Attaque Linéaire

I. Définition :

La cryptanalyse linéaire est une technique inventée par Mitsuru Matsui, chercheur chez Mitsubishi Electric. Elle date de 1993 et fut développée à l'origine pour casser l'algorithme de chiffrement symétrique DES. Ce type de cryptanalyse se base sur un concept antérieur à la découverte de Matsui : les expressions linéaires probabilistes. Ces dernières ont été étudiées par Henri Gilbert et Anne Tardy-Corfdir dans le cadre d'une attaque sur FEAL.

La cryptanalyse linéaire est plus efficace que la cryptanalyse différentielle, mais moins pratique pour la simple et bonne raison que l'on part du principe que l'attaquant ne dispose pas de la boîte noire symbolisant l'algorithme de chiffrement, et qu'il ne peut pas soumettre ses propres textes.

La cryptanalyse linéaire consiste à faire une approximation linéaire de l'algorithme de chiffrement en le simplifiant. En augmentant le nombre de couples disponibles, on améliore la précision de l'approximation et on peut en extraire la clé. Tous les nouveaux algorithmes de chiffrement doivent veiller à être résistants à ce type d'attaque.

DES n'était pas conçu pour empêcher ce genre de méthode, les tables de substitution (S-Boxes) présentent en effet certaines propriétés linéaires, alors qu'elles étaient justement prévues pour ajouter une non-linéarité à DES.

Elle a par la suite été appliquée avec succès sur plusieurs algorithmes comme LOKI, FEAL ou une version simplifiée de Serpent. Les algorithmes plus récents comme AES (Rijndael), IDEA, et bien d'autres encore, sont insensibles à une attaque linéaire. La complexité de l'attaque est dans ces cas largement supérieure à celle d'une recherche exhaustive [1].

II. Principe :

Le principe général de cette attaque se base sur les équations linéaires binaires (booléennes).

Equations linéaires :

Une expression linéaire (équation linéaire) est une expression qui s'écrit :

$X_1 \oplus X_2 \oplus .. \oplus X_n = Y_1 \oplus Y_2 \oplus .. \oplus Y_n$ avec le symbole $\oplus$ signifie le Ou exclusif (XOR), et $X_1,..X_n$, $Y_1,..Y_n$ sont des variables booléennes (qui peuvent avoir uniquement les valeurs 0 ou 1).

Cette équation peut être écrite : $X_1 \oplus X_2 \oplus .. \oplus X_n \oplus Y_1 \oplus Y_2 \oplus .. \oplus Y_n = 0$.

Les tables de substitution (S-Box) :

Une table de substitution prend en général une variable de m bits en entrée et produit une sortie de n bits, les entrées et les sorties n'ont pas forcément la même taille. Elles sont utilisées dans les chiffres symétriques et sont conçues généralement pour être non linéaires, cependant la combinaison entre quelques entrées avec quelques sorties peut exprimer une certaine linéarité [1][2].

Si ***S*** est une fonction de substitution, ainsi si $\boldsymbol{Y = S(X) \Rightarrow X = S^{-1}(Y).}$

Exemple :

Soit une table de substitution S représentée par le tableau ci-dessous :

Input	000	001	010	011	100	101	110	111
Output	010	100	000	111	001	110	101	011

Cette boite prend en entrée un nombre hexadécimal composé de 3 bits $X_1X_2X_3$ et fournie en sortie un autre nombre hexadécimal constitué de 3 bits aussi $Y_1Y_2Y_3$.

Soient les deux équations linéaires suivantes :

- $X_1 \oplus X_2 \oplus X_3 = Y_1 \oplus Y_2$ (1)
- $X_2 \oplus X_3 = Y_3$ (2)

La figure ci-dessous donne les différents cas pour lesquels les deux équations sont satisfaites :

Première équation

X	Y	$X_1 \oplus X_2 \oplus X_3$	$Y_1 \oplus Y_2$
000	010	0	1
001	100	1	1
010	000	1	0
011	111	0	0
100	001	1	0
101	110	0	0
110	101	0	1
111	011	1	1

Deuxième équation

X	Y	$X_2 \oplus X_3$	Y_3
000	010	0	0
001	100	1	0
010	000	1	0
011	111	0	1
100	001	0	1
101	110	1	0
110	101	1	1
111	011	0	1

On remarque que la probabilité de satisfaction de la 1ère équation est 4/8, et pour la 2ème équation est égale à 2/8.

