CHAPITRE 5.

CRYPTANALYSE DES PROTOCOLES À CLÉ SECRÈTE

http://math.univ-lyon1.fr/~roblot/masterpro.html

Confusion et diffusion

En théorie de l'information, **Shannon** définit deux notions que tout bon cryptosystème doit posséder : la **confusion** et la **diffusion**.

Confusion. C'est le fait que la méthode de calcul du message crypté à partir du message clair doit être suffisamment complexe. C'est-à-dire qu'il ne doit pas exister de relation simples entre les bits du message clair et les bits du message crypté.

Diffusion. C'est le fait qu'une différence, même minime, entre deux messages clairs doit entraîner une très grande différence entre les messages cryptés. Ainsi, chaque bit du message clair doit contribuer au calcul de chaque bit du message crypté.

Soit un cryptosystème donné par la fonction booléenne vectorielle $(f_1, \ldots, f_m)$ sur $\{0,1\}^n$.

Confusion. La confusion correspond au fait que les fonctions f_i n'ont pas d'expression "simple". Si on pose

$$\deg(f_1, \ldots, f_m) = \min(\deg f_1, \ldots, \deg f_m)$$

où $\deg f_i$ est le degré de la représentation algébrique de f_i, on veut avoir

$$\deg(f_1, \ldots, f_m) \approx n.$$

Diffusion. La diffusion correspond au fait que les bits de sortie, donc les valeurs des f_i, varient de manière imprévisible quand on modifie l'entrée. Donc pour tout $x \in \{0,1\}^n$ et tout $\delta \in \{0,1\}^n$ avec $\delta \neq 0$, on veut avoir

$$\mathrm{Prob}\big(f_i(x) = f_i(x \oplus \delta)\big) \approx 1/2$$

pour $1 \leq i \leq m$.

Cryptanalyse linéaire

C'est une attaque à **texte clair connu** contre les protocoles de cryptographie dont la confusion est faible.

Texte clair connu. L'attaquant dispose de un ou plusieurs message(s) clair(s) avec le(s) message(s) crypté(s) correspondant, tous cryptés avec la même clé. L'attaquant cherche à retrouver (de l'information sur) la clé.

Idée. Trouver des *relations linéaires* de dépendance de *probabilités exceptionnelles* entre les bits d'entrée et de sortie.

En effet, une relation linéaire ne peut pas être vraie pour *tous* les messages w sinon **le protocole a une faiblesse.**

Exercice. Montrer que le protocole de *Vigenère* est linéaire.

Fonctions linéaires

Définition. Une fonction booléenne f dont la représentation algébrique est de degré 1 est dite **affine**. Si, de plus, on a $f(0...0) = 0$, on dit que f est **linéaire**.

Fonctions linéaires et produit scalaire. Les fonctions linéaires de $\{0,1\}^n$ sont exactement les fonctions de la forme

$$f(x) = x * v \text{ où } v \in \{0,1\}^n$$

Les fonctions affines de $\{0,1\}^n$ sont exactement les fonctions de la forme

$$f(x) = (x * v) \oplus w \text{ où } v \in \{0,1\}^n \text{ et } w \in \{0,1\}$$

Exercice. Démontrer ces deux résultats et en déduire le nombre de fonctions booléennes linéaires et affines.

Exercice. Soit f une fonction linéaire booléenne sur $\{0,1\}^3$ avec $f(010) = 0$, $f(111) = 1$ et $f(011) = 0$. Retrouver la fonction f.

Résistance linéaire

Exercice. Soit f une fonction booléenne sur $\{0,1\}^n$ et soit $v \in \{0,1\}^n$. Montrer que

$$d(f, x * v) = 2^n - d(f, (x * v) \oplus 1)$$

Montrer que, pour tout $x \in \{0,1\}^n$, on a

$$1 - (-1)^{f(x) \oplus (x * v)} = \begin{cases} 0 & \text{si } f(x) = x * v \\ 2 & \text{sinon} \end{cases}$$

En sommant sur tous les x, en déduire que

$$d(f, x * v) = 2^{n-1} - \frac{1}{2}\tilde{f}(v)$$

Résistance linéaire. On définit la résistance linéaire de f par

$$\mathcal{RL}(f) = 2^{n-1} - \frac{1}{2} \max_{v \in \{0,1\}^n} \left|\tilde{f}(v)\right|$$

Exercice. Montrer que $\mathcal{RL}(f)$ est la distance minimale entre f et une fonction booléenne affine.

Propriétés de la résistance linéaire

Soit f un fonction booléenne sur $\{0,1\}^n$.

