

Soutenance à mi-parcours: Cryptanalyse algébrique avec oracle

Christopher Goyet

THALES Communications

Équipe SALSA/LIP6/UPMC

THALES



UPMC
UNIVERSITÉ PARIS UNIVERSITÉS

Introduction

Cryptanalyse :

- évalue la sécurité des cryptosystèmes
- basée sur un problème sous-jacent supposé "difficile"

Exemple cryptographie asymétrique

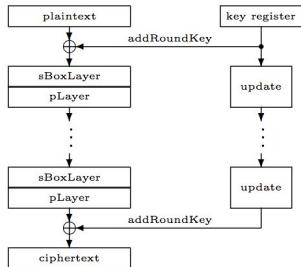
Cryptosystème RSA :

- basé sur un problème de théorie des nombres :
calculer la racine e -ème modulo $N = pq$

Attaque générale : factorisation de N

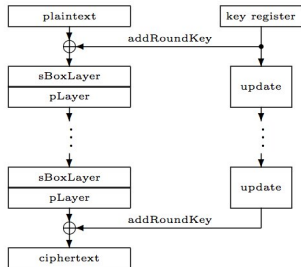
- Meilleur complexité reste : $O(e^{c \ln(x)^\alpha})$, $0 < \alpha < 1$
- Dernier record 2010 : $N \approx 2^{768}$

Exemple : chiffrement par bloc



Exemple : AES

Exemple : chiffrement par bloc



Exemple : AES

Cryptanalyse algébrique

$$\left\{ \begin{array}{l} x_1 x_2 + x_1 k_2 + x_1 + x_2 k_1 + x_3 + x_4 s_4 + \\ s_1 s_4 + s_3 s_4 + s_3 + s_4 k_4 + s_4 + k_1 k_2 + k_1 + k_3, \\ x_4 + s_1 s_3 + s_2 + s_4 + k_4 + 1, \\ \vdots \\ s_{21} + s_{52} y_{124} + s_3 y_{124} + y_{121} y_{124} + y_{121} + \\ y_{123} y_{124} + y_{124} k_{122} + y_{124} k_{123} + y_{124} + k_{121} \end{array} \right. \Leftarrow$$

résolution \Rightarrow clef secrète

Une approche différente

Au lieu de chercher à résoudre des problèmes difficiles en général...

Quels contextes \Rightarrow problèmes faciles ?

RSA : exemple de problème difficile devenant facile dans un contexte particulier

- ① 1985, Rivest, Shamir :
LSB de p connu \Rightarrow factorisation $N = pq$ temps polynomial

RSA : exemple de problème difficile devenant facile dans un contexte particulier

- ① 1985, Rivest, Shamir :
LSB de p connu \Rightarrow factorisation $N = pq$ temps polynomial
- ② 1996, Coppersmith :
MSB de p connu \Rightarrow factorisation $N = pq$ temps polynomial

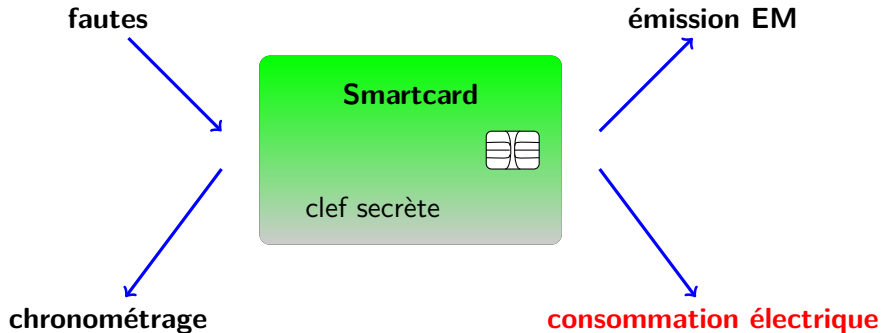
RSA : exemple de problème difficile devenant facile dans un contexte particulier

- ❶ 1985, Rivest, Shamir :
LSB de p connu \Rightarrow factorisation $N = pq$ temps polynomial
- ❷ 1996, Coppersmith :
MSB de p connu \Rightarrow factorisation $N = pq$ temps polynomial
- ❸ 1998, Boneh, Durfee, Frankel :
fraction exposant secret connu \Rightarrow factorisation $N = pq$ temps polynomial
- ❹ ...

