Quelques aspects de la cryptographie et du codage - Partie 2

Claude Carlet (Université Paris 8-MAATICAH et INRIA)

Sommaire

- Rappels
- Utilisation des fonctions booléennes dans les schémas par blocs
- Bornes sur la nonlinéarité
- Fonctions optimales connues
- Une notion de nonlinéarité pour les boîtes S des schémas par flots

Rappels

Critères cryptographiques sur les fonctions booléennes

1. toute fonction cryptographique f doit avoir un *degré élevé*.

2. toute fonction cryptographique f doit être *équilibrée* (i.e. équidistribuée sur $\{0,1\}$).

3. toute fonction cryptographique doit être *à grande distance de Hamming*

$$d_H(f,l) = \#\{x \in F_2^n;\ f(x) \neq l(x)\} = w_H(f+l)$$

des fonctions affines.

La *nonlinéarité* $\mathcal{NL}(f)$ de f est sa distance minimale aux fonctions affines.

La nonlinéarité peut être caractérisée à l'aide de la *transformation de Walsh*.

$$\widehat{f_\chi}(a) = \sum_{x \in F_2^n} (-1)^{f(x)+a\cdot x}.$$

Relation de *Parseval* :

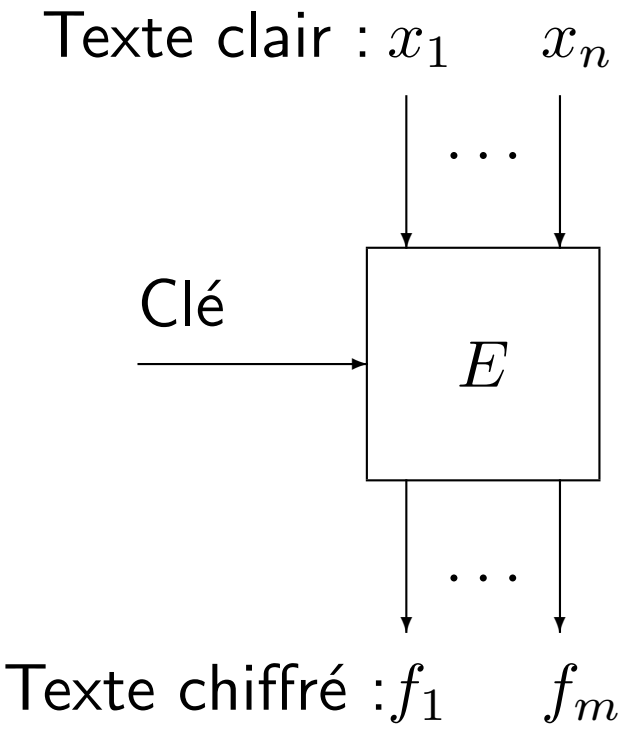
$$\sum_{a \in F_2^n} \widehat{f_\chi}^2(a) = 2^{2n}.$$

On appelle *courbes* les fonctions de nonlinéarité $2^{n-1} - 2^{n/2-1}$ (n pair).

Alors $\widehat{f_\chi}(a) = \pm 2^{n/2}$ pour tout a (condition indépendante du choix du produit scalaire), et la fonction définie par $\widetilde{f}(a) = 0$ si $\widehat{f_\chi}(a) = 2^{n/2}$ et $\widetilde{f}(a) = 1$ sinon est elle-même courbe.

Une fonction est courbe si et seulement si toutes ses dérivées $f(x) + f(x + a)$ sont équilibrées : provient du fait que la transformée de Fourier de la *fonction d'autocorrélation* de f : $a \mapsto \sum_{x \in F_2^n} (-1)^{f(x)+f(x+a)}$ est égale au carré de la transformée de Walsh de f.

Utilisation des fonctions booléennes dans les schémas par blocs



Fonctions vectorielles $F(x) = (f_1(x), \ldots, f_m(x)) : F_2^n \mapsto F_2^m$.
On appelle F une (n, m)-*fonction vectorielle*.

Forme algébrique normale :

$$F(x_1, \cdots, x_n) = \sum_{I \subseteq \{1,\ldots,n\}} a_I \left(\prod_{i \in I} x_i \right); \quad a_I \in F_2^m.$$

Degré : le degré maximal des f_j ; $j = 1, \ldots, m$.
En cryptographie, le degré *minimal* des combinaisons linéaires non nulles des f_j a également de l'importance.

Une (n, m)-fonction F est dite *équilibrée* si ses sorties sont uniformément distribuées sur F_2^m.

CNS : les combinaisons linéaires non nulles, $v \cdot F$, $v \neq 0$, des fonctions coordonnées de F sont équilibrées.

