Introduction à la cryptographie

Vanessa VITSE

Université Grenoble Alpes

M1 Maths 2021

Section 8

Générateurs de nombres pseudo-aléatoires

PRNG

Nécessité de générer des bits aléatoires

- génération de clefs ou de nonces
- pour transformer une graine secrète en une longue suite de bits pseudo-aléatoire (chiffrement par flux)
- mais aussi dans d'autres disciplines (simulations, Monte-Carlo, etc...)

PRNG

Nécessité de générer des bits aléatoires

- génération de clefs ou de nonces
- pour transformer une graine secrète en une longue suite de bits pseudo-aléatoire (chiffrement par flux)
- mais aussi dans d'autres disciplines (simulations, Monte-Carlo, etc...)

seulement du pseudo-aléatoire réalisable sur machine

Définition

Un PRNG est un algorithme qui prend en entrée une valeur initiale (la graine) et retourne une suite de nombres

- qualité d'un PRNG à l'aide tests statistiques permettant de le distinguer d'une vraie suite aléatoire (partie non traitée dans ce cours)
- limitation par nature puisque le PRNG est déterministe !

Période du PRNG

Propriété

La suite (u_n) produite par un PRNG (qui utilise une quantité finie de mémoire) est ultimement périodique :

$$\exists T > 0, N \in \mathbb{N}, \forall n \geqslant N, \quad u_{n+T} = u_n.$$

Sécurité

- avoir une longue période (relativement à l'usage mémoire) ; exemple du PRNG Mersenne Twister de période $2^{19937} - 1$ avec mémoire 2.5 Ko
- étant donnés $u_k, u_{k+1}, \ldots, u_{k+m}$, il doit être difficile de retrouver de l'information sur le terme suivant u_{k+m+1} même si m est grand

En pratique : on utilise des générateurs d'aléa plus compliqués, les PRNG ne sont qu'une première approche de ce problème

Registres à décalage linéaire (LFSR)

- exemple classique de PRNG
- étaient utilisés dans les chiffrements par flux (CSS pour DVD, ou RC4 pour web)
- peu coûteux (utiles dans les circuits embarqués) mais doivent être combinés avec fonctions booléennes pour la sécurité

Définition

Un LFSR de longueur L est un registre à décalage contenant une suite $(s_i, \ldots, s_{i+L-1})$ de L bits et une fonction de rétroaction linéaire :

- à chaque cycle d'horloge, le LFSR retourne le bit de poids faible
- le nouveau bit de poids fort du registre est donné par

$$s_{i+L} = c_1 s_{i+L-1} + c_2 s_{i+L-2} + \cdots + c_{L-1} s_{i+1} + c_L s_i$$

où les $c_i \in \{0; 1\}$ sont appelés *coefficients de rétroaction* du LFSR.

