TD7 - Révisions Cryptanalyses Linéaire et Différentielle (Solutions)

Responsable : M. Bombar

Cette semaine, je vous propose un TD de révision, sans machine, pour souffler un peu, et en prévision de l'examen (pas de panique, on n'est qu'à la moitié du cours !).

1 Cryptanalyse Différentielle de FEAL - Examen 2021

1.1 Notations

Dans cet exercice, si $x \in \mathbb{F}_2^n$ et $y \in \mathbb{F}_2^m$ sont deux chaînes de bits, représentés par des vecteurs binaires, alors $x \,||\, y \in \mathbb{F}_2^{n+m}$ est le vecteur binaire représentant leur **concaténation**.

Par ailleurs, les boîtes S de FEAL agissant sur des octets parfois vus comme des éléments de $\mathbb{Z}/256\mathbb{Z}$, ou bien comme des vecteurs de 8 bits dans $\mathbb{F}_2^8$, on notera $\oplus$ l'addition dans l'espace vectoriel $\mathbb{F}_2^8$ (donc le XOR bit à bit), et $\boxed{+}$ l'addition modulo 256.

Enfin, pour n un entier non nul, on notera $\mathbf{0}_n \in \mathbb{F}_2^n$ pour désigner le vecteur nul de longueur n. Par exemple $\mathbf{0}_3 = 000$.

1.2 Le chiffrement FEAL

FEAL (*Fast Data Encipherment Algorithm*) est un chiffrement par blocs de type Feistel à 4 tours qui utilise des clés de 64 bits pour chiffrer des blocs de 64 bits. Il a été proposé par A. Shimizu et S. Miyaguchi en 1987 dans le but de remplacer le chiffrement DES et ses clés de 56 bits.

Les clés. FEAL utilise 2 clés K_0 et K_5 de 64 bits, et 4 sous clés de 16 bits K_1, K_2, K_3, K_4.

Les boîtes S. FEAL utilise deux boîtes S différentes notées S_0 et S_1 prenant en entrée 16 bits représentés comme une paire d'entiers dans $\{0, \ldots, 255\}$ et produisant 8 bits en sortie. Elles sont définies comme suit :

$$S_i(x, y) = (x \boxed{+} y \boxed{+} i \mod 256) <<< 2.$$

où $<<< 2$ désigne une rotation de 2 bits vers la gauche (de façon cyclique).

La fonction de tour. FEAL utilise 4 tours de type Feistel dont les fonctions de tours, notées F_{K_i} pour $1 \leq i \leq 4$, prennent en entrée 32 bits et ressortent 32 bits. Soit M un bloc de 64 bits à chiffrer. On pose $X_0 \overset{\text{def}}{=} M \oplus K_0 = (L_0 \,||\, R_0)$, où L_0 (resp. R_0) désigne les 32

bits les plus à gauche (resp. les plus à droite) de X_0. On rappelle que dans un schéma de Feistel en notant pour $1 \leq i \leq 4$ $X_i = (L_i \,||\, R_i)$ les 32 bits en sortie du tour i, on a

$$L_i = R_{i-1}, \qquad R_i = L_{i-1} \oplus F_{K_i}(R_{i-1}).$$

Le chiffré est alors

$$C = (R_4 \,||\, L_4) \oplus K_5.$$

Pour $X = (x_0 \,||\, x_1 \,||\, x_2 \,||\, x_3)$ de 32 bits, vu comme la concaténation de 4 octets, et $K = (K^L \,||\, K^R)$ une clé de 16 bits, la fonction $F_K(X)$ est définie par

$$F_K(X) \stackrel{\text{def}}{=} S_0(x_0, u) \,||\, u \,||\, v \,||\, S_1(x_3, v),$$

avec

$$u = S_1(x_0 \oplus x_1 \oplus K^L, x_2 \oplus x_3 \oplus K^R)$$

et

$$v = S_0(x_2 \oplus x_3 \oplus K^R, u).$$

(Q1) Montrer que pour tout $(x, y) \in \mathbb{F}_2^8 \times \mathbb{F}_2^8$, on a