En effet :

$$\#F^{-1}(b) = 2^{-m} \sum_{v \in F_2^m, x \in F_2^n} (-1)^{v \cdot (F(x)+b)} =$$

$$2^{-m} \sum_{v \in F_2^m} (-1)^{v \cdot b} \sum_{x \in F_2^n} (-1)^{v \cdot F(x)}$$

est la transformée de Fourier de la fonction $v \mapsto \sum_{x \in F_2^n} (-1)^{v \cdot F(x)}$.

La *nonlinéarité* d'une (n, m)-fonction F est le paramètre qui quantifie sa résistance à l'attaque linéaire (Matsui Eurocrypt'93) :

$$N_F = \min_{v \in F_2^{m*}} N_{v \cdot F} \quad (1)$$

$$= 2^{n-1} - \frac{1}{2} \max_{v \in F_2^{m*}, u \in F_2^n} |\sum_{x \in F_2^n} (-1)^{v \cdot F(x) + u \cdot x}|. \quad (2)$$

La nonlinéarité est un invariant affine (à gauche et à droite).

La construction d'une (n, m)-fonction vectorielle hautement non-linéaire ne se réduit pas à la construction de m fonctions booléennes hautement non-linéaires.

Bornes sur la nonlinéarité

On a : $N_F \leq 2^{n-1} - 2^{n/2-1}$ (borne générale).

Pour tout n pair et tout $m \leq n/2$, il existe F (dite courbe) telle que $N_F = 2^{n-1} - 2^{n/2-1}$.

Exemples avec $F_2^n \sim F_{2^n}$ et $x \cdot y = tr(xy)$:

• Nyberg : $m = n/2$, $F_2^m \sim F_{2^m}$, $F_2^n \sim (F_{2^m})^2$, $\pi : F_{2^m} \mapsto F_{2^m}$ permutation, $g : F_{2^m} \mapsto F_{2^m}$ quelconque et $F(x, y) = x \times \pi(y) + g(y)$.

• Exemple de modification en une fonction équilibrée : $F(x, y) = \begin{cases} \frac{x}{y} & \text{si } y \neq 0 \\ x & \text{si } y = 0 \end{cases}$. On a $N_F = 2^{n-1} - 2^m$.

Une (n, m)-fonction vectorielle est courbe si et seulement si toutes ses dérivées $F(x) + F(x + a)$ sont équilibrées.

Mais les fonctions courbes n'existent que pour $m \leq n/2$

(n pair) : $F^{-1}(b) = 2^{-m} \sum_{x \in F_2^n; v \in F_2^m} (-1)^{v \cdot (F(x)+b)} = 2^{n-m} + 2^{\frac{n}{2}-m} \sum_{v \in F_2^m; v \neq 0} (-1)^{\widetilde{F_v}(0) \oplus v \cdot b}$ où $F_v(x) = v \cdot F(x)$.

Borne de Sidelnikov-Chabaud-Vaudenay :

$$N_F \leq 2^{n-1} - \frac{1}{2}\sqrt{3 \times 2^n - 2 - 2\frac{(2^n-1)(2^{n-1}-1)}{2^m-1}}.$$

Idée de la preuve :

$$\max_{v\in F_2^{m*},u\in F_2^n}\left(\sum_{x\in F_2^n}(-1)^{v\cdot F(x)+u\cdot x}\right)^2 \geq$$

$$\frac{\sum_{v\in F_2^{m*},u\in F_2^n}\left(\sum_{x\in F_2^n}(-1)^{v\cdot F(x)+u\cdot x}\right)^4}{\sum_{v\in F_2^{m*},u\in F_2^n}\left(\sum_{x\in F_2^n}(-1)^{v\cdot F(x)+u\cdot x}\right)^2}$$

$$\forall v\in F_2^m;\ \sum_{u\in F_2^n}\left(\sum_{x\in F_2^n}(-1)^{v\cdot F(x)+u\cdot x}\right)^2 = 2^{2n}$$

et

$$\sum_{v\in F_2^m, u\in F_2^n} \left(\sum_{x\in F_2^n} (-1)^{v\cdot F(x)+u\cdot x} \right)^4 =$$

$$\sum_{x,y,z,t\in F_2^n} \left[\sum_{v\in F_2^m} (-1)^{v\cdot(F(x)+F(y)+F(z)+F(t))} \right] \left[\sum_{u\in F_2^n} (-1)^{u\cdot(x+y+z+t)} \right.$$

$$= 2^{n+m} \#\{(x,y,z)/\, F(x)+F(y)+F(z)+F(x+y+z)=0\} \geq$$

$$2^{n+m} \#\{(x,y,z)/\, x=y \text{ ou } x=z \text{ ou } y=z\}.$$

La borne de Sidelnikov-Chabaud-Vaudenay n'améliore la borne générale que pour $m \geq n$. Elle est elle-même améliorée pour m grand par rapport à n (C.C.- C. Ding).

