TP2 : Attaque intégrale contre 4 tours de l'AES

Christina Boura
christina.boura@irif.fr

10 janvier 2025

Le but de ce TP est d'implémenter une attaque dite *intégrale* ou *square attack* contre 4 tours du chiffrement `AES`. Cette attaque repose sur une propriété qui est vérifiée avec une probabilité de 1 sur trois tours du chiffrement, quelle que soit la clé secrète. Exploiter cette propriété, appelée *distingueur*, permet à l'attaquant de retrouver la clé secrète du dernier tour avec une complexité bien inférieure à celle d'une recherche exhaustive.

1 Attaque intégrale

En 1997, Joan Daemen, Lars Knudsen et Vincent Rijmen ont présenté un nouveau chiffrement par bloc appelé `SQUARE` [1]. Lors de la conception de cet algorithme, les auteurs ont découvert une nouvelle attaque de type clair-choisi capable de casser six tours de `SQUARE`. En réponse à cette faiblesse, ils ont renforcé l'algorithme en ajoutant deux tours supplémentaires et publié la description avec les détails de cette nouvelle attaque, qui a ensuite été référencée sous le nom de *square attaque*.

Initialement, cette attaque n'a été appliquée qu'aux chiffrements de type SPN (*Substitution-Permutation Network*). Par la suite, Stephan Lucks a généralisé cette technique pour l'appliquer à des chiffrements non SPN [3] en la nommant *attaque par saturation* et l'a utilisé pour attaquer `Twofish`, un chiffrement par bloc de type Feistel. Aujourd'hui, cette cryptanalyse est davantage connue sous le nom de *cryptanalyse intégrale*, grâce à l'article de Lars Knudsen et David Wagner [2], qui ont structuré et popularisé cette approche en rassemblant les diverses techniques dans un cadre théorique unifié.

L'attaque intégrale est basée sur la notion suivante.

Définition 1 (Knudsen et Wagner, 2002)**.** *Pour tout multi-ensemble S d'éléments de $\mathbb{F}_2^n$, l'*intégrale de *S est définie comme la somme de tous les éléments de S, c'est-à-dire*

$$\bigoplus_{x \in S} x.$$

Dans une attaque intégrale, l'attaquant cherche à suivre l'évolution d'un multi-ensemble de départ bien choisi et à detecter des intégrales qui valent 0 après un certain nombre de tours. Nous allons décrire ici l'attaque classique qui s'applique sur les chiffrements orientés mots, dont la plupart des chiffrements de type SPN font partie. On note w le nombre de mots de l'état. Par exemple, $w = 16$ dans l'`AES` puisque les données sont représentées sous la forme d'une matrice d'octets 4×4. Nous notons N le nombre de clairs/chiffrés utilisés simultanément dans l'attaque. Ces textes clairs sont choisis de façon que le multi-ensemble de tous leur i-èmes mots ($1 \leq i \leq w$) vérifie une propriété spécifique pour au moins une valeur de i. Dans la plupart des cas, l'attaquant utilise un nombre de clairs/chiffrés égal au nombre de mots possibles.

L'attaquant suit l'évolution du multi-ensemble à travers le chiffrement, en se focalisant indépendamment sur chaque mot de l'état. Il distingue trois cas classiques selon que tous les i-ièmes mots sont égaux, distincts ou que leur somme soit égale à 0 :

1. $\mathcal{C}$ (constant) : Si un mot i est noté par le symbole $\mathcal{C}$, cela signifie que tous les éléments du multi-ensemble sont égaux pour le mot i.
2. $\mathcal{A}$ (all) : Si le mot i est noté par le symbole $\mathcal{A}$, alors tous les éléments du multi-ensemble sont distincts pour le mot i.
3. $\mathcal{B}$ (balanced) : Si le mot i est noté par le symbole $\mathcal{B}$, alors la somme de tous les i-ièmes mots est égale à 0.

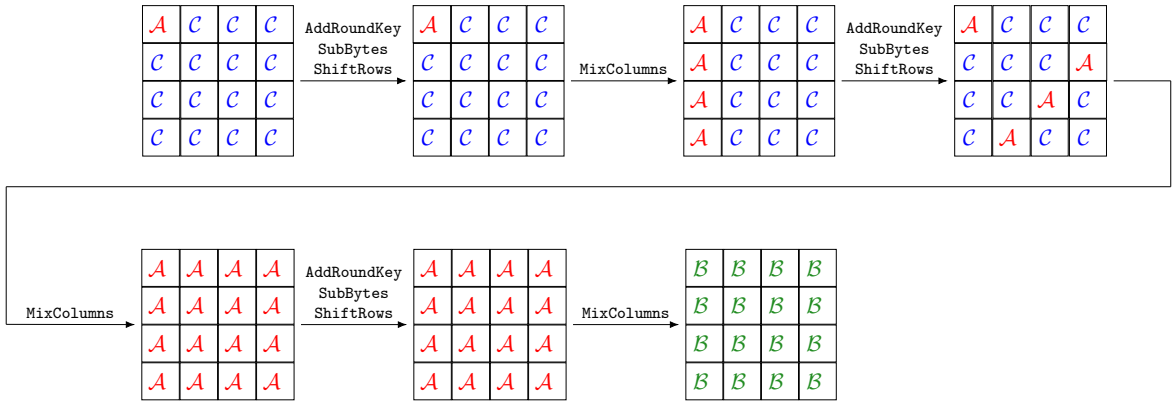

FIGURE 1 – Distingueur intégral sur 3 tours de l'AES.

2 Distingueur intégral sur 3 tours de l'AES

L'attaque intégrale contre 4, 5 et 6 tours de l'AES repose sur le distingueur simple suivant (voir également la figure 1) : soient 256 messages clairs prenant toutes les $2^8 = 256$ valeurs possibles sur le premier octet, et étant égaux à une valeur constante sur les 15 octets restants. Après trois tours de chiffrement, si l'on somme les 256 textes chiffrés octet par octet, la somme (l'intégrale) est égale à 0.

3 Travail à réaliser

1. Expliquer en détail le fonctionnement du distingueur présenté dans la figure 1.
2. Décrire comment exploiter ce distingueur pour retrouver la sous-clé du 4$^{\text{e}}$ tour de l'AES en adoptant une approche de type *diviser pour mieux régner*.
   Rappel : Le dernier tour de l'AES ne contient pas de `MixColumns`.
3. Implémenter cette attaque dans le langage de programmation de votre choix.

Références

[1] Joan Daemen, Lars R. Knudsen, and Vincent Rijmen. The block cipher Square. In Eli Biham, editor, *FSE '97*, volume 1267 of *Lecture Notes in Computer Science*, pages 149–165. Springer, 1997.

[2] Lars R. Knudsen and David A. Wagner. Integral cryptanalysis. In Joan Daemen and Vincent Rijmen, editors, *FSE 2002*, volume 2365 of *Lecture Notes in Computer Science*, pages 112–127. Springer, 2002.

[3] Stefan Lucks. The saturation attack - A bait for Twofish. In Mitsuru Matsui, editor, *FSE 2001*, volume 2355 of *Lecture Notes in Computer Science*, pages 1–15. Springer, 2001.