$$S_0(x \oplus 1000\ 0000, y) = S_0(x, y) \oplus 0000\ 0010.$$

(Q2) Soient $M, M^\star \in \mathbb{F}_2^{64}$ deux messages clairs tels que

$$M \oplus M^\star = 1000\ 0000\ 1000\ 0000\ \mathbf{0}_{48}.$$

On note $(L_2 \,||\, R_2)$ (resp. $(L_2^\star \,||\, R_2^\star)$) l'entrée du troisième tour lors du chiffrement de M (resp. de $M^\star$). Que vaut la différence

$$(L_2 \,||\, R_2) \oplus (L_2^\star \,||\, R_2^\star)?$$

(Q3) Soit $K = K^L \,||\, K^R$ une sous clé de 16 bits, avec K^L et K^R sur 8 bits. Montrer que pour tout X de 32 bits, on a

$$F_K(X) = F_{\mathbf{0}_{16}}(X \oplus (0000\ 0000 \,||\, K^L \,||\, K^R \,||\, 0000\ 0000)).$$

(Q4) On note $K_5 \stackrel{\text{def}}{=} K_5^L \,||\, K_5^R$ avec K_5^L, K_5^R de 32 bits, et de même $K_4 = K_4^L \,||\, K_4^L$ avec K_4^L, K_4^R de 8 bits.
Déduire des deux questions précédentes une attaque utilisant 2 clairs choisis et permettant de retrouver la valeur de $K_5^R \oplus (0000\ 0000 \,||\, K_4^L \,||\, K_4^R \,||\, 0000\ 0000)$ en évaluant un certain nombre de fois N la fonction de tour F. Quelle est la valeur de N ?
Indication : Raisonner autour de $F_{K_4}(L_4) \oplus F_{K_4}(L_4^\star)$.

Solutions

1. Tout d'abord, remarquons que par $\mathbb{F}_2$ linéarité de l'opération $<<$, on a

$$S_0(x,y) \oplus S_0(z,t) = \Big((x \boxed{+} y) \oplus (z \boxed{+} t)\Big) << 2$$

On écrit $x = x_7x_6x_5x_4x_3x_2x_1x_0$. Alors,

$$x \oplus 1000\ 0000 = \bar{x_7}x_6x_5x_4x_3x_2x_1x_0$$

donc la différence $(x \boxed{+} y) \oplus (x \oplus 1000\ 0000 \boxed{+} y)$ est nulle sur tous les octets x_i pour $0 \leq i \leq 6$. Par ailleurs, quelle que soit la valeur de la retenue, celle-ci n'affecte pas la somme modulo 256. Par conséquent, on a

$$(x \boxed{+} y) \oplus (x \oplus 1000\ 0000 \boxed{+} y) = 1000\ 0000$$

d'où après rotation de 2 bits vers la gauche

$$S_0(x,y) \oplus S_0(x \oplus 1000\ 0000, y) = 0000\ 0010.$$

2. On note $M = (L_0, R_0)$ et $M^\star = (L_0^\star, R_0^\star)$. Par hypothèse

$$L_0 \oplus L_0^\star = 1000\ 0000\ 1000\ 0000\ \mathbf{0}_{16}$$

et

$$R_0 \oplus R_0^\star = \mathbf{0}_{32}.$$

Par ailleurs, après deux tours de Feistel, on a

$$L_2 = L_0 \oplus F_{K_1}(R_0).$$

d'où

$$L_2 \oplus L_2^\star = L_0 \oplus L_0^\star \oplus F_{K_1}(R_0) \oplus F_{K_1}(R_0^\star).$$

Or, on vient de rappeler que R_0 avait une différentielle nulle. Autrement dit, $R_0 = R_0^\star$, et donc $F_{K_1}(R_0) = F_{K_1}(R_0^\star)$.
Finalement,

$$L_2 \oplus L_2^\star = L_0 \oplus L_0^\star = 1000\ 0000\ 1000\ 0000\ \mathbf{0}_{16}.$$