Elle n'est atteinte que pour $n = m$ impair. Les fonctions atteignant cette borne $N_F \leq 2^{n-1} - 2^{\frac{n-1}{2}}$ s'appellent *presque courbes* (AB).

Ce sont les fonctions qui "résistent" le mieux à l'attaque linéaire de Matsui.

Elles jouent aussi un rôle en séquences pour le CDMA :

une fonction puissance $F(x) = x^d$ sur F_{2^n} est AB si et seulement si les corrélations croisées $\sum_{i=0}^{2^n-2}(-1)^{s_{di}+s_{i+t}}$ $(0 \leq t \leq 2^n - 2)$ de la m-séquence $s_i = tr(\alpha^i)$ (où α est primitif) avec la séquence décimée s_{di} sont optimales.

Toute fonction AB est en même temps *presque parfaitement non-linéaire* (APN) :

pour tout $a \in F_2^{n*}$ et tout $b \in F_2^m$, l'équation $F(x) + F(x+a) = b$ admet au plus 2 solutions. Les fonctions APN sont celles qui "résistent" le mieux à l'attaque différentielle de Biham et Shamir.

Fonctions optimales connues

Seuls exemples connus jusqu'à récemment de fonctions AB : quelques fonctions puissances sur les corps finis :

les fonctions puissance x^s suivantes sur le corps à 2^n éléments :

- *Gold* : $s = 2^h + 1$ avec $\gcd(h, n) = 1$ et $1 \leq h \leq \frac{n-1}{2}$.
- *Kasami* : $s = 2^{2h} - 2^h + 1$ avec $\gcd(h, n) = 1$ et $2 \leq h \leq \frac{n-1}{2}$.
- *Welch* : $s = 2^{(n-1)/2} + 3$.
- *Niho* :
  $s = 2^{(n-1)/2} + 2^{(n-1)/4} - 1$, where $n \equiv 1 \pmod 4$
  $s = 2^{(n-1)/2} + 2^{(3n-1)/4} - 1$, where $n \equiv 3 \pmod 4$.

Fonctions APN connues :

- $s = 2^n - 2$ (n impair) ;
- $s = 2^h + 1$ avec $\gcd(h, n) = 1$ et $1 \leq h \leq \frac{n-1}{2}$;
- $s = 2^{2h} - 2^h + 1$ avec $\gcd(h, n) = 1$ et $2 \leq h \leq \frac{n-1}{2}$;
- $s = 2^{\frac{4n}{5}} + 2^{\frac{3n}{5}} + 2^{\frac{2n}{5}} + 2^{\frac{n}{5}} - 1$, avec n divisible par 5 ;
- $s = 2^{(n-1)/2} + 3$ (n impair) ;
- $s = 2^{(n-1)/2} + 2^{(n-1)/4} - 1$, where $n \equiv 1 \pmod 4$;
- $s = 2^{(n-1)/2} + 2^{(3n-1)/4} - 1$, where $n \equiv 3 \pmod 4$.

Nouvelles fonctions (L. Budaghian, C.C., A. Pott) :

Si deux fonctions F et G ont leurs graphes G_F et G_G affinement équivalents (i.e. s'il existe un automorphisme affine A tel que $G_F = A(G_G)$) alors F est AB (resp. APN) si et seulement si G l'est.

L'équivalence linéaire des graphes est une relation plus large que l'équivalence affine des fonctions.

De nouvelles classes de fonctions AB ou APN, équivalentes en ce sens aux fonctions de Gold, mais affinement inéquivalentes à toutes fonctions puissances, ont été déduites de cette observation.