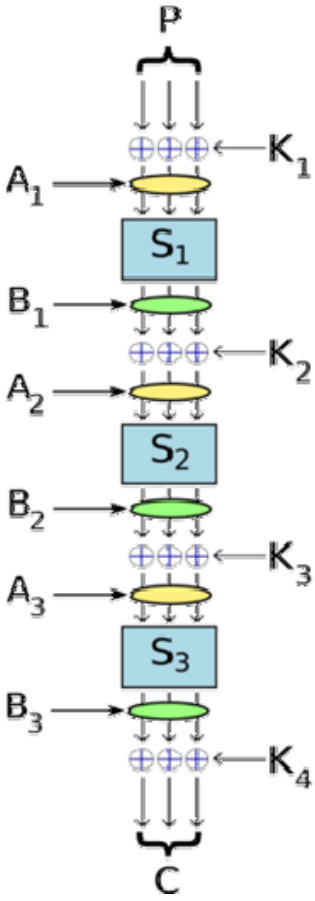
L'approche de cryptanalyse linéaire consiste à chercher des approximations qui ont des probabilités d'occurrence très élevées ou bien très faibles.

Il faut noter qu'il ne devrait y avoir des approximations définies sur l'ensemble des entrées et sorties à la fois avec une haute ou faible probabilité d'occurrence, sinon l'algorithme de chiffrement est considéré comme trivialement faible.

III. Exemple d'application de la cryptanalyse linéaire :

Considérons un algorithme de chiffrement très simple qui prend 3 bit en entrée et donne 3 bit chiffrés en sortie. Le processus se déroule sur 3 tours et utilise 4 sous-clés [2].

Soit P la donnée en clair de 3 bits et soit le résultat final C chiffré de 3 bits.



Tour1 :

Le texte P est chiffré avec la sous-clé K_1 (XOR), on obtient le texte A_1.

- $A_1 = P \oplus K_1$

Le résultat passe dans une table de substitution S_1 :

- $B_1 = S_1(A_1)$.

Tour2 :

Mixage du résultat du 1er tour par la sous-clé K_2, puis substitution par la table S2. On obtient ainsi :

- $A_2 = B_1 \oplus K_2$
- $B_2 = S_2(A_2)$.

Tour3 :

Le même processus est appliqué et on obtient :

- $A_3 = B_2 \oplus K_3$
- $B_3 = S_3(A_3)$.

Finalisation :

A la fin un dernier mixage est appliqué avec la sous-clé K_4.

- $C = B_3 \oplus K_4$

On suppose avoir les approximations suivantes :

- $S_1 : X_1 \oplus X_2 \oplus X_3 = Y_2$.
- $S_2 : X_2 = Y_1 \oplus Y_3$.

1er tour :

- $A_1 = P \oplus K_1$ donc $A_1 = [A_{1,1} = P_{1,1} \oplus K_{1,1}, A_{1,2} = P_{1,2} \oplus K_{1,2}, A_{1,3} = P_{1,3} \oplus K_{1,3}]$.
- $B_1 = S_1(A_1) \Rightarrow B_{1,2} = A_{1,1} \oplus A_{1,2} \oplus A_{1,3}$

  $\Rightarrow B_{1,2} = (P_{1,1} \oplus K_{1,1}) \oplus (P_{1,2} \oplus K_{1,2}) \oplus (P_{1,3} \oplus K_{1,3})$. ***(I)***

2ème tour :

- $A_2 = B_1 \oplus K_2$, donc $A_2 = [A_{2,1} = B_{1,1} \oplus K_{2,1}, A_{2,2} = B_{1,2} \oplus K_{2,2}, A_{2,3} = B_{1,3} \oplus K_{2,3}]$.
- $B_2 = S_2(A_2)$ et à partir de l'approximation 2 ($X_2 = Y_1 \oplus Y_3$) donc ($A_{2,2} = B_{2,1} \oplus B_{2,3}$)

  $\Rightarrow B_{2,1} \oplus B_{2,3} = B_{1,2} \oplus K_{2,2}$. En remplaçant $B_{1,2}$ par sa valeur dans (I) on obtient :

  $B_{2,1} \oplus B_{2,3} = ((P_{1,1} \oplus K_{1,1}) \oplus (P_{1,2} \oplus K_{1,2}) \oplus (P_{1,3} \oplus K_{1,3})) \oplus K_{2,2}$. ***(II)***

3ème tour :

- $A_3 = B_2 \oplus K_3$ donc $A_3 = [A_{3,1} = B_{2,1} \oplus K_{3,1}, A_{3,2} = B_{2,2} \oplus K_{3,2}, A_{3,3} = B_{2,3} \oplus K_{3,3}]$

  $\Rightarrow (B_{2,1} = A_{3,1} \oplus K_{3,1}$ et $B_{2,3} = A_{3,3} \oplus K_{3,3}) \Rightarrow B_{2,1} \oplus B_{2,3} = (A_{3,1} \oplus K_{3,1}) \oplus (A_{3,3} \oplus K_{3,3})$.