Solutions

De l'autre côté :

$$\begin{aligned} R_2 \oplus R_2^\star &= \Big(R_0 \oplus F_{K_2}(R_1)\Big) \oplus \Big(R_0^\star \oplus F_{K_2}(R_1^\star)\Big) \\ &= F_{K_2}(R_1) \oplus F_{K_2}(R_1^\star) \end{aligned}$$

Il faut alors regarder plus précisément l'expression de

$$F_{K_2}(R_1) = (S_0(x_0, u) \,||\, u \,||\, v \,||\, S_1(x_3, v)).$$

où $R_1 = x_0 \,||\, x_1 \,||\, x_2 \,||\, x_3$, et remarquons que

$$\begin{aligned} R_1 \oplus R_1^\star &= \Big(L_0 \oplus F_{K_1}(R_0)\Big) \oplus \Big(L_0^\star \oplus F_{K_2}(R_0^\star)\Big) \\ &= L_0 \oplus L_0^\star \\ &= (1000\ 0000) \,||\, (1000\ 0000) \,||\, (0000\ 0000) \,||\, (0000\ 0000). \end{aligned}$$

puisque $R_0 \oplus R_0^\star = 0$.
En particulier,

$$(x_2 \oplus x_2^\star) \oplus (x_3 \oplus x_3^\star) = 0$$

et

$$(x_0 \oplus x_0^\star) \oplus (x_1 \oplus x_1^\star) = 1000\ 0000 \oplus 1000\ 0000 = 0.$$

Par conséquent,

$$u \stackrel{\text{def}}{=} S_1(x_0 \oplus x_1 \oplus K^L, x_2 \oplus x_3 \oplus K^R)$$

a une différentielle nulle, et de même pour v, puis (x_3, v). Finalement, le seul élément qui a une différentielle non nulle dans F_{K_2} est $S_0(x_0, u)$ qui a une différentielle de 0000 0010 d'après la question $Q1$ et puisque $x_0 \oplus x_0 = 1000\ 0000$.
On en déduit que

$$(L_2 \,||\, R_2) \oplus (L_2^\star \,||\, R_2^\star) = (1000\ 0000\ 1000\ 0000\ \mathbf{0}_{16} \,||\, 0000\ 0010\ \mathbf{0}_{24}).$$

Solutions

(Q3) Simple calcul.

Solutions

En sortie du 4ème tour d'un schéma de Feistel, l'état interne est de la forme

$$(L_4 \,||\, R_4) = (L_4 \,||\, L_3 \oplus F_4(R_3))$$

mais $R_3 = L_4$ et $L_3 = R_2$ par construction. On en déduit donc :

$$R_4 = R_2 \oplus F_4(L_4).$$

Le chiffré étant

$$(C_L, C_R) = (R_4 \oplus K_5^L \,||\, L_4 \oplus K_5^R)$$

on en déduit que

$$R_4 = C_L \oplus K_5^L$$

et par conséquent

$$F_4(L_4) = C_L \oplus K_5^L \oplus R_2.$$

On en déduit

$$\begin{aligned} F_4(L_4) \oplus F_4(L_4^\star) &= \left(C_L \oplus C_L^\star\right) \oplus (K_5^L \oplus K_5^L) \oplus (R_2 \oplus R_2^\star) \\ &= \left(C_L \oplus C_L^\star\right) \oplus (0000\ 0010\ \mathbf{0}_{24}). \end{aligned}$$