Exercice. Montrer que la résistance linéaire est invariante par *transformation affine*, c'est-à-dire, si pour $w \in \{0,1\}^n$, on pose $g(x) = f(x \oplus w)$. Alors, on a $\mathcal{RL}(g) = \mathcal{RL}(f)$

Valeur optimale. Pour n pair, on a

$$\mathcal{RL}(f) \leq 2^{n-1} - 2^{n/2-1}$$

Fonction courbe. Pour n pair, une fonction booléenne f sur $\{0,1\}^n$ est dite *courbe* si on a

$$\left|\tilde{f}(x)\right| = 2^{n/2} \qquad \text{pour tout } x \in \{0,1\}^n$$

Pour une telle fonction, la résistance linéaire est maximale.

Exercice. Construction de fonctions courbes

Soit m un entier positif.

1. Soient $x_1, x_2, y_1, y_2 \in \{0,1\}^m$. Montrer que
$$(x_1 \cdot x_2) * (y_1 \cdot y_2) = (x_1 * y_1) \oplus (x_2 * y_2)$$
2. Soit g une fonction booléenne sur $\{0,1\}^m$ et π une FBV de $\{0,1\}^m$ dans $\{0,1\}^m$. On définit une fonction booléenne f sur $\{0,1\}^{2m}$ en posant
$$f(x_1 \cdot x_2) = (x_1 * \pi(x_2)) \oplus g(x_2) \text{ où } x_1, x_2 \in \{0,1\}^m$$
    1. Pour $x_1 \in \{0,1\}^m$, on pose $P(x_1) = \{y_1 \in \{0,1\}^m \text{ tel que } \pi(y_1) = x_1\}$. Montrer que
    $$\tilde{f}(x) = 2^m \sum_{u_2 \in P(x_1)} (-1)^{g(u_2) \oplus (x_2 * u_2)} \text{ où } x = x_1 \cdot x_2$$
    2. En déduire que f est courbe si π est une permutation.
    3. En déduire que f n'est pas courbe si π n'est pas une permutation.
    4. Construire une fonction courbe sur $\{0,1\}^4$.

Résistance linéaire des fonctions booléennes vectorielles

Définition. Soit F une FBV de $\{0,1\}^n$ dans $\{0,1\}^m$, on définit la résistance linéaire de f par

$$\mathcal{RL}(F) = \min_{\substack{v\in\{0,1\}^m \\ v\neq 0\ldots 0}} \mathcal{RL}(v * F)$$

où $v * F$ est la fonction booléenne définie par $(v * F)(x) = v * f(x)$.

Exercice. Calculer la résistance vectorielle de la FBV suivante

$$F(a,b,c) = \big((a \wedge b) \oplus \neg c, (\neg a \wedge b) \vee (a \oplus b \oplus c)\big)$$

Résultat. On a les propriétés suivantes :

$$\mathcal{RL}(F) = 2^{n-1} - \frac{1}{2} \max_{\substack{u\in\{0,1\}^n \\ v\in\{0,1\}^m\setminus\{0\ldots 0\}}} \left|\widetilde{(v * f)}(u)\right|$$

$$\mathcal{RL}(F) \leq 2^{n-1} - 2^{n/2-1} \quad \text{(pour } n \text{ pair)}$$

Biais et principe d'empilement

Définitions. On appelle *variable aléatoire binaire* une variable aléatoire à valeurs dans $\{0,1\}$. On appelle **biais** de la variable aléatoire binaire X la quantité

$$\varepsilon(X) = P(X=0) - 1/2$$

On a $\varepsilon(X) = 0$ ssi X est totalement aléatoire et plus la valeur de $\varepsilon(X)$ est grande (en valeur absolue), moins les valeurs de X sont aléatoires.

Exercice.

1. Soient X_1 et X_2 deux variables aléatoires binaires. Montrer que si X_1 et X_2 sont indépendantes, on a alors

$$\varepsilon(X_1 \oplus X_2) = \frac{1}{2}\varepsilon(X_1)\varepsilon(X_2)$$

2. Soient X_1, X_2 et X_3 trois variables aléatoires binaires indépendantes. Montrer que $X_1 \oplus X_2$ et $X_2 \oplus X_3$ sont indépendantes. En déduire que

$$\varepsilon(X_1 \oplus X_3) = 2\,\varepsilon(X_1 \oplus X_2)\,\varepsilon(X_2 \oplus X_3)$$

Biais linéaire d'une FBV

Biais linéaire. Soit F une FBV de $\{0,1\}^n$ dans $\{0,1\}^m$. Pour $u \in \{0,1\}^n$ et $v \in \{0,1\}^m$, on définit le biais linéaire suivant (u,v) de F par

$$\varepsilon(F;u,v) = \varepsilon(\big(F(x) * v\big) \oplus (x * u))$$

Propriétés.