Attaque par canaux auxiliaires

Implémentation d'algorithmes cryptographiques (carte à puce, FPGA, ...)

⇒ **vulnérabilités physiques**



"A correct implementation of a strong protocol is not necessarily secure"
(Kocher, 1999)

Une approche différente

Au lieu de chercher à résoudre des problèmes difficiles en général...

Quels contextes \Rightarrow problèmes faciles ?

\rightsquigarrow **information supplémentaire**

Une approche différente

Au lieu de chercher à résoudre des problèmes difficiles en général...

Quels contextes \Rightarrow problèmes faciles ?

\rightsquigarrow **information supplémentaire**

Modélisation :



Cryptanalyse avec Oracle

- Comment ? \rightsquigarrow permet d'accélérer attaques
- Quoi ? Combien ? \rightsquigarrow temps polynomial

Attaque algébrique par canaux auxiliaires

Conférence internationale :



[COSADE 2011](#)

Analysis of the Algebraic Side Channel Attack

Autre exposé :

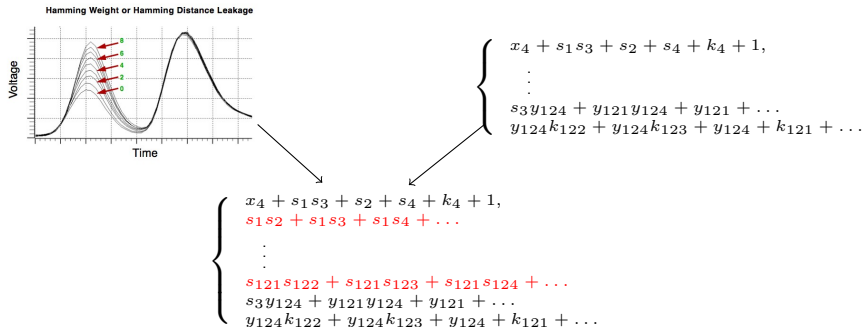


[Journées nationales Codage et Cryptographie 2011](#)

Article détaillé prochainement soumis à un journal

Attaque algébrique par canaux auxiliaires (ASCA)

Nouveau type d'attaque récemment proposé par Renaud, Standaert et Veyrat-Charvillon (CHES 2009, Inscrypt2009)



Idée principale

- 1 Phase Online : mesures de fuites physiques
- 2 Phase Offline : attaque algébrique : modélisation + résolution

Attaque algébrique par canaux auxiliaires

Avantages

- moins d'observations qu'une DPA classique
- étape de résolution apparemment très **rapide** (avec SAT-solver)
- peut fonctionner avec contre-mesures par masquage

Attaque algébrique par canaux auxiliaires

Avantages

- moins d'observations qu'une DPA classique
- étape de résolution apparemment très **rapide** (avec SAT-solver)
- peut fonctionner avec contre-mesures par masquage

Cependant, l'efficacité dépend de

- l'appareil ciblé et la qualité des traces
- du modèle de fuite
- de la quantité d'information disponible
- du système d'équations (modélisation)
- des **heuristiques** utilisées par le **SAT-solver**
- ...

⇒ résultats des expériences très difficiles à expliquer et à prédire

État de l'art



Algebraic Side-Channel Attacks

Renauld, Standaert, Inscrypt 2009



Algebraic Side-Channel Attacks on the AES : Why Time also Matters in DPA

Renauld, Standaert, Veyrat-Charvillon, CHES 2009



Blind Differential Cryptanalysis for Enhanced Power Attacks

Handschuh, Preneel, Selected Areas in Cryptography 2006



Multi-Linear cryptanalysis in Power Analysis Attacks

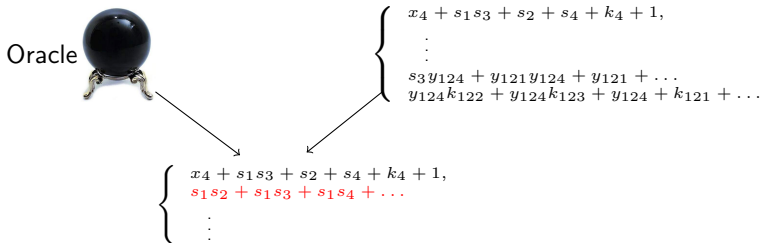
Roche, Tavernier, 2009



...