  On remplace $B_{2,1} \oplus B_{2,3}$ par sa valeur dans (II) on obtient:

  $(A_{3,1} \oplus K_{3,1}) \oplus (A_{3,3} \oplus K_{3,3}) = ((P_{1,1} \oplus K_{1,1}) \oplus (P_{1,2} \oplus K_{1,2}) \oplus (P_{1,3} \oplus K_{1,3})) \oplus K_{2,2}$.

En regroupant les termes on obtient ainsi l'équation finale :

$(K_{1,1} \oplus K_{1,2} \oplus K_{1,3} \oplus K_{2,2} \oplus K_{3,1} \oplus K_{3,3}) \oplus (P_{1,1} \oplus P_{1,2} \oplus P_{1,3}) \oplus (A_{3,1} \oplus A_{3,3}) = 0$. ***(III)***

On met $\sum K = (K_{1,1} \oplus K_{1,2} \oplus K_{1,3} \oplus K_{2,2} \oplus K_{3,1} \oplus K_{3,3})$, on obtient ainsi :

$\sum K \oplus (P_{1,1} \oplus P_{1,2} \oplus P_{1,3}) \oplus (A_{3,1} \oplus A_{3,3}) = 0$. (***IV***)

On a maintenant une approximation qui dépend de :

- Une partie des trois clés intermédiaires.
- Le texte en clair.
- Une partie de l'entrée de la dernière table de substitution.

Par l'application du lemme du Piling-Up de Matsui (expliqué en bas), on fixe la valeur de $\sum K$ à 0 ou à 1, et on calcule la probabilité que cette approximation soit valable.

Récupération des clés :

On a sous la main une approximation des 3 premiers tours de cet algorithme de chiffrement, mais il manque la clé du dernier tour K_4. C'est ici qu'interviennent les messages chiffrés dans cette analyse.

On prend un message chiffré C et on essaye de le déchiffrer en essayant toutes les valeurs possibles de la sous-clé K4 (attaque par force brute sur K_4).

Donc on calcule $C \oplus K_4$. En fait cela correspond à la sortie de la 3ème table de substitution c'est-à-dire à B_3 ($C = B_3 \oplus K_4 \Rightarrow B_3 = C \oplus K_4$).

On calcule ensuite la substitution inverse de S_3, c'est-à-dire $S_3^{-1}(C \oplus K_4)$. Or cette valeur correspond à A_3 donc $A_3 = S_3^{-1}(C \oplus K_4)$. On peut ainsi avoir une estimation de la validité des clés testées en comparant la valeur exacte retournée par la substitution inverse et l'approximation linéaire sur tout ou une partie des bits.

Avec un grand nombre de paires de messages on peut rendre plus précise les estimations.

Pour découvrir les autres clés intermédiaires on continue l'attaque en remontant progressivement dans les tours jusqu'à arriver à la 1ère sous-clé.

Sur des chiffrements plus complexes comme DES, on ne s'intéresse qu'à une partie des sous-clés afin de diminuer la complexité de l'attaque. Une étude plus poussée permet de déterminer quels bits de la dernière sous-clé ont vraiment une influence sur l'approximation linéaire. Dans son exemple avec un DES de 8 tours, Matsui indique que, malgré la présence de la dernière clé (de 48 bits) dans l'équation, seuls 6 bits de cette dernière clé influencent le terme où elle apparaît.

Plusieurs autres techniques ont été développées pour améliorer les performances de cette cryptanalyse.

IV. Lemme Piling-Up :

Le lemme Piling-Up est un résultat statistique introduit par Mitsuru Matsui en 1993 dans le cadre de la cryptanalyse linéaire. Ce lemme permet de quantifier le biais linéaire (***Linear Bias***) présent dans une approximation linéaire d'un algorithme de chiffrement symétrique par bloc [3].

Formulation mathématique :

Une équation linéaire dans le cadre de la cryptanalyse linéaire se présente sous la forme d'un ou-exclusif de variables binaires.

Soient N variables binaires, aléatoires et indépendantes X_1, X_2, .., X_N. La probabilité que l'équation linéaire définie sur cet ensemble de variables soit vraie est définie par :

$$P(X_1 \oplus X_2 \oplus .. \oplus X_N = 0) = \frac{1}{2} + 2^{N-1} \prod_{i=1}^{N} \varepsilon i$$

Avec ε_i est le biais linéaire de la variable aléatoire X_i

Raisonnement :

Soit $P(X_i = 0)$, la probabilité que la variable binaire X_i soit égale à 0. Cette probabilité soit égale à 1 si $X_i = 0$ et égale à 0 si $X_i = 1$.

Dans le cadre du lemme Piling-Up, on a donc affaire à des variables aléatoires, binaires et considérées comme indépendantes.

Si on considère tout d'abord le lemme pour deux variables aléatoires :

$$P(X_1 = i) = \begin{cases} p_1 & pour\ i = 0 \\ 1 - p_1 & pour\ i = 1 \end{cases} \qquad P(X_2 = j) = \begin{cases} p_2 & pour\ j = 0 \\ 1 - p_2 & pour\ j = 1 \end{cases}$$

Soit l'équation $X_1 \oplus X_2 = 0$.

$P(X_1 \oplus X_2 = 0) \Leftrightarrow P(X_1 = X_2)$ (grâce aux propriétés du XOR).

$X_1 = X_2 = 0$ et $X_1 = X_2 = 1$ sont des événements mutuellement exclus et de ce fait :

$$P(X_1 = X_2) = P(X_1 = X_2 = 0) + P(X_1 = X_2 = 1)$$

$$= P(X_1 = 0)P(X_2 = 0) + P(X_1 = 1)P(X_2 = 1).$$

$$= p_1p_2 + (1-p_1)(1-p_2) = p_1p_2 + 1 - p_1 - p_2 + p_1p_2$$

$P(X_1 \oplus X_2 = 0) = 2p_1p_2 - p_1 - p_2 + 1$. ($V$)

Soit $p_1 = \frac{1}{2} + \varepsilon_1$ et $p_2 = \frac{1}{2} + \varepsilon_2$ où ε_1 et ε_2 sont respectivement les biais des probabilités des deux variables X_1 et X_2.

Ce biais permet de quantifier le degré de déviation de la probabilité par rapport à ½.

On remplaçant p1 et p2 par leurs valeurs dans la formule (V) on obtient :

$$P(X_1 \oplus X_2 = 0) = 2(\tfrac{1}{2} + \varepsilon_1)(\tfrac{1}{2} + \varepsilon_2) - (\tfrac{1}{2} + \varepsilon_1) - (\tfrac{1}{2} + \varepsilon_2) + 1$$

$$= 2(\tfrac{1}{4} + \tfrac{1}{2}\varepsilon_1 + \tfrac{1}{2}\varepsilon_2 + \varepsilon_1\varepsilon_2) - \tfrac{1}{2} - \varepsilon_1 - \tfrac{1}{2} - \varepsilon_2 + 1$$

$$= \tfrac{1}{2} + \varepsilon_1 + \varepsilon_2 + 2\varepsilon_1\varepsilon_2 - \varepsilon_1 - \varepsilon_2 - \tfrac{1}{2} - \tfrac{1}{2} + 1$$

$$= \tfrac{1}{2} + 2\varepsilon_1\varepsilon_2.$$

Ainsi le biais linéaire de la formule précédente $(X_1 \oplus X_2 = 0)$ est $2\varepsilon_1\varepsilon_2$.

En généralisant la formule pour N variables aléatoires binaires et indépendantes on obtient :

$$P(X_1 \oplus X_2 \oplus .. \oplus X_N = 0) = \frac{1}{2} + 2^{N-1} \prod_{i=1}^{N} \varepsilon i$$

Si un seul biais linéaire ε_i est égal à 0 (une variable X_i est non biaisée), la probabilité de toute la formule est égale à ½ (toute la formule sera non biaisée).

Apport du lemme Piling-Up sur la cryptanalyse linéaire :

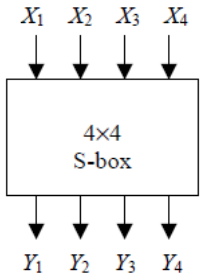
Le biais linéaire peut être positif ou négatif et quantifie l'écart par rapport à une distribution uniforme où l'espérance d'une variable aléatoire binaire est ½.

Plus ce biais est important, plus un algorithme de chiffrement est susceptible d'être attaqué via la cryptanalyse linéaire [4].

Exercice :

Soit la table de substitution (de 4x4 bits) suivante :

X	Y	X	Y
0000	1110	1000	0011
0001	0100	1001	1010
0010	1100	1010	0110
0011	0001	1011	1100
0100	0010	1100	0101
0101	1111	1101	1001
0110	1011	1110	0000
0111	1000	1111	0111



Calculer le biais linéaire pour les équations suivantes :

- $X_2 \oplus X_3 \oplus Y_1 \oplus Y_3 \oplus Y_4 = 0$
- $X_1 \oplus X_4 \oplus Y_2 = 0$
- $X_3 \oplus X_4 \oplus Y_1 \oplus Y_4 = 0$

Calculer la probabilité de satisfaction d'une analyse basée sur ces 3 approximations.

Références :

[1] Christopher Swenson « Modern Cryptanalysis- Techniques for advanced breaking » Wiley Publishing 2008 ISBN: 978-0-470-13593-8.

[2] Wikipedia-Linear-cryptanalysis « https://en.wikipedia.org/wiki/Linear_cryptanalysis».

[3] Wikipedia-Piling-up « https://en.wikipedia.org/wiki/Piling-up_lemma ».

[4] Vijay Yella «Linear and Differential Cryptanalysis of Substitution Permutation Networks»