$$\mathcal{RL}(F) = 2^n \left(\frac{1}{2} - \max_{\substack{u \in \{0,1\}^n \\ v \in \{0,1\}^m \setminus \{0\ldots0\}}} |\varepsilon(F;u,v)| \right)$$

Pour n pair, il existe toujours $u \in \{0,1\}^n$ et $v \in \{0,1\}^m \setminus \{0\ldots0\}$ tels que

$$|\varepsilon(F;u,v)| \geq \frac{1}{2^{n/2}+1}$$

Exercice. Démontrer ces propriétés.

Exercice. Biais linéaire d'une S-boîte

On considère la S-boîte (FBV bijective) sur 4 bits suivante :

x	$S(x)$	x	$S(x)$	x	$S(x)$	x	$S(x)$
0000	1110	0100	0010	1000	0011	1100	0101
0001	0100	0101	1111	1001	1010	1101	1001
0010	1101	0110	1011	1010	0110	1110	0000
0011	0001	0111	1000	1011	1100	1111	0111

1. Calculer le biais $\varepsilon(S; u, v)$ pour les couples (u, v) suivants

$$(0110, 1011), \quad (1001, 0100), \quad (0011, 1001)$$

2. Que peut-on en déduire sur la résistance linéaire de S ?

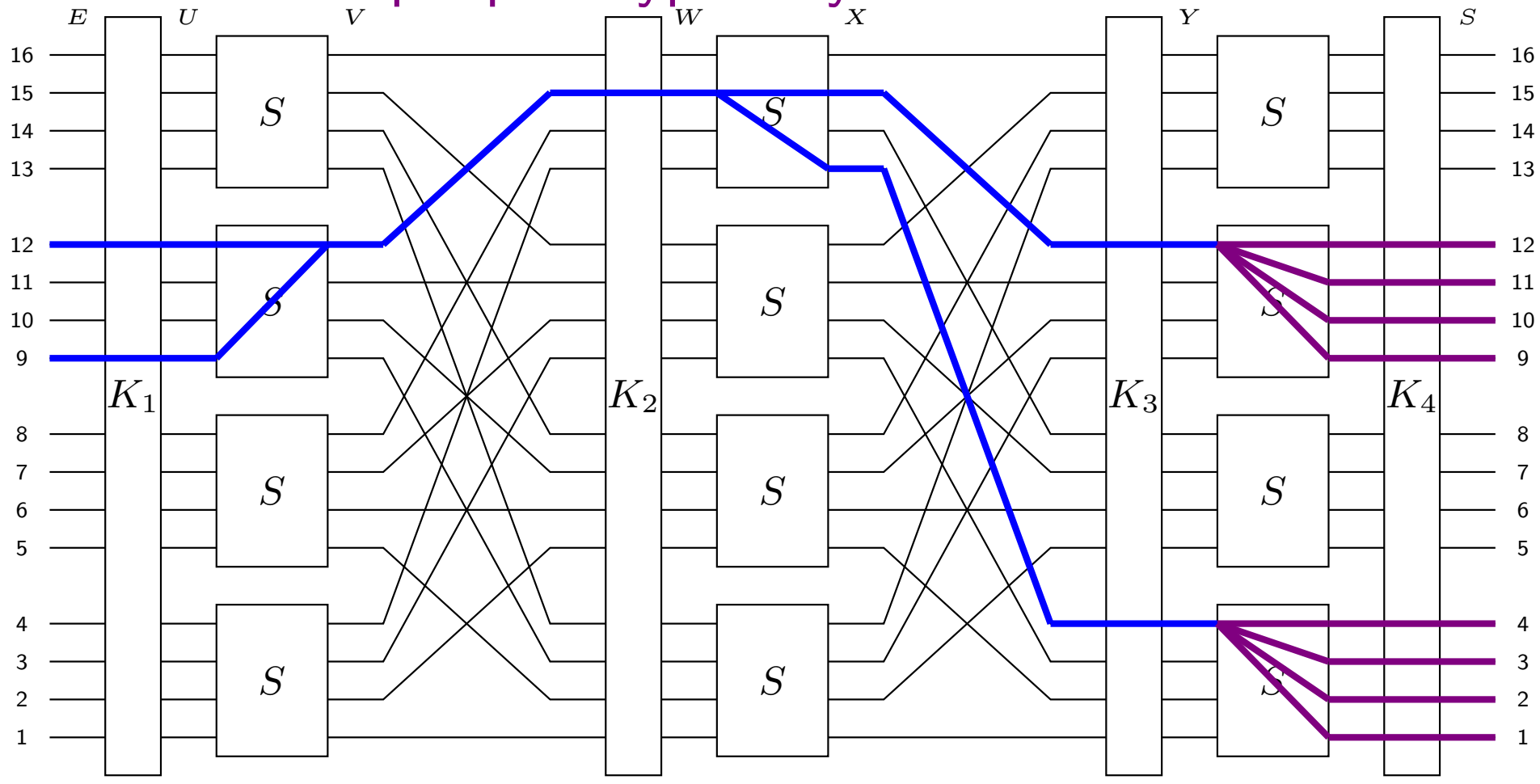
Exemple sur 4 bits : Table d'approximation linéaire