Analyse de la phase algébrique

Dans le but d'expliquer l'efficacité de l'étape de résolution



Objectifs

- impact du modèle d'oracle ? HW dans notre cas
- combien requêtes à oracle nécessaires ?
- certaines réponses plus intéressantes ?
- quelles parties du chiffrement cibler ?

Nécessite méthode de résolution plus stable et prévisible qu'avec SAT-solver **sans heuristiques** \implies **Bases Gröbner**

Objectif : analyse de la phase algébrique

Modèle d'Oracle :

- poids de Hamming sur 8-bits à chaque étape
- supposés sans erreur

PRESENT	PRESENT+Oracle
SAT-Solver = ∞ ❌	SAT-Solver $\simeq 1s$ ✓ (CHES 2009)
Base Gröbner = ∞ ❌	Gröbner basis (F4) $\simeq 20min$ ✓ (our work)

∞ : >3jours

SAT-Solver = Heuristiques \Rightarrow

~~analyse~~

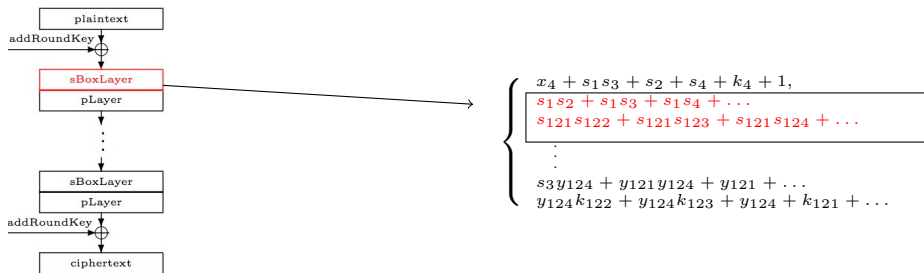
Gröbner basis = Résolution Algébrique \Rightarrow

analyse théorique
confirmée par expériences

Étude locale

- boîtes-S fournissent la résistance aux attaques algébriques
- seules parties non-linéaires
- représentés par systèmes d'équations de haut degré

Principal critère = **Immunité Algébrique** des boîtes-S



Immunité Algébrique (Ars, Carlet, Courtois, ...)

Principal critère attaques algébriques = **Immunité Algébrique**

Notations : soient

- $S : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$ une n -bits boîte-S.
- X_1, \dots, X_n et Y_1, \dots, Y_n respectivement ces bits d'entrées et de sorties.
- $F_i(X_1, \dots, X_n)$, $1 \leq i \leq n$ les fonctions définissant sorties de S

Définition de l'Immunité Algébrique

Soit $I_S = \langle \{Y_i - F_i(X_1, \dots, X_n), X_i^2 - X_i, Y_i^2 - Y_i, i \in \{1 \dots n\}\} \rangle$.

Alors l'**Immunité Algébrique** de S est défini par

$$AI(S) = \min\{\deg(P), P \in I_S \setminus \{0\}\}$$

Le **nombre** de telles relations de bas degré est aussi un invariant important

Immunité Algébrique (Ars, Carlet, Courtois, ...)

Comment calculer l'**Immunité Algébrique** pour une boîte-S donnée S ?
Il suffit de calculer une base de Gröbner avec l'ordre **DRL** de

$$I_S = \langle \{Y_i - F_i(X_1, \dots, X_n), X_i^2 - X_i, Y_i^2 - Y_i, i \in \{1 \dots n\}\} \rangle$$

En effet, on a

Proposition

La base de Gröbner réduite G_S de I_S pour n'importe quel ordre du degré contient une base linéaire des relations de plus bas degré S (i.e. les polynômes $P \in I_S$ tels que $\deg(P) = AI(S)$).

