur le modèle de Merkle-Damgard, qu'il estime « super secure ». Voici son idée.

La fonction DumbHash Un message $X = x_1||\cdots||x_n$ est découpé en blocs de 256 bits, sauf peut-être x_n qui est complété par des 0. La fonction de hachage de votre jeune collègue va alors itérer une fonction de compression $C : \mathbb{F}_2^{512} \to \mathbb{F}_2^{256}$ de la forme $C(K, M) = E_K(M)$, où $E.(\cdot)$ est un chiffrement par blocs appliqué avec une clé K de 256 bits sur un bloc M de 256 bits également. Afin de démarer la construction, la fonction de hachage possède une valeur initiale correspondant à une clé formée uniquement de 0. Plus précisément, la fonction de hachage va calculer une suite $z_1, \ldots, z_n$ définie par

$$\begin{aligned} z_1 &= C(0^{256}, x_1) = E_{0^{256}}(x_1) \\ z_{i+1} &= C(z_i, x_{i+1}). \end{aligned}$$

et le haché complet de X est alors $H(X) = z_n$.

Le problème Cependant, votre ingénieur n'a apparemment pas bien appris ses cours sur les fonctions de hachage, et en particulier n'applique pas la transformation de Davies-Meyer ($z_{i+1} = E_{m_{i+1}}(z_i) \oplus z_i$). L'objectif de ce problème est d'expliquer à votre ingénieur que sa fonction n'est pas résistante aux collisions, indépendamment du choix du chiffrement par blocs E (que l'on suppose connu de l'attaquant), en construisant une attaque.

(Q14) On suppose que l'on connaît deux valeurs consécutives z_i et z_{i+1}. Montrez que l'on peut facilement retrouver un bloc x_{i+1} tel que $z_{i+1} = C(z_i, x_{i+1})$.

On va alors se servir de cette propriété pour construire des collisions. On se fixe un message m et on calcule $h \stackrel{\text{def}}{=} H(m)$.

(Q15) Montrez qu'il est possible de trouver un message $X = x_1||x_2$, différent de m, et tel que $H(X) = h = H(m)$.

Indication

Si ça peut simplifier l'écriture, vous pouvez choisir m de la forme $m = (m_1||m_2)$.

Vous présentez votre attaque à votre ingénieur, mais celui-ci vous rétorque qu'il a en effet oublié quelque chose : dans la construction de Merkle-Damgard, le padding demande de rajouter un bloc supplémentaire x_{n+1} représentant la longueur du message X, avant padding.

(Q16) Cette solution rend-elle la fonction de hachage résistante aux collisions ? Si oui, justifiez, si non expliquez comment obtenir une collision.

4 Cryptanalyse de la Structure SASAS

On considère un chiffrement par bloc de 128 bits de la forme SPN, possédant 3 couches de 16 boîtes S inversibles sur 8 bits, ainsi que deux couches affines. Un exemple est représenté en Figure 2, et une telle fonction de chiffrement est notée

$$F_{SASAS} = S^2 \circ A^1 \circ S^1 \circ A^0 \circ S^0.$$

On suppose de plus que **tous** les composants de ce schéma sont paramétrés par une clé, et donc **inconnus**. Par conséquent, un attaquant modélise toutes les boîte S comme des applications inversibles aléatoires, et toutes les boîtes A par des applications affines aléatoires.

Explication du Design Une telle construction reçoit de temps en temps de l'intérêt pour former ce que l'on appelle un « cryptosystème en boîte blanche ». L'idée est d'offrir les mêmes possibilités que la cryptographie à clé publique en permettant à tout le monde de chiffrer, mais sans autoriser le déchiffrement, tout en ayant les performances de la cryptographie symétrique. On peut imaginer par exemple que F_{SASAS} est implémentée sur un circuit matériel que l'attaquant possède, qu'il peut utiliser pour chiffrer, mais n'a pas accès aux différents composants qui sont nécessaires pour déchiffrer.

On rappelle qu'une application A sur 128 bits est affine si elle envoie des éléments de $x \in \mathbb{F}_2^{128}$ en

$$A(x) = L(x) + B,$$

où $L\colon \mathbb{F}_2^{128} \to \mathbb{F}_2^{128}$ est une application linéaire, et $B \in \mathbb{F}_2^{128}$ est une constante.
De même, un sous-espace affine E de $\mathbb{F}_2^n$ est un sous-ensemble de la forme

$$E = \{x + b \mid x \in V\} = V + b$$

où V est un sous-espace vectoriel de $\mathbb{F}_2^n$ et b est une constante. On rappelle que V est entièrement caractérisé par E. La dimension de E est par définition la dimension de V.

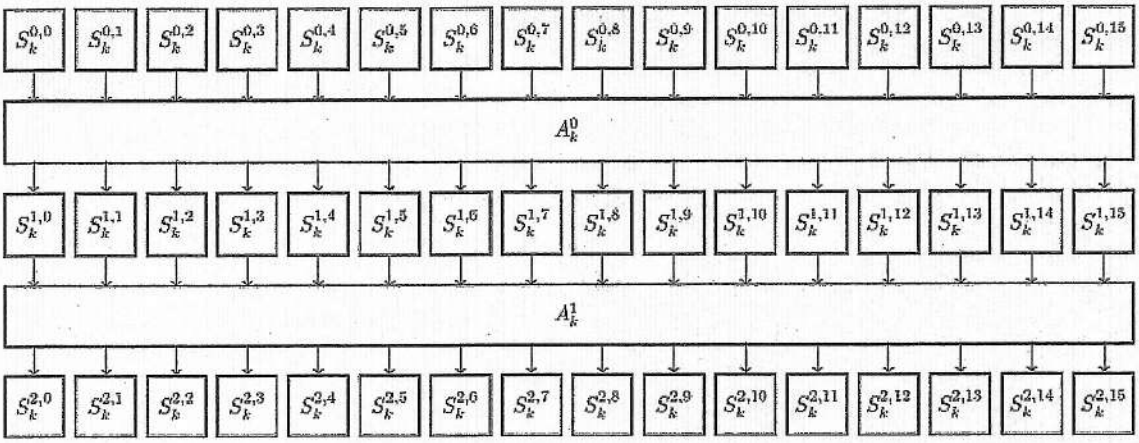

FIGURE 2 – Exemple de chiffrement par bloc de type SASAS.

Objectif Le but de ce problème est de montrer qu'un chiffrement par blocs possédant cette structure n'est en réalité absolument pas sûr. Pour cela, on va monter une attaque algébrique à clairs choisis de complexité extrêmement faible pour retrouver toutes les boîtes S (ou en tout cas des boîtes S équivalentes) de la dernière couche alors mêmes qu'elles ont été supposées différentes et aléatoires. Notons que dans notre modèle d'attaque, un attaquant peut chiffrer les messages de son choix, donc accéder à ces clairs choisis est automatique.

4.1 L'Algèbre des Fonctions Booléennes

Soit I_n l'idéal de l'anneau $\mathbb{F}_2[X_1, \ldots, X_n]$ engendré par les $X_i^2 - X_i$ pour $1 \leq i \leq n$. Soit $\mathbb{B}_n$ le quotient $\mathbb{F}_2[X_1, \ldots, X_n]/I_n$. On identifie tous les éléments de $\mathbb{B}_n$ avec l'unique représentant dans la classe d'équivalence qui ne contient pas de carrés, c'est-à-dire dont tous les monômes sont de la forme $\prod_{i \in S} X_i$ pour un certain sous-ensemble $S \subset \{1, \ldots, n\}$ de variables.

(Q17) Soit $P \in \mathbb{B}_n$. Montrez que la fonction $f_P \colon \mathbb{F}_2^n \to \mathbb{F}_2$ qui à $(x_1, \ldots, x_n)$ associe $P[X_1 = x_1, \ldots, X_n = x_n]$ est bien définie, c'est-à-dire qu'elle ne dépend pas du choix du représentant.

(Q18) Rappelez pourquoi toute fonction $\mathbb{F}_2^n \to \mathbb{F}_2$ est polynomiale.

(Q19) En déduire qu'il y a une bijection entre les fonctions booléennes $\mathbb{F}_2^n \to \mathbb{F}_2$ et l'anneau $\mathbb{B}_n$.

Dans la suite, on identifie alors fonctions Booléennes et polynômes. Le degré d'une fonction vectorielle $\mathbb{F}_2^n \to \mathbb{F}_2^m$ est le maximum des degrés de chaque bit de sortie (en tant que polynôme sans carré).

(Q20) Soit $M \stackrel{\text{def}}{=} X_{i_1} \cdots X_{i_k}$ un monôme de degré ℓ. Montrez que $M^{-1}\{1\}$ est une intersection de ℓ hyperplans affines.

(Q21) Soit E un sous-espace affine de $\mathbb{F}_2^n$ de dimension $k > 0$, et soit P un polynôme de degré $k - 1$. Montrez que

$$\sum_{x \in E} P(x) = 0.$$

Indication

Remarquez que cette somme est une somme dans $\mathbb{F}_2$, et commencez par traiter le cas où P est un monôme.

(Q22) Soit P un polynôme de degré n. Montrez que

$$\sum_{x \in \mathbb{F}_2^n} P(x) = 1.$$

(Q23) En déduire qu'une fonction $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ bijective est de degré au-plus $n - 1$.

4.2 L'attaque

Pour $a \in \mathbb{F}_2^{120}$, on note $\mathcal{M}_a \stackrel{\text{def}}{=} \{a||x : x \in \mathbb{F}_2^8\}$ l'ensemble des messages de prefixe a. Soit $F_{ASAS} \stackrel{\text{def}}{=} A^1 \circ S^1 \circ A^0 \circ S^0$ le chiffrement sans la dernière couche S et soit $a \in \mathbb{F}_2^{120}$.

(Q24) Soit $\mathcal{E}_a \stackrel{\text{def}}{=} \mathcal{M}_a$. Montrez que $\mathcal{E}_a$ est un espace affine. Quelle est sa dimension ?

(Q25) Quel est le degré d'une des boîtes $S_k^{i,j}$?

(Q26) Montrez que $A^1 \circ S^1 \circ A^0$ est de degré au plus 7.

(Q27) Montrez que

$$\sum_{x \in \mathcal{M}_a} F_{ASAS}(x) = 0,$$

où cette somme est bien à valeurs dans $\mathbb{F}_2^{128}$.

On se fixe à présent sur une seule boîte S de la dernière couche. Par exemple, $S_k^{2,0}$. Soit T son inverse, et soit $X_z \stackrel{\text{def}}{=} T(z)$ pour $z \in \mathbb{F}_2^8$ qu'on identifie à une inconnue, puisque l'on ne connaît pas $S_k^{2,0}$. En particulier, notre objectif est de déterminer les valeurs de X_z pour tout $z \in \mathbb{F}_2^8$, ce qui détermine entièrement la boîte S.

(Q28) Déduire de la question précédente une équation linéaire reliant tous les X_z.

(Q29) Expliquez comment obtenir un système linéaire déterminant entièrement les X_z.

(Q30) Conclure en donnant un algorithme permettant de retrouver toutes les boîtes S de la dernière couche. Quelle est sa complexité ?

(Q31) Cet algorithme retrouve-t-il nécessairement les boîtes S du schéma original ?