Vision Hardware d'un LFSR

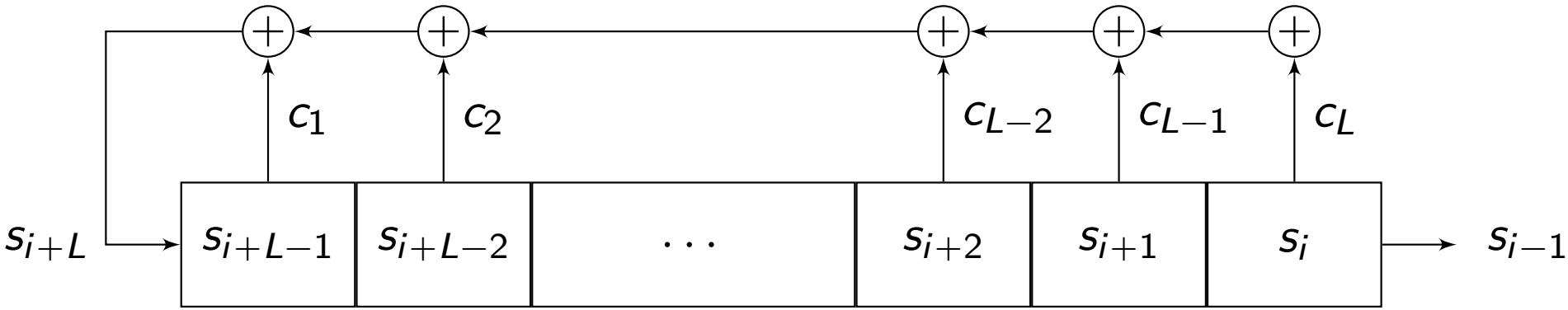
Algorithm 1: LFSR

Input : $(s_0, \ldots, s_{L-1}) \in \mathbb{F}_2^L \backslash \{0\}$

for $i = 0, \ldots, N-1$ **do**

  output s_i

  $s_{i+L} \leftarrow c_1 s_{i+L-1} + c_2 s_{i+L-2} + \cdots + c_{L-1} s_{i+1} + c_L s_i$



Polynômes de rétroaction

Définition

Le polynôme de rétroaction du LFSR est

$$C(X) = 1 - c_1 X - c_2 X^2 - \cdots - c_L X^L,$$

son polynôme réciproque $P(X) = X^L - c_1 X^{L-1} - \cdots - c_{L-1} X - c_L$ est appelé *polynôme caractéristique*.

Proposition

- La suite (s_i) est ultimement périodique de période $T \leqslant 2^L - 1$
- Si $c_L \neq 0$, alors la suite est périodique.

Polynômes de rétroaction

Définition

Le polynôme de rétroaction du LFSR est

$$C(X) = 1 - c_1 X - c_2 X^2 - \cdots - c_L X^L,$$

son polynôme réciproque $P(X) = X^L - c_1 X^{L-1} - \cdots - c_{L-1} X - c_L$ est appelé *polynôme caractéristique*.

Proposition

- La suite (s_i) est ultimement périodique de période $T \leqslant 2^L - 1$
- Si $c_L \neq 0$, alors la suite est périodique.

Remarque : si $c_L = 0$ la suite est générée par une séquence plus courte...
dans la suite on suppose $c_L \neq 0$

Périodicité du LFSR

Soit $R_i = (s_i, \ldots, s_{i+L-1}) \in \mathbb{F}_2^L$ l'état du registre à la i-ème itération.

- si $\exists i_0$ tel que $R_{i_0} = (0, \ldots, 0)$, alors $s_i = 0$ et $T = 1$
- autrement il y a $2^L - 1$ valeurs possibles du registre donc parmi $R_0, \ldots, R_{2^L-1}$ au moins deux sont égales et $0 < T \leqslant 2^L - 1$
- si $c_L \neq 0$, alors $s_i = \frac{1}{c_L}\left(s_{i+L} - c_1 s_{i+L-1} - \cdots - c_{L-1} s_{i+1}\right)$ donc le i-ème terme est déterminé par les L suivants et $s_i = s_{i+T}$ est alors vrai pour tout i.

Quelques propriétés des LFSR de période maximale

Soit un LFSR d'état initial $R_0 \in \mathbb{F}_2^L$ dont la sortie est une suite (s_i) de période $2^L - 1$

Remarque

La période étant $2^L - 1$, on a $\{R_0, R_1, \ldots, R_{2^L-2}\} = \mathbb{F}_2^L \backslash \{(0, \ldots, 0)\}$ (registres distincts et non nuls).

Période et polynôme de rétroaction

La période ne dépend que du polynôme de rétroaction et non de la graine : Soit $R'_0 \neq (0, \ldots, 0)$, la sortie (s'_i) correspondante est de période $2^L - 1$ et il existe $\tau \in \mathbb{N}$ tel que $(s'_i) = (s_{i+\tau})$.

Preuve

Soit R'_0 un registre initial non nul. Alors il existe $\tau \in \{0, \ldots, 2^L - 2\}$ tel que $R'_0 = R_\tau$ et $\forall i, R'_i = R_{i+\tau}$, i.e. $(s'_i) = (s_{i+\tau})$.

Résultats statistiques

Équidistribution

Dans une période :

- la suite (s_i) prend exactement 2^{L-1} fois la valeur 1 et $2^{L-1}-1$ fois la valeur 0
- plus généralement toute séquence de k bits, $1 \leqslant k \leqslant L$, apparaît exactement 2^{L-k} fois (ou $2^{L-k}-1$ fois pour la séquence de k zéros)

Résultats statistiques

Équidistribution

Dans une période :

- la suite (s_i) prend exactement 2^{L-1} fois la valeur 1 et $2^{L-1}-1$ fois la valeur 0
- plus généralement toute séquence de k bits, $1 \leqslant k \leqslant L$, apparaît exactement 2^{L-k} fois (ou $2^{L-k}-1$ fois pour la séquence de k zéros)

Preuve

- Soit $(b_0, \ldots, b_{k-1}) \neq (0, \ldots, 0)$ avec $1 \leqslant k \leqslant L$.
  Alors il y a 2^{L-k} éléments de $\mathbb{F}_2^L \backslash \{0\}$ démarrant par $(b_0, \ldots, b_{k-1})$.
  D'après la remarque, il y a donc 2^{L-k} valeurs de i telles que R_i commence par cette séquence, donc telles que $(s_i, \ldots, s_{i+k-1}) = (b_0, \ldots, b_{k-1})$.
- Même type de raisonnement avec la séquence de k bits à 0

Résultats statistiques

Auto-corrélation

$\forall \tau \neq 0 \bmod 2^L - 1, |\{i \in \{0; 2^L - 2\} : s_i \neq s_{i+\tau}\}| = 2^{L-1}$
et $|\{i \in \{0; 2^L - 2\} : s_i = s_{i+\tau}\}| = 2^{L-1} - 1$

Résultats statistiques

Auto-corrélation

$\forall \tau \neq 0 \bmod 2^L - 1, |\{i \in \{0; 2^L - 2\} : s_i \neq s_{i+\tau}\}| = 2^{L-1}$
et $|\{i \in \{0; 2^L - 2\} : s_i = s_{i+\tau}\}| = 2^{L-1} - 1$

Preuve : soit $\tau \neq 0 \bmod 2^L - 1$.

- $R_0 \neq R_\tau$ donc $R'_0 = R_0 - R_\tau$ est un registre non nul
- la suite $(s'_i) = (s_i - s_{i+\tau})$ est la sortie correspondante du LFSR avec état initial R'_0 (car satisfait la même relation de récurrence que (s_i)) $\rightarrow$ c'est donc un décalage de (s_i)
- Par équidistribution elle contient 2^{L-1} bits à 1 et $2^{L-1} - 1$ bits à 0 dans une période.

Représentation de Fibonacci

Soient $R_i = \begin{pmatrix} s_i \\ s_{i+1} \\ \vdots \\ s_{i+L-1} \end{pmatrix}$ et $A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \\ c_L & \dots & \dots & c_2 & c_1 \end{pmatrix} \in \mathcal{M}_L(\mathbb{F}_2)$ la matrice compagnon du polynôme caractéristique du LFSR.

Alors $R_{i+1} = AR_i = A^i R_0$ et pour tout $i \in \mathbb{N}$

$$s_i = (1\ 0 \dots 0) \begin{pmatrix} s_i \\ s_{i+1} \\ \vdots \\ s_{i+L-1} \end{pmatrix} = (1\ 0 \dots 0) A^i R_0.$$

Représentation de Galois

Idée : faire un changement de base pour obtenir de l'information sur la période du LFSR

- $\mathcal{B}_c = (e_1, \ldots, e_L)$ la base canonique de $\mathbb{F}_2^L$
- $f \in \mathrm{End}(\mathbb{F}_2^L)$ tel que $A = \mathrm{Mat}_{\mathcal{B}_c} f$

Alors $\mathcal{B} = (e_L, f(e_L), \ldots, f^{L-1}(e_L))$ est une autre base de $\mathbb{F}_2^L$ car la matrice de passage de $\mathcal{B}_c$ vers $\mathcal{B}$ est

$$M = \begin{pmatrix} 0 & \ldots & 0 & 1 \\ \vdots & \ddots & \ddots & * \\ 0 & \ddots & \ddots & \vdots \\ 1 & * & \ldots & * \end{pmatrix}, \quad \text{et} \quad M^{-1}AM = \mathrm{Mat}_{\mathcal{B}}(f) = \begin{pmatrix} 0 & \ldots & \ldots & 0 & * \\ 1 & \ddots & & \vdots & \vdots \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & * \\ 0 & \ldots & 0 & 1 & * \end{pmatrix}$$

est une matrice compagnon qui a le même polynôme caractéristique que A donc $M^{-1}AM = {}^tA$.

Représentation de Galois

Idée : faire un changement de base pour obtenir de l'information sur la période du LFSR

- $\mathcal{B}_c = (e_1, \ldots, e_L)$ la base canonique de $\mathbb{F}_2^L$
- $f \in \mathrm{End}(\mathbb{F}_2^L)$ tel que $A = \mathrm{Mat}_{\mathcal{B}_c} f$

Alors $\mathcal{B} = (e_L, f(e_L), \ldots, f^{L-1}(e_L))$ est une autre base de $\mathbb{F}_2^L$ car la matrice de passage de $\mathcal{B}_c$ vers $\mathcal{B}$ est

$$M = \begin{pmatrix} 0 & \ldots & 0 & 1 \\ \vdots & \ddots & \ddots & * \\ 0 & \ddots & \ddots & \vdots \\ 1 & * & \ldots & * \end{pmatrix}, \quad \text{et} \quad M^{-1}AM = \mathrm{Mat}_{\mathcal{B}}(f) = \begin{pmatrix} 0 & \ldots & \ldots & 0 & c_L \\ 1 & \ddots & & \vdots & \vdots \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & c_2 \\ 0 & \ldots & 0 & 1 & c_1 \end{pmatrix}$$

est une matrice compagnon qui a le même polynôme caractéristique que A donc $M^{-1}AM = {}^tA$.

Représentation de Galois

Idée : faire un changement de base pour obtenir de l'information sur la période du LFSR

Rappel : pour tout $i \in \mathbb{N}$ on a $s_i = (1\ 0 \dots 0) A^i R_0$
Donc $s_i = (1\ 0 \dots 0)\,(M\,{}^tA M^{-1})^i R_0 = (1\ 0 \dots 0)\, M({}^tA)^i M^{-1} R_0$

Comme $(1\ 0 \dots 0)\, M = (0 \dots 0\ 1)$, on trouve

La suite récurrente linéaire (s_i) peut s'obtenir en considérant la dernière coordonnée des vecteurs $Y_i = ({}^tA)^i Y_0$ dont on note les coordonnées $(q_{i,0}, q_{i,1}, \dots, q_{i,L-1})$ avec $Y_0 = M^{-1} R_0$.

Représentation de Galois

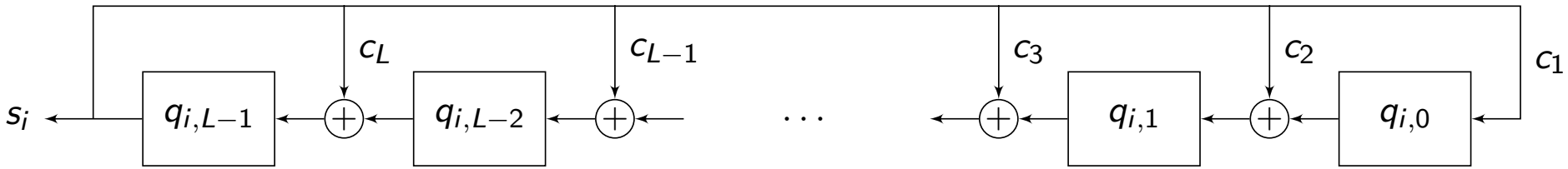
La suite récurrente linéaire (s_i) peut s'obtenir en considérant la dernière coordonnée des vecteurs $Y_i = ({}^tA)^i Y_0$ dont on note les coordonnées $(q_{i,0}, q_{i,1}, \ldots, q_{i,L-1})$ avec $Y_0 = M^{-1} R_0$.

Interprétation hardware (LFSR "dual") :

$$\begin{pmatrix} q_{i+1,0} \\ q_{i+1,1} \\ \vdots \\ q_{i+1,L-1} \end{pmatrix} = \begin{pmatrix} 0 & \ldots & \ldots & 0 & c_L \\ 1 & \ddots & & \vdots & \vdots \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & c_2 \\ 0 & \ldots & 0 & 1 & c_1 \end{pmatrix} \begin{pmatrix} q_{i,0} \\ q_{i,1} \\ \vdots \\ q_{i,L-1} \end{pmatrix}$$

Représentation de Galois

X^L=c1X^{L-1}+...+cL



Soit $Q_i = q_{i,0} + q_{i,1}X + \cdots + q_{i,L-1}X^{L-1}$, alors

$$Q_{i+1} = XQ_i \bmod P, \text{ où } P(X) = X^L - c_1X^{L-1} - \cdots - c_L$$

Autrement dit, l'endomorphisme associé à ${}^tA$ dans la base $(1, X, X^2, \ldots, X^{L-1})$ est

$$\begin{array}{ccc} \mathbb{F}_2[X]/\langle P(X)\rangle & \mapsto & \mathbb{F}_2[X]/\langle P(X)\rangle \\ Q & \mapsto & XQ \end{array}$$

et Q_i est le reste dans la division euclidienne de X^iQ_0 par $P(X)$.

La suite pseudo-aléatoire (s_i) s'obtient en considérant le coefficient de degré $L-1$ dans la suite de polynômes $(Q_i) = (X^iQ_0 \bmod P(X))$

LFSR de période maximal

L'état interne d'un LFSR est donné par le polynôme $Q_i = X^i Q_0 \bmod P(X)$

- la période du LFSR est le plus petit entier $T > 0$ tel que $X^T Q_0 = Q_0 \bmod P(X)$
- si $Q_0 \wedge P = 1$, alors la période est l'ordre (multiplicatif) de X dans le groupe des éléments inversibles de l'anneau $\mathbb{F}_2[X]/(P(X))$

LFSR de période maximal

L'état interne d'un LFSR est donné par le polynôme $Q_i = X^i Q_0 \bmod P(X)$

- la période du LFSR est le plus petit entier $T > 0$ tel que $X^T Q_0 = Q_0 \bmod P(X)$
- si $Q_0 \wedge P = 1$, alors la période est l'ordre (multiplicatif) de X dans le groupe des éléments inversibles de l'anneau $\mathbb{F}_2[X]/(P(X))$

Théorème

Un LFSR est de période maximale $2^L - 1$ si et seulement si son polynôme caractéristique $P(X)$ est irréductible et la classe de X est un générateur du groupe multiplicatif de $\mathbb{F}_2[X]/(P(X))$.

Indication pour la preuve :
montrer que $\mathrm{ord}(X) \leqslant |\mathbb{F}_2[X]/(P(X))|^\times \leqslant 2^L - 1$ et

$$|\mathbb{F}_2[X]/(P(X))|^\times \leqslant 2^L - 1 \iff P \text{ est irréductible sur } \mathbb{F}_2[X].$$

LFSR de période maximal

L'état interne d'un LFSR est donné par le polynôme $Q_i = X^i Q_0 \bmod P(X)$

- la période du LFSR est le plus petit entier $T > 0$ tel que $X^T Q_0 = Q_0 \bmod P(X)$
- si $Q_0 \wedge P = 1$, alors la période est l'ordre (multiplicatif) de X dans le groupe des éléments inversibles de l'anneau $\mathbb{F}_2[X]/(P(X))$

Théorème

Un LFSR est de période maximale $2^L - 1$ si et seulement si son polynôme caractéristique $P(X)$ est irréductible et la classe de X est un générateur du groupe multiplicatif de $\mathbb{F}_2[X]/(P(X))$.

Un tel polynôme P est appelé *polynôme primitif*.
Pour en trouver un de degré L, chercher un générateur de $(\mathbb{F}_{2^L})^*$ et prendre son polynôme minimal.

Exercices

1. Montrer que le polynôme $Q = X^4 + X^3 + X^2 + X + 1 \in \mathbb{F}_2[X]$ est irréductible. Est-ce que la classe x de X est un générateur du groupe multiplicatif de $\mathbb{F}_{2^4} = \mathbb{F}_2[X]/(Q)$?

2. Soit $y = 1 + x \in \mathbb{F}_{2^4}$. Montrer que y est un générateur de $\mathbb{F}_{2^4}^*$ et trouver son polynôme minimal P.

3. Calculer la sortie du LFSR correspondant et vérifier qu'il satisfait les propriétés statistiques vues en cours.

Période d'un LFSR

Soit $P(X) = \prod_{j=1}^{m} P_j(X)^{\alpha_j}$ la décomposition en facteurs premiers d'un polynôme caractéristique d'un LFSR.

Théorème des restes chinois

$\mathbb{F}_2[X]/(P(X)) \simeq \prod_{j=1}^{r} \mathbb{F}_2[X]/(P_j(X)^{\alpha_j})$

- l'ordre multiplicatif de X dans $\mathbb{F}_2[X]/(P(X))$ est le ppcm des ordres de X dans $\mathbb{F}_2[X]/(P_j(X)^{\alpha_j})$
- si P n'a pas de facteurs carrés, tous les $\mathbb{F}_2[X]/(P_j(X))$ sont des corps et l'ordre de X dans chacun de ces corps est un diviseur de $2^{\deg P_j} - 1$

Exercice

1. Soit $P = X^5 + X^4 + 1 = (X^3 + X + 1)(X^2 + X + 1)$. Trouver la période de la sortie du LFSR correspondant quand $Q_0 = 1$ et quand $Q_0 = X^3 + 1$.

2. Quand P n'est pas sans carré, il peut être plus difficile de déterminer la période. Deux exemples :
   1. Soit $P = (X + 1)^2 = X^2 + 1$. Selon la valeur Q_0, trouver toutes les périodes possibles.
   2. Même question avec $P = (X^2 + X + 1)^2 = X^4 + X^2 + 1$.

3. Déterminer toutes les périodes possibles de la sortie d'un LFSR de longueur $L = 4$.

Cryptanalyse d'un LFSR

Sécurité d'un LFSR

Un LFSR n'est pas sûr cryptographiquement :
on peut retrouver son polynôme de rétroaction à partir de $2L$ bits consécutifs (algèbre linéaire ou mieux algorithme de Berlekamp-Massey)

Cryptanalyse d'un LFSR

Sécurité d'un LFSR

Un LFSR n'est pas sûr cryptographiquement :
on peut retrouver son polynôme de rétroaction à partir de $2L$ bits consécutifs (algèbre linéaire ou mieux algorithme de Berlekamp-Massey)

Soit $(s_i)_{0 \leqslant i < 2L}$ des bits en sortie d'un LFSR de longueur $\leqslant L$ et $P = \sum_{k=0}^{L} a_k X^k$ son polynôme caractéristique. Alors le système

$$\begin{pmatrix} u_0 & u_1 & \dots & u_L \\ u_1 & u_2 & \dots & u_{L+1} \\ \vdots & & & \vdots \\ u_{L-1} & u_L & \dots & u_{2L-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_L \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

permet de retrouver P

Usage en cryptographie

Plusieurs solutions :

- combiner non linéairement les sorties de plusieurs LFSR (chiffrement A5/1 et E0 pour GSM eet Bluetooth)

- *shrinking generator* : étant donnés deux LFSR A et S, renvoyer seulement les bits de A pour lesquels le bit correspondant de S vaut 1
  Exemple si les sorties de A et S sont

$$S : \ 1001010111010101001$$
$$A : \ 0011011110010100$$

  la sortie du shrinking generator est 011110000