Exemple de la boîte-S de l'AES

L'Immunité algébrique de la fonction inverse sur \mathbb{F}_{2^8} (e.g. AES S-box) est égale à **2**. En effet, elle peut être représentée par 39 équations quadratiques sur \mathbb{F}_2 (Courtois 2002)

Une nouvelle notion d'Immunité Algébrique

ASCA \Rightarrow considérer **informations supplémentaires**

Notations

Pour toute valeur possible ℓ retournée par l'oracle, on note

- $E_\ell(X_1, \dots, X_n, Y_1, \dots, Y_n)$ équations données par l'oracle représentant l'information ℓ
- $I_\ell = \langle E_\ell(X_1, \dots, X_n, Y_1, \dots, Y_n) \cup \{Y_i - F_i(X_1, \dots, X_n), X_i^2 - X_i, Y_i^2 - Y_i, i \in \{1 \dots n\}\} \rangle$

Définition d'Immunité Algébrique avec Oracle

Le plus petit degré des relations dans I_ℓ est appelé **Immunité Algébrique avec Oracle** de la boîte S . Il est noté $AI(S, \ell)$ et le nombre de telles relations est noté $\#AI(S, \ell)$.

Exemple du poids de Hamming (HW)

Hypothèse : Oracle renvoie

- HW de l'entrée de S
- HW de la sortie de S
- $\ell = (w_{in}, w_{out})$

\Rightarrow l'idéal I_ℓ contient au moins 2 **polynômes linéaires** indépendants :

$$X_1 + \cdots + X_n + (w_{in} \bmod 2) \in I_\ell$$

$$Y_1 + \cdots + Y_n + (w_{out} \bmod 2) \in I_\ell$$

Propositions

\forall boîte S , et $\forall \ell \in \{0, \dots, n\}^2$

$$AI_{HW}(S, \ell) = 1$$

$$\#AI_{HW}(S, \ell) \geq 2$$

La boîte- S est-elle **linéarisée** pour autant ?

Exemple HW ($\ell = (w_{in}, w_{out})$)

\Rightarrow l'idéal I_ℓ contient au moins 2 **polynômes linéaires** indépendants :

$$X_1 + \cdots + X_n + (w_{in} \bmod 2) \in I_\ell$$

$$Y_1 + \cdots + Y_n + (w_{out} \bmod 2) \in I_\ell$$

ne nous aident pas beaucoup pour résoudre notre système :

- pas de relations linéaires entre l'entrée et la sortie
- étape de substitution reste donc **non linéaire**

Mais, nous savons maintenant que cette information supplémentaire peut apporter des équations linéaires !!

Y en a-t-il d'autres de plus intéressantes ?

Exemple HW ($\ell = (w_{in}, w_{out})$)

Exemple trivial : $w_{in} = 0$

\forall boîte S , si $w_{in} = 0$ alors $X_1 = X_2 = \dots = X_n = 0$
et les Y_i sont donnés par

$$Y_1, \dots, Y_n = S(0, \dots, 0) = y_1, \dots, y_n$$

$\#AI_{HW}(S, \ell) = 2n$ est **maximal** dans ce cas et
la boîte correspondante est **entièrement décrite** par des relations linéaires

Exemple HW ($\ell = (w_{in}, w_{out})$)

Exemple trivial : $w_{in} = 0$

\forall boîte S , si $w_{in} = 0$ alors $X_1 = X_2 = \dots = X_n = 0$
et les Y_i sont donnés par

$$Y_1, \dots, Y_n = S(0, \dots, 0) = y_1, \dots, y_n$$

$\#AI_{HW}(S, \ell) = 2n$ est **maximal** dans ce cas et
la boîte correspondante est **entièrement décrite** par des relations linéaires

Exemple de PRESENT : $\#AI_{HW}(S, (w_{in}, w_{out}))$

$w_{in} w_{out}$	0	1	2	3	4	5	6	7	8
0					16				
1					9				
2			15	15	8	13	15		
3			9	5	9	5	9		
4	16	15	14	2	11	3	12	13	16
5		13	13	2	7	10	11	13	
6		15	12	15	7	15	14		
7			13		13				
8			16						

Suivant la réponse
de l'oracle,
beaucoup
d'équations linéaires
peuvent apparaître

Autre invariant

Définition

\forall boîte S , \forall valeur $\ell = (w_{in}, w_{out})$, nous définissons

$$N_S(\ell) = \#V(I_\ell)$$

- simple à décrire, lecture directe
- utile pour caractérisation

Autre invariant

Définition

\forall boîte S , \forall valeur $\ell = (w_{in}, w_{out})$, nous définissons

$$N_S(\ell) = \#V(I_\ell)$$

- simple à décrire, lecture directe
- utile pour caractérisation

Proposition

Soit n la taille de S . Si $AI(S, \ell) = 1$ et $N_S(\ell)$ est non nul alors

$$\#AI(S, \ell) \geq 2n + 1 - N_S(\ell)$$

$N_S(\ell)$ petit \rightsquigarrow beaucoup de relations linéaires

Boîte-S de PRESENT

Hypothèse : bus 8-bits et oracle **poids de Hamming**

$w_{in} w_{out}$	0	1	2	3	4	5	6	7	8
0					16				
1					9				
2			15	15	8	13	15		
3			9	5	9	5	9		
4	16	15	14	2	11	3	12	13	16
5		13	13	2	7	10	11	13	
6		15	12	15	7	15	14		
7			13		13				
8			16						

FIGURE: $\#AI_{HW}(S, w_{in}, w_{out})$

$w_{in} w_{out}$	0	1	2	3	4	5	6	7	8
0					1				
1					8				
2			2	2	18	4	2		
3			8	12	8	20	8		
4	1	2	3	24	7	22	6	4	1
5		4	4	16	12	8	8	4	
6		2	6	2	12	2	4		
7			4		4				
8			1						

FIGURE: $N_S(w_{in}, w_{out})$

Observations

- confirme que N_S petit $\Rightarrow \#AI$ grand
- permet de trier les réponses par importance
- la plupart donnant beaucoup de relations linéaires : $\mathbb{E}(\#AI_{HW}) = 7.9$

Boîte-S de PRESENT

Hypothèse : bus 8-bits et oracle **poids de Hamming**

$w_{in}w_{out}$	0	1	2	3	4	5	6	7	8
0					16				
1					9				
2			15	15	8	13	15		
3			9	5	9	5	9		
4	16	15	14	2	11	3	12	13	16
5		13	13	2	7	10	11	13	
6		15	12	15	7	15	14		
7			13		13				
8			16						

FIGURE: $\#AI_{HW}(S, w_{in}, w_{out})$

$w_{in}w_{out}$	0	1	2	3	4	5	6	7	8
0					1				
1					8				
2			2	2	18	4	2		
3			8	12	8	20	8		
4	1	2	3	24	7	22	6	4	1
5		4	4	16	12	8	8	4	
6		2	6	2	12	2	4		
7			4		4				
8			1						

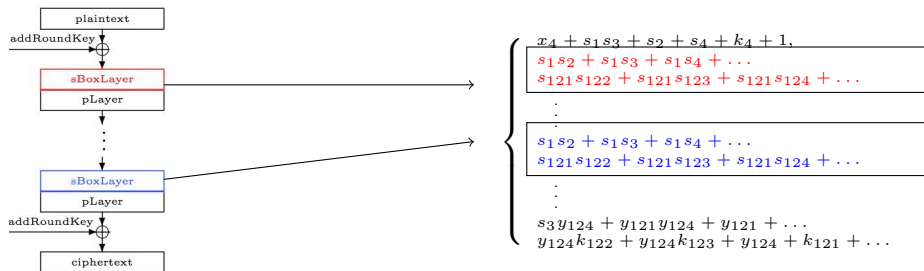
FIGURE: $N_S(w_{in}, w_{out})$

Observations

- confirme que N_S petit $\Rightarrow \#AI$ grand
- permet de trier les réponses par importance
- la plupart donnant beaucoup de relations linéaires : $\mathbb{E}(\#AI_{HW}) = 7.9$

Stratégie de résolution efficace

Calcul direct par base de Gröbner \rightsquigarrow ❌



Résultats :

Calculs successifs de bases de Gröbner (F4)

→ meilleur contrôle sur le degré

→ stratégie de résolution efficace \simeq 20min ✓

\rightsquigarrow implémenté en magma

Raisons du succès

Explication de la réussite de l'attaque lorsque :

- chiffrement par blocs très simple : PRESENT
- Oracle donne **tous les poids de Hamming** sur 8-bits à tous les tours
- supposés sans erreurs

Raisons :

- $AI_{HW} = 1$
- $\mathbb{E}(\#AI_{HW}) = 7,9$
- $\mathbb{P}(\#AI_{HW} \geq 8) \approx \frac{1}{2} \rightsquigarrow$ distribution uniforme

$\Rightarrow \mathbb{E}(\text{couche de substitution entière}) \approx 64$

Attaque efficace sous hypothèses plus faibles suivantes :

- informations supplémentaires sur **3 ou 4 tours seulement** ?
- sans le **clair ni le chiffré** ?

Raisons du succès

Explication de la réussite de l'attaque lorsque :

- chiffrement par blocs très simple : PRESENT
- Oracle donne **tous les poids de Hamming** sur 8-bits à tous les tours
- supposés sans erreurs

Raisons :

- $AI_{HW} = 1$
- $\mathbb{E}(\#AI_{HW}) = 7,9$
- $\mathbb{P}(\#AI_{HW} \geq 8) \approx \frac{1}{2} \rightsquigarrow$ distribution uniforme

$\Rightarrow \mathbb{E}(\text{couche de substitution entière}) \approx 64$

Attaque efficace sous hypothèses plus faibles suivantes :

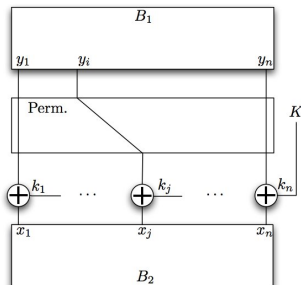
- informations supplémentaires sur **3 ou 4 tours seulement** ?
- sans le **clair ni le chiffré** ?

Peu d'informations consécutives ou clair/chiffré inconnus

Retour à l'étude locale des boîtes-S :

$N_S(\ell)$ petit \Rightarrow forte linéarisation

$N_S(\ell)$ très petit (≤ 6) \Rightarrow bits d'entrée/sortie fixés !!



\rightsquigarrow bits d'autres sous-clefs facilement déduits à travers le keyschedule

Boîtes-S plus résistantes ?

Nécessaires :

- peu de bits fixés
- faible linéarisation

↪ maximiser N_S pour tout poids de Hamming

Choix de la classe : N_S pour tout poids de Hamming

$$N_S(w_{in}, w_{out}) = \#(HW^{-1}(w_{in}) \cap S^{-1}(HW^{-1}(w_{out})))$$

Alors S doit vérifier

$$HW^{-1}(w_{in}) = S^{-1}(HW^{-1}(w_{out}))$$

et donc

$$w_{in} = w_{out} \text{ ou } w_{in} = n - w_{out}$$

Boîtes-S plus résistantes ?

Exemple d'une telle boîte-S sur 4-bits :

x	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
$S(x)$	0	B	5	C	E	6	9	8	7	5	3	1	A	2	4	F

$HW(x)$	$HW(S(x))$
0	0
1	3
2	2
3	1
4	4

↪ meilleure résistance confirmée par les expériences

Caractérisation (Carlet) :

$$S(x) = \pi(x) + f(HW(x))(1, \dots, 1)$$

- $\pi(x)$ = permutation stable sur HW constant
- f = fonction booléenne t.q. $\forall x \in \{0, \dots, n\}, f(x) = f(n - x)$

Cependant, $\text{nonlinéarité}(S) \simeq 0 \Rightarrow$ très faible contre cryptanalyse linéaire

Expériences

Expériences contre PRESENT et AES

Analyse confirmée par expériences :

GB

- rejette réponses grand N_S ✓
- rejette réponses petit N_S ✗
- sur tours non consécutifs ✗

Expériences

Expériences contre PRESENT et AES

Analyse confirmée par expériences :

	GB	SAT-solver
• rejette réponses grand N_S	✓	✓
• rejette réponses petit N_S	✗	✗
• sur tours non consécutifs	✗	✗

Analyse est valide avec bases Gröbner **et** SAT-solver

Premier bilan

- Bonne compréhension de l'influence des informations supplémentaires
 - ▶ Nouvelle notion d'immunité algébrique
 - ▶ Résultats des expériences expliqués
 - ▶ Tri par importance des informations supplémentaires

Perspectives

- Autres classes boîtes-S résistantes contre ASCA et autres cryptanalyses classiques
- Gérer les erreurs

Autre modèle d'oracle apportant moins d'information ?

⇒ même analyse sur Distance Hamming

Boîtes-S de PRESENT

Hypothèses : bus 8-bits et oracle renvoyant **DISTANCE Hamming**

Définition :

$$d = HD(x, S(x)) = HW(x \oplus S(x))$$

HD modèle :

- $AI_{HD}(d) = 1$
- $\#AI_{HD}(d) \geq 1$
- $\mathbb{E}(\#AI_{HD}) = 2,3$
- $\mathbb{P}(\#AI_{HD} = 1) \approx \frac{7}{10}$

d	0	1	2	3	4	5	6	7	8
$N_S(d)$	0	0	16	56	81	64	30	8	1
$\#AI_{HD}(S, d)$	0	0	10	3	1	1	1	9	16
Bits fixés	0	0	0	0	0	0	0	0	16

FIGURE: HD avec boîtes-S de PRESENT

Beaucoup moins que dans le modèle HW

↪ systèmes non linéarisés (ou très peu), pas de bits fixés

↪ prévoit résolution beaucoup plus difficile ↪ confirmé par expériences

Attaque algébrique par collisions et par fautes



Algebraic Methods in Side-Channel Collision Attacks and Practical Collision Detection

Bogdanov, Kizhvatov, Pyshkin, Indocrypt 2008



...



A Differential Fault Attack Technique against SPN Structures, with Application to the AES and KHAZAD

Piret, Quisquater, CHES 2003



Piret and Quisquater's DFA on AES Revisited

Giraud, Thillard, eprint 2010



...

Attaque algébrique par collisions ou par fautes

Idée : généraliser en attaques algébriques avec oracle

- méthode systématique exploiter les informations supplémentaires
- réduisant au minimum le nombre de fautes/collisions nécessaires
- étendre à d'autres tours
- modèles de fautes plus complexes (plusieurs fautes, ...)



Modélisation :

$$\left\{ \begin{array}{l} x_1 x_2 + x_1 k_2 + x_1 + x_2 k_1 + x_3 + x_4 s_4 + \\ \quad s_1 s_4 + s_3 s_4 + s_3 + s_4 k_4 + s_4 + k_1 k_2 + k_1 + k_3, \\ x_4 + s_1 s_3 + s_2 + s_4 + k_4 + 1, \\ \vdots \\ c_1 + 1, c_2, \dots, c_{128}, \end{array} \right.$$

Attaque algébrique par collisions ou par fautes

Expériences contre AES

Résultats :

- ① Attaque DFA Piret et Quisquater tour 7 ✓
- ② Collisions tours extrêmes ✓
- ③ même nombre nécessaires
- ④ autres tours ✗

↪ pas d'améliorations notables

Perspectives :

- ① Autres cryptosystèmes : DES (Courtois, 2010), ...
- ② Autres modèles de fautes (Kim, 2011)
- ③ Généralisation par méthodes algébriques

Attaque algébrique par collisions ou par fautes

Expériences contre AES

Résultats :

- ① Attaque DFA Piret et Quisquater tour 7 ✓
- ② Collisions tours extrêmes ✓
- ③ même nombre nécessaires
- ④ autres tours ✗

↪ pas d'améliorations notables

Perspectives :

- ① Autres cryptosystèmes : DES (Courtois, 2010), ...
- ② Autres modèles de fautes (Kim, 2011)
- ③ Généralisation par méthodes algébriques

Conclusion - Perspectives

Conclusion - Perspectives

Conclusion

- Analyse des attaques ASCA
- Généralisation algébrique des attaques par fautes et par collisions

Perspectives

- Oracle en cryptographie asymétrique
- Travail en cours sur (EC)DSA avec information implicite
- Attaques basées sur méthodes de Coppersmith et LLL