– AB : $F : \mathbb{F}_{2^m} \longrightarrow \mathbb{F}_{2^m}$, $m > 3$ impair,

$F(x) = x^{2^i+1}+(x^{2^i}+x)tr(x^{2^i+1}+x),\ 1 \leq i < \frac{m+1}{2},\ gcd(m,i)$

– APN : $F : \mathbb{F}_{2^m} \longrightarrow \mathbb{F}_{2^m}$, $m \geq 4$ pair,

$F(x) = x^{2^i+1}+(x^{2^i}+x+1)tr(x^{2^i+1}),\ 1 \leq i < \frac{m}{2},\ gcd(m,i) = 1$

– APN : $F : \mathbb{F}_{2^m} \longrightarrow \mathbb{F}_{2^m}$, m divisible par 6,

$F(x) = [x+tr_{m/3}(x^{2(2^i+1)}+x^{4(2^i+1)})+tr(x)tr_{m/3}(x^{2^i+1}+x^{2^{2i}(2^i$

with $gcd(m,i) = 1$, $1 \leq i < \frac{m}{2}$

– AB : $F : \mathbb{F}_{2^m} \longrightarrow \mathbb{F}_{2^m}$, $m > 3$ impair et divisible par 3,

$$F(x) =$$
$$x^{2^i+1} + tr(x^{2^i+1}) + T(x)^6 + T(x)^2 tr(x) + xtr(x) tr_{m/3}(x)^{2^{i+1}}$$
$$+tr(x) tr_{m/3}(x)^{2(2^i+1)} + xT(x)(tr_{m/3}(x)^6 + tr_{m/3}(x) + 1) + x^{2^i} tr(x) tr_{m/3}(x)^2$$
$$+T(x)^{2^i+1}(x + tr_{m/3}(x)) + x^{2^i} T(x)^{2^{i+1}}(tr_{m/3}(x)^4 + tr_{m/3}(x)^3 + 1)$$
$$+T(x)(tr_{m/3}(x)^4 + tr_{m/3}(x)^{2^{i-1}}) + x^{2^i} T(x)^{2^{i-1}}(tr_{m/3}(x)^5 + tr_{m/3}(x)^2 + 1)$$
$$+T(x)^{2^{2i}+1}(x^{2^i} + tr_{m/3}(x)^{2^i}) + T(x)^4(tr_{m/3}(x) + tr_{m/3}(x)^{2^i} + tr_{m/3}(x)^{4(2^i+1)})$$
$$+xT(x)^{2^i}(tr_{m/3}(x)^{2^i+1} + tr_{m/3}(x)^{2^{2i}} + 1),$$

où $T(x) = tr_{m/3}(x^{2^i+1})$, $gcd(m, i) = 1$, $1 \leq i < \frac{m+1}{2}$.

Une notion de nonlinéarité pour les boîtes S des schémas par flots

Pour accélérer le générateur de pseudo-aléas, on peut utiliser une fonction vectorielle pour combiner les sorties des LFSRs.

La fonction F doit être équilibrée, et sans corrélation d'ordre élevé.

Une (n, m)-fonction F est dite *sans corrélation d'ordre* t si sa distribution de valeurs ne change pas lorsque l'on fixe au plus t variables x_i.

Elle est dite t-*résiliente* si elle est équilibrée et sans corrélation d'ordre t.

Sont équivalents :

(1) F est sans corrélation d'ordre t (resp. t-résiliente) ;

(2) pour tout $v \in (F_2^m)^*$, la fonction Booléenne $x \mapsto v \cdot F(x)$ est sans corrélation d'ordre t (resp. t-résiliente) ;

(3) pour toute fonction g Booléenne sur F_2^m, la fonction Booléenne $g \circ F$ est sans corrélation d'ordre t.

On appelle *nonlinéarité uniforme* de F et l'on note UN_F la distance de Hamming minimale entre toutes les fonctions booléennes $g \circ F$ ($g \in {\mathcal{B}_m}^*$) et toutes les fonctions booléennes affines non-constantes ℓ définies sur F_2^n.

Remarque : si $m = n$, alors $UN_F = 0$ pour toute (n, n)-fonction bijective F, alors qu'on peut avoir $N_F = 2^{n-1} - 2^{\frac{n-1}{2}}$.

Bornes sur UN_F :

Si F est équilibrée, on a $UN_F \leq N_F \leq 2^{n-1} - 2^{n/2-1}$.
De plus, si $m < n$, alors

$$UN_F \leq 2^{n-1} - \frac{1}{2} A_{n,m},$$

où

$$A_{n,m} =$$

$$\frac{2^{2m}-2^m}{2^n-1} - \sqrt{\frac{2^{2n}-2^{2n-m}}{2^n-1} + \left(\frac{2^{2m}-2^m}{2^n-1}-1\right)^2} + 1.$$

Cette borne est meilleure que $2^{n-1} - 2^{n/2-1}$ si $m \geq n/2$.

Pour la fonction équilibrée $F(x,y) = \begin{cases} \frac{x}{y} & \text{si } y \neq 0 \\ x & \text{si } y = 0 \end{cases}$, on a

$$UN_F = N_F = 2^{n-1} - 2^{n/2}.$$