$2^4 \times \varepsilon(S;u,v)$		u 1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
	1	0	−2	−2	0	0	−2	+6	+2	+2	0	0	+2	+2	0	0
	2	0	−2	−2	0	0	−2	−2	0	0	+2	+2	0	0	−6	+2
	3	0	0	0	0	0	0	0	+2	−6	−2	−2	+2	+2	−2	−2
	4	+2	0	−2	−2	−4	−2	0	0	−2	0	+2	+2	−4	+2	0
	5	−2	−2	0	−2	0	+4	+2	−2	0	−4	+2	0	−2	−2	0
	6	+2	−2	+4	+2	0	0	+2	0	−2	+2	+4	−2	0	0	−2
v	7	−2	0	+2	+2	−4	+2	0	−2	0	+2	0	+4	+2	0	+2
	8	0	0	0	0	0	0	0	−2	+2	+2	−2	+2	−2	−2	−6
	9	0	−2	−2	0	0	−2	−2	−4	0	−2	+2	0	+4	+2	−2
	A	+4	−2	+2	−4	0	+2	−2	+2	+2	0	0	+2	+2	0	0
	B	+4	0	−4	+4	0	+4	0	0	0	0	0	0	0	0	0
	C	−2	+4	−2	−2	0	+2	0	+2	0	+2	+4	0	+2	0	−2
	D	+2	+2	0	−2	+4	0	+2	−4	−2	+2	0	+2	0	0	+2
	E	+2	+2	0	−2	−4	0	+2	−2	0	0	−2	−4	+2	−2	0
	F	−2	−4	−2	−2	0	+2	0	0	−2	+4	−2	−2	0	+2	0

Exercice.

1. Montrer que la somme de chaque ligne et chaque colonne est ± 8.
2. Pour quelles valeurs de (u, v), le biais linéaire est-il maximal (en valeur absolue) ?
3. Quelle est la résistance linéaire de S ?

Exercice. Attaque par cryptanalyse linéaire



On considère le cryptosystème sur 16 bits ci-dessus qui comporte 3 rondes composé de la même S-boîte répétée et d'addition de sous-clés.

Exercice. Attaque par cryptanalyse linéaire (suite)

1. Justifier pourquoi il est nécessaire d'ajouter une sous-clé au début et à la fin du cryptosystème.
2. Déterminer les biais linéaires $\varepsilon(S; u, v)$ pour les valeurs suivantes

$$(1001, 1000) \quad \text{et} \quad (0100, 0101)$$

3. On note $E_1, \ldots, E_{16}$ les bits d'entrée, $S_1, \ldots, S_{16}$ les bits de sortie, et plus généralement U_i, V_i, etc., les bits aux différentes étapes (cf. diagramme). Les mots sont composés en prenant les bits de droite à gauche. Ainsi, les mots à l'entrée et à la sortie de la première S-boîte en bas à gauche sont $U_4U_3U_2U_1$ et $V_4V_3V_2V_1$.

   En utilisant la question précédente, trouver une relation linéaire entre les bits E_9, E_{12}, W_{15} et certains bits des sous-clés K_1, K_2 vérifiée avec une grande probabilité. Puis, une relation linéaire vérifiée avec une grande ou faible probabilité entre E_9, E_{12}, Y_4 et Y_{12}.

Exercice. Attaque par cryptanalyse linéaire (suite)

4. Soit (E, S) un couple formé d'un message clair et d'un message crypté. On calcule Y_4 et Y_{12} en partant de S et en remontant le diagramme. Calculer la probabilité que

$$E_9 \oplus E_{12} = Y_4 \oplus Y_{12}$$

en distinguant les cas où K_4 est ou n'est pas la sous-clé ayant servi à crypter S.

5. En déduire comment, à partir d'un grand nombre de couples messages clairs / messages cryptés avec la même clé, on peut obtenir des informations sur certains bits de la sous-clé K_4.