Cryptanalyse

Cours 2 - Chiffrement par Flot (Stream Ciphers)

Maxime Bombar

13 Septembre 2024

Introduction

Rappel de la semaine dernière : One-Time-Pad

- Chiffrement sûr au sens de la théorie de l'information (*Perfect secrecy*)
- Clés nécessitent une grosse entropie :
  e.g. clés aléatoires et aussi longues que le message
- