Université de Limoges

PROJET STATISTIQUES

Analyse de canaux auxiliaires

Thomas LAVAUR - Théo BONNET

encadré par
Mr. Christophe CLAVIER et Mr. Francisco SILVA

Master 1 cryptis - 2019/2020

1 Introduction

Le but de l'article est de s'intéresser aux attaques par canaux auxiliaires. Plus précisément sur la façon de pouvoir récupérer une information, comme une clé de chiffrement par exemple, sans connaître les donnés d'entrée ni de sortie. Pour cela, l'article réintroduit la méthode des slices proposée par Linge puis propose une amélioration à celle-ci. Dans un second temps, il est montrer qu'une implémentation de cette attaque est possible et qu'elle fonctionne contre les masque booléen de premier-ordre. Finalement des contre mesures pour ce protéger de ce genre d'attaque sont alors proposées. Tout au long, des exemples d'applications sur AES servent à illustrer les propos théoriques qui sont donc appuyer par des cas concrets.
Tout au long de l'article, on supposera connue la façon de repérer les moments "intéressants" dans les consommation de courant.

2 Méthode des slices

La première attaque proposée par Linge repose sur les probabilités jointes. En effet, lorsqu'on suppose que les entrées et les sorties suivent des distributions uniformes, indépendantes de la clé, mais si on regarde la distribution jointe des deux, celle-ci varie fortement en fonction de la clé. Cette liaison est aussi observée lorsqu'on passe aux poids de Hamming. A savoir que les probabilités d'apparition des couples $(hw(m), hw(y))$ varie avec la clé choisie (où $hw(m)$ est le poids de Hamming du message d'entrée et $hw(y)$ celui de la sortie après le passage dans la S-Box).

Dès lors qu'on est capable d'extraire de la consommation d'électricité ces deux poids de Hamming, on peut alors construire cette distribution des probabilités jointes. Ainsi, lors de la comparaison avec les modèles théoriques (ceux qu'on construit en supposant que le message suit une loi uniforme), si on fait une écoute de suffisamment de chiffrement (ou de l'opération de transformation qui nous intéresse), on peut supposer que l'information cherchée est la même que celle utilisée dans le modèle théorique qui se rapproche le plus de celui observé. On peut effectivement supposer que lorsque le nombre de chiffrement observé tend vers l'infini, la distribution des messages d'entrée tend vers une distribution uniforme.

L'article rappel donc les points importants qu'avait proposé Linge dans son article pour cette méthode qui sont au nombre de deux :

-Le premier étant de savoir comment transformer les mesures de courant en poids de Hamming. On sait que celui-ci est image d'une fonction affine par rapport à la consommation. Si la fonction est croissante (on résonnera de façon similaire si décroissante), on peut supposer que les valeurs de consommation sont rangées dans le même ordre que celui des poids de Hamming.
Puisque théoriquement, la proportion de chaque poids de Hamming observé suit une loi normale, on peut donner la même proportion de poids de Hamming correspondants. C'est donc pour cela que l'on décide de dire que les $\frac{\text{nombre d'écoute}}{256}\binom{8}{0}$ seront de poids le plus petit (car la fonction est croissante), en ainsi de suite, que les $\frac{\text{nombre d'écoute}}{256}\binom{8}{1}$ suivant seront de poids 1...

-Le second, une fois les poids de Hamming supposés correctes, est de savoir quelle distance est la meilleure à utiliser pour choisir le modèle théorique le plus proche de la distribution empirique. Pour cela, Linge à comparer de nombreuses distances différentes et comparer les meilleurs taux de réussite pour chacune et garda la meilleure.

3 Méthode du maximum de vraisemblance

La deuxième méthode du maximum de vraisemblance exploite mieux les probabilités jointes que la méthode précédente. Le principe est de calculer la probabilité que chaque clé soit à l'origine du chiffrement en utilisant les mesures théoriques.

Contrairement à la première méthode, la déduction des poids de Hamming à partir des mesures de courant n'est pas la même. Dans cette méthode, on estime des poids de Hamming à valeur réelle ce qui permet d'éviter des erreurs d'arrondissement qui pourrait intervenir dans la méthode des slices par exemple. On peut utiliser deux méthodes d'estimation :

- La régression linéaire où l'on doit estimer les coefficients α et β du modèle linéaire du courant suivant $l(m) = \alpha \times hw(m) + \beta$. Pour cela il est nécessaire de connaître des couples $(l(m), hw(m))$.

- L'analyse de variance fonctionne sur le même principe, c'est-à-dire la recherche des coefficients α et β. Le principe est de calculer la variance de la mesure pour un très grand nombre d'exécutions et ensuite à partir des relevés de variance en déduire la valeur de α a partir de la formule $Var(l(m)) = \alpha^2 Var(hw(m)) + Var(\omega)$ pour une mesure de courant bruitée (ω étant le bruit). Il ne reste ensuite plus qu'à en déduire β à partir des mesures au point d'intérêt.

Une fois que nous avons les poids de Hamming $(hw(m), hw(y))$, en utilisant la formule de Bayes et le théorème des probabilités totales on peut calculer pour chacune des 256 clefs leur probabilité d'être la bonne. Une fois cela fait on peut donc regarder la clef avec la plus forte probabilité.

4 Comparaison des deux méthodes en pratique

L'article propose ensuite une comparaison des deux modèles présentés précédemment. Le schéma de gauche présenté en page 10 met en scène cinq utilisations, il y en a trois à partir de la méthode des slices mais avec l'utilisation de différentes distances (produit interne en bleu, euclidienne en gris et le khi-deux de Pearson en vert). La quatrième en trait plein rouge est la méthode du maximum de vraisemblance à partir de poids de Hamming entier obtenu grâce à la méthode des slices et la cinquième en pointillés rouges la méthode du maximum de vraisemblance à partir de valeurs de poids de Hamming réelles.

Ce graphique nous montre la classement moyen de la méthode par rapport à son nombre d'exécutions et on peut très clairement voir que ce sont les deux méthodes basées sur le maximum de vraisemblance qui sont les meilleurs. A partir de poids de Hamming à valeurs entières générés par la méthode des slices, c'est la méthode du maximum de vraisemblance qui est la plus fiable. Et en utilisant cette méthode sur des poids de Hamming approchées à des valeurs entières et sur des valeurs réelles, on voit que c'est l'utilisation des valeurs réelles qui est la meilleure.

Le deuxième graphique à droite montre que si on ajoute un autre point de mesure (par exemple $x = m \oplus k$) on peut encore améliorer la méthode du maximum de vraisemblance. Dessus on peut voir les courbes pleines correspondant à l'utilisation de deux points de mesure pour différents niveaux de bruits et des courbes en tirets correspondant à l'utilisation de trois points de mesure pour les mêmes valeurs de bruits que les courbes pleines. Ainsi on peut voir que dans les trois cas, l'étude de trois points de mesures est bénéfique par rapport à seulement deux.

Par la suite, l'article nous propose une légère comparaison entre deux variantes. La première, que nous avons vu depuis le début, l'analyse des mesures de courant aux points m et $y = S(m \oplus k)$. Par rapport à la deuxième, l'analyse des mesures de courant aux points m et $x = m \oplus k$. Une des premières qui nous est expliquée est que la variante $m - x$ nous permet de récupérer moins d'informations que la variante $m - y$ mais en contre partie elle récupère plus efficacement l'information $hw(k)$ que la variante $m - y$ ne récupère k. Cela vient notamment du fait qu'il y ait seulement neuf modèles à comparer dans la variante $m - x$ et que, comme

nous le montre la figure 4, les modèles sont très distincts les uns par rapport aux autres. La partie basse de la figure nous montre également que cette variante possède un excellent classement pour la récupération de $hw(k)$.

5 Supports protégés par masquage booléen

Jusqu'ici, les deux méthodes présentées visaient à attaquer un support ou le chiffrement se faisait normalement. Cependant, de nombreux supports de chiffrement adoptent des contre-mesures en ajoutant un masque à différents endroits du chiffrement. On s'intéresse donc par la suite aux applications possibles des deux méthodes sur ces contre-mesures.
Plusieurs masquages sont alors étudiés. Dans le cadre d'une comparaison qui repose sur les probabilités jointes dans les deux méthodes, ce qui va changer est la construction du modèle empirique. Lors de nos mesures, les distributions jointes des mesures seront faussées par le masque booléen et le modèle construit ne correspondra plus au modèle théorique de la clé utilisé dans le chiffrement.

On a alors plusieurs possibilités : si m et y sont masqués par la même valeur, on peut construire les modèles théoriques pour chaque valeur de masque et comparer par la suite. Dans ce cas, pour des clés différentes on peut avoir deux valeurs de masque qui donne des modèles très très proches, mais il reste différents les uns des autres. On peut évidement raisonner de même pour le cas ou l'on masque m, x et y. Et il est montré que, comme pour précédemment, la variante $m - x - y$ est plus efficace car plus précise. Si le masque est le même pour m, x et y alors la méthode étudiée dans le document fonctionne parfaitement de la même façon et ne ralentira que très peu l'attaquant.

Finalement, si m est masqué de la façon suivante : $m \oplus u = m'$ et $x \oplus u = x'$ on montre que les distributions de couples $(HW(m), HW(x))$ sont les mêmes que $(HW(m'), HW(x'))$ car il s'agit simplement d'une permutation dans la S-Box. Toute cette partie théorique est appliquée par la suite en pratique et les résultats exploités dans la suite du papier.

6 Attaque avec la connaissance du clair

Désormais, nous allons considérer que nous connaissons le texte fourni en entrée et donc adapter les attaques pour prendre en compte cela. Pour commencer l'attaque est maintenant seulement faisable au premier tour, néanmoins l'attaque s'effectue de la même manière que précédemment. La figure 7 nous montre bien que la méthode du maximum de vraisemblance avec la connaissance du texte m est plus efficace que la CPA classique pour différents niveaux de bruit.

Deuxièmement, on considère le cas d'une implémentation masquée où l'on doit aussi essayer de déterminer le point d'intérêt u (u correspondant au masquage booléen) qui n'est pas forcément très aisé. La partie droite de la figure nous montre la comparaison entre la variante $m - y$ et l'attaque CPA pour l'implémentation masquée. On peut surtout remarquer que la variante $m - y$ est plus rapide que l'attaque CPA.

7 Expérience concrète

Cette partie nous donne les résultats d'expériences basées sur deux appareils différents, un premier possédant une implémentation naïve (non masquée) et le deuxième possédant une implémentation masquée. Le tableau 1 nous montre le classement des méthodes avec le produit interne, la division euclidienne et la méthode du maximum de vraisemblance sur l'appareil avec une implémentation naïve et le texte inconnu. Pour 1000 traces, sur les 16 octets différents on voit clairement que la méthode du maximum de vraisemblance est la meilleure. De la même manière le second tableau compare l'implémentation SO-CPA et l'implémentation du maximum de vraisemblance sur l'appareil avec le masquage booléen et la connaissance du texte. De façon pratique, on observe que les deux méthodes possèdent de très bons résultats

assez proches pour 200 traces, même si théoriquement la méthode du maximum de vraisemblance semblait la meilleure, on voit que pratiquement les différences de performance sont moins grandes.

8 Applications et contre-mesures proposées

Cette partie vise à appliquer la méthode du maximum de vraisemblance dans 3 cas différents. On montre que l'on peut retrouver la clé de session d'un chiffrement AES ou DES par exemple. La méthode du maximum de vraisemblance apporte ici en plus de la rapidité, une attaque possible sur les masquage booléen de premier ordre. Cela n'était pas possible avec l'attaque classique de Linge et al. Ces trois exemples montre donc les possibilités d'applications et que l'attaque est réalisable en pratique, souvent pour récupérer des clés ou morceaux de clés sur des processus de chiffrements.

Finalement, une mesure pour contrer cette attaque serait de ne pas utiliser le même masque pour les valeurs d'entrée et de sortie de la S-Box. Il est aussi proposé d'adopter toute mesure capable de complexifier la partie de retro-engineering qui vise à identifier l'endroit dans les mesures de courant où se passe l'action voulue. On pense entre autre aux méthodes permettant de rendre aléatoire la durée des opérations arithmétiques etc.

9 Conclusion

Cette article a tout d'abord montré que la meilleure attaque concernant l'analyse de la distribution jointe est celle basée sur le maximum de vraisemblance à travers plusieurs expérimentations théoriques ou pratiques. Il propose également différentes variantes pouvant s'appliquer selon le type d'appareils qu'il soit naïf ou qu'il utilise le masquage booléen ; de plus, il nous est proposé des moyens de rendre plus difficile la recherche de la clef pour ces implémentations. Pour conclure, nous avons globalement compris la grande majorité de ce qui était expliqué même si nous avons eu un peu de mal avec le concept de masque du premier et second ordre.