D'un autre côté,

$$F_4(L_4) = F_{\mathbf{0}_{16}}(L_4 \oplus (0000\ 0000 \,||\, K_4^L \,||\, K_4^R \,||\, 0000\ 0000))$$

mais

$$L_4 = C_R \oplus K_5^R.$$

D'où finalement, pour un couple de messages $M, M^\star$ de différence

$$M \oplus M^\star = 1000\ 0000\ 1000\ 0000\ \mathbf{0}_{48},$$

on a

$$\begin{aligned} &F_{\mathbf{0}_{16}}\left(C_R \oplus K_5^R \oplus (0000\ 0000 \,||\, K_4^L \,||\, K_4^R \,||\, 0000\ 0000)\right) \\ \oplus &F_{\mathbf{0}_{16}}\left(C_R^\star \oplus K_5^R \oplus (0000\ 0000 \,||\, K_4^L \,||\, K_4^R \,||\, 0000\ 0000)\right) \\ &= \left(C_L \oplus C_L^\star\right) \oplus (0000\ 0010\ \mathbf{0}_{24}). \end{aligned}$$

qui est une équation ne dépendant **que** des chiffrés et de la valeur recherchée.

On peut alors montrer qu'en itérant l'attaque précédente sur tous les tours, on peut retrouver une clé équivalente de FEAL en 2^{35} évaluations de la fonction F, à l'aide de plusieurs propriétés différentielles.

2 Cryptanalyse Linéaire de SAFER - Examen 2023

SAFER (*Secure And Fast Encryption Routine*) est un chiffrement par blocs de type réseau de substitutions-permutations (SPN) proposé par J. Massey en 1993 qui utiliser une clé de 64 bits pour chiffrer des blocs de 64 bits. Puisqu'il s'agit d'un SPN, l'état interne est aussi de 64 bits, vus comme 8 octets, chacun identifiés à un élément de $\mathbb{Z}/256\mathbb{Z}$ identifié à l'ensemble $I \stackrel{\text{def}}{=} \{0, \ldots, 255\}$ muni de l'addition modulo 256 notée $\boxed{+}$ comme dans l'exercice précédent. Si x est un bloc de 64 bits, on note $x^{(1)}, x^{(2)}, \ldots, x^{(8)}$ ces 8 octets. Le tour i avec $1 \leq i \leq 6$ fait intervenir deux sous clés k_{2i-1} et k_{2i}, de 64 bits chacune. On suppose que ces clés sont obtenues à partir d'une clé maître $k = (k^{(1)}, \ldots, k^{(8)})$ de sorte que $k_{2i-1}^{(j)}$ et $k_{2i}^{(j)}$ ne dépendent que de $k^{(j)}$.

L'addition des clés avec l'état interne se fait au niveau de chacun des 8 octets en utilisant soit l'addition $\boxed{+}$ modulo 256, soit le xor bit à bit $\oplus$. On désigne par S une permutation fixée non linéaire sur 8 bits, c'est-à-dire une bijection de I, et on note L la fonction

$$L\colon \begin{cases} I \times I & \to & I \times I \\ (a,b) & \mapsto & (L_1(a,b), L_2(a,b)) \stackrel{\text{def}}{=} (2a \boxed{+} b \mod 256, a \boxed{+} b \mod 256). \end{cases}$$

Cette fonction prend donc deux octets en entrée et ressort deux octets en sortie. on la schématise par

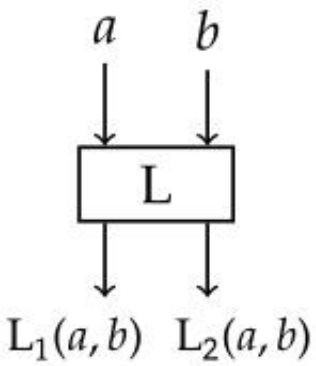

FIGURE 1 – Représentation de la boîte L de SAFER.

Let tour i transforme un bloc de 64 bits x_i en un bloc de 64 bits x_{i+1} par le schéma de la Figure 2.

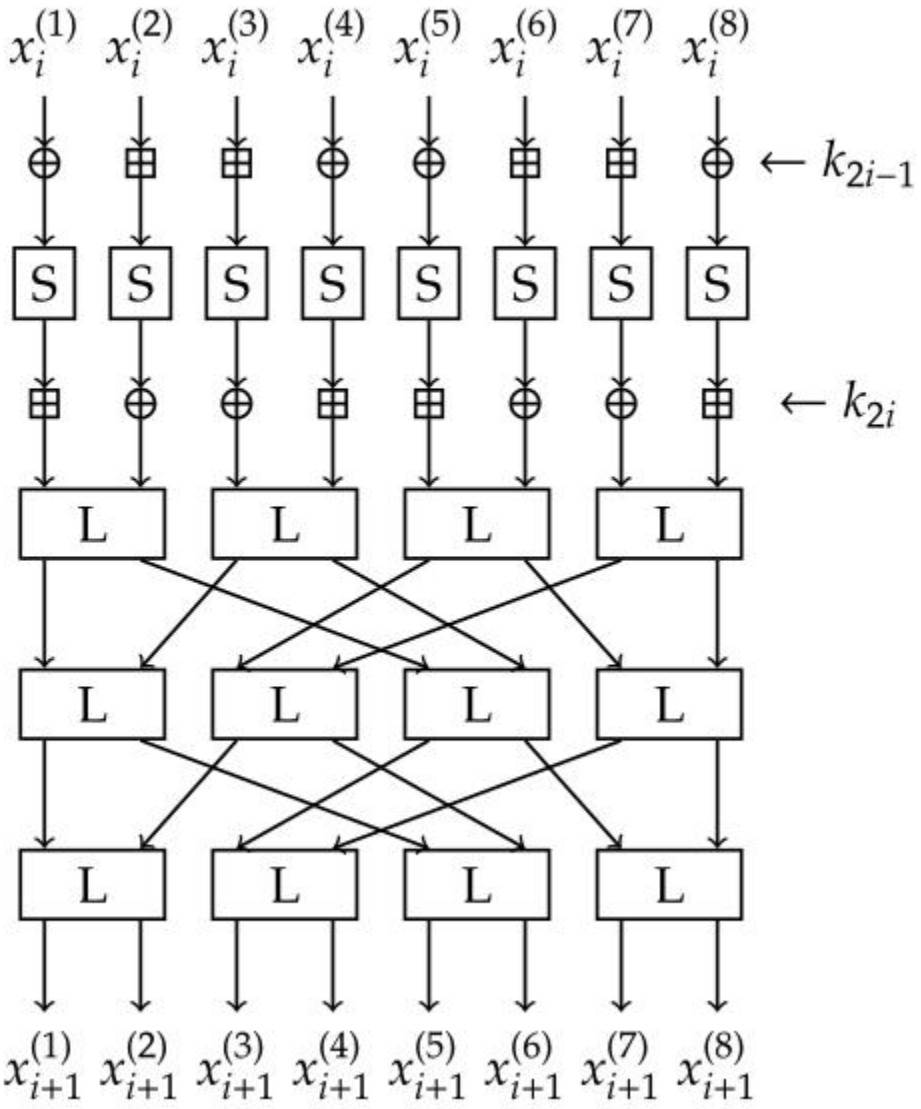

FIGURE 2 – Fonction de tour de SAFER.

On note u_i le bloc de 64 bits après l'application des fonctions S, c'est-à-dire

— $u_i^{(1)} = S(x_i^{(1)} \oplus k_{2i-1}^{(1)})$

— $u_i^{(2)} = S(x_i^{(2)} \boxed{+} k_{2i-1}^{(2)})$
— ...

De même, on note z_i le bloc de 64 bits après l'ajout de la clé k_{2i}, c'est-à-dire
— $z_i^{(1)} = k_{2i}^{(1)} \boxed{+} u_i^1$
— $z_i^{(2)} = k_{2i}^{(2)} \oplus u_i^1$
— ...

Le chiffrement complet prend en entrée un message clair $m = x_1$ de 64 bits et retourne un chiffré c de 64 bits obtenu en effectuant 6 tours comme la Figure 2 avec des clés $k_1, k_2, \ldots, k_{11}, k_{12}$, puis on ajoute à x_7 une clé k_{13} de la même façon que les clés d'indice impair k_{2i-1} sont ajoutées pour le tour i : $c^{(1)} = x_7^{(1)} \oplus k_{13}^{(1)}, c^{(2)} = x_7^{(2)} \boxed{+} k_{13}^{(2)}, \ldots$.

(Q5) Du point de vue de la sécurité, quel est le but des fonctions S ? Quel est celui des trois rangées de fonctions L ?

Solutions

Les boîtes S apportent la confusion, alors que les boîtes L qui sont linéaires, apportent la diffusion.

(Q6) Montrez que pour $1 \leq i \leq 6$ on a

$$z_i^{(3)} + z_i^{(4)} \equiv x_{i+1}^{(3)} + x_{i+1}^{(4)} \mod 2 \quad (1)$$

Solutions

On a

$$(x_{i+1}^{(3)}, x_{i+1}^{(4)}) = L(L_1(L_2(z_i^{(1)}, z_i^{(2)}), L_2(z_i^{(3)}, z_i^{(4)}), L_1(L_2(z_i^{(5)}, z_i^{(6)}), L_2(z_i^{(7)}, z_i^{(8)})))$$

Or, $L_1(a,b) = 2a \boxed{+} b \mod 256 \equiv b \mod 2$, d'où

$$(x_{i+1}^{(3)}, x_{i+1}^{(4)}) \equiv L(L_2(z_i^{(3)}, z_i^{(4)}), L_2(z_i^{(7)}, z_i^{(8)}) \mod 2$$

Par ailleurs,

$$L_1(a,b) + L_2(a,b) = 3a + 2b \mod 256 \equiv a \mod 2$$

d'où finalement

$$\begin{aligned} x_{i+1}^{(3)} + x_{i+1}^{(4)} &\equiv L_2(z_i^{(3)}, z_i^{(4)}) \mod 2 \\ &\equiv (z_i^{(3)} + z_i^{(4)} \mod 256) \mod 2 \\ &\equiv z_i^{(3)} + z_i^{(4)} \mod 2. \end{aligned}$$

Dans les deux questions suivantes, on suppose que

$$\mathbb{P}_a\Big(a \equiv S(a) \mod 2\Big) = \frac{1}{2}(1+\varepsilon), \qquad 0 \leq \varepsilon \leq 1.$$

(Q7) Pour $1 \leq i \leq 6$ et $j \in \{3,4\}$, on note $\kappa_i^{(j)} \stackrel{\text{def}}{=} k_{2i-1}^{(j)} + k_{2i}^{(j)}$. Quelle est la probabilité (sur x_i) que

$$x_i^{(3)} + x_i^{(4)} + z_i^{(3)} + z_i^{(4)} \equiv \kappa_i^{(4)} + \kappa_i^{(4)} \mod 2$$

Solutions

Commençons par remarquer que quel que soit l'opération d'ajout de clé ($\oplus$ ou $\boxed{+}$), lorsqu'on les réduit modulo 2, elles sont toutes les deux équivalentes à l'addition de la clé modulo 2.

Notons $y_i^{(j)} \stackrel{\text{def}}{=} x_i^{(j)} + k_{2i-1}^{(j)}$ pour $1 \leq i \leq 6$ et $j \in \{3,4\}$. Considérons les variables aléatoires

$$Y_i^j \stackrel{\text{def}}{=} \Big(S(y_i^{(j)}) - y_i^j \mod 2 \Big).$$

Ce sont des fonctions **déterministes** des $x_i^{(j)}$. En particulier, ce sont des variables aléatoires à valeurs dans $\mathbb{F}_2$, et **mutuellement indépendantes** s'il en était de même pour les $x_i^{(j)}$, ce qu'on suppose pour la suite. Par ailleurs, par hypothèse sur la boîte S, on a

$$\mathbb{P}(Y_i^{(j)} = 0) = \frac{1}{2}(1-\varepsilon).$$

En particulier, on a la relation $u_i^{(j)} \equiv Y_i^{(j)} + x_i^{(j)} + k_{2i-1}^{(j)} \mod 2$. D'où

$$z_i^{(j)} = k_{2i}^{(j)} + u_i^{(j)} \equiv k_{2i}^{(j)} + k_{2i-1}^{(j)} + x_i^{(j)} + Y_i^{(j)} \mod 2$$

où encore

$$z_i^{(3)} + z_i^{(4)} + x_i^{(3)} + x_i^{(4)} \equiv \kappa_i^{(3)} + \kappa_i^{(4)} + Y_i^{(3)} + Y_i^{(4)}.$$

Puisque les $Y_i^{(j)}$ sont indépendantes, on en déduit d'après le **piling-up** lemma que

$$z_i^{(3)} + z_i^{(4)} + x_i^{(3)} + x_i^{(4)} \equiv \kappa_i^{(3)} + \kappa_i^{(4)} \tag{2}$$

avec probabilité $\frac{1}{2}\left(1+\varepsilon^2\right)$.

(Q8) Soit $m = (m^{(1)}, \ldots, m^{(8)})$ un bloc clair de 64 bits, et soit $c = (c^{(1)}, \ldots, c^{(8)})$ le chiffré correspondant. On note $\kappa \stackrel{\text{def}}{=} \sum_{i=1}^{13} \kappa_i^{(3)} + \kappa_i^{(4)}$. Déterminer une relation existant entre $m^{(3)}, m^{(4)}, c^{(3)}, c^{(4)}, \kappa$ avec probabilité > 0, et donnez sa probabilité (où l'aléa est sur m).

(Q9) En déduire une attaque à clairs connus permettant de retrouver de l'information sur la clé de ce chiffrement.

Solutions

En combinant les équations 1 et 2, on en déduit qu'à chaque tour

$$x_{i+1}^{(3)} + x_{i+1}^{(4)} + x_i^{(3)} + x_i^{(4)} = \kappa_i^{(3)} + \kappa_i^{(4)}$$

avec probabilité $\frac{1}{2}\left(1-\varepsilon^2\right)$. Ainsi, sur plusieurs tours on en déduit (en sommant sur tous les tours, et en utilisant l'hypothèse d'indépendance ainsi que le piling-up lemma) que

$$\mathbb{P}_m\Big(c^{(3)} + c^{(4)} + m^{(3)} + m^{(4)} = \kappa\Big) = \frac{1}{2}\left(1 \pm \varepsilon^{12}\right).$$

On peut alors monter une attaque par cryptanalyse linéaire dès lors que $\varepsilon >> 0$:

- Étant donnée une collection de $n = \Omega\left(\frac{1}{\varepsilon^{24}}\right)$ couples clairs/chiffrés
- Pour tous les couples d'octets possibles (k_3, k_4) pour la clé, un attaquant va initialiser un compteur $c(k_3, k_4)$ et calculer le nombre de couples qui vérifient la relation.
- Si ce compteur $c(k_3, k_4)$ est significativement plus grand ou significativement plus petit que les autres valeurs, alors c'est que (k_3, k_4) sont vraissemblablement les deux octets recherchés de la clé.

Remarque 1. *En réalité, l'équation 1 n'est pas la seule relation linéaire entre les entrées et sorties de la couche de boîtes L. Par exemple*

- $x_{i+1}^{(2)} + x_{i+1}^{(6)} \equiv z_i^{(2)} + z_i^{(6)} \mod 2$
- $x_{i+1}^{(5)} + x_{i+1}^{(7)} \equiv z_i^{(5)} + z_i^{(7)} \mod 2$
- $x_{i+1}^{(3)} + x_{i+1}^{(7)} \equiv z_i^{(5)} + z_i^{(6)} \mod 2$
- $x_{i+1}^{(5)} + x_{i+1}^{(6)} \equiv z_i^{(2)} + z_i^{(4)} \mod 2$
- $x_{i+1}^{(2)} + x_{i+1}^{(4)} \equiv z_i^{(3)} + z_i^{(7)} \mod 2$

Remarque 2. *En pratique, le choix de la permutation S était tel que $\varepsilon = 0$ et donc cette attaque ne s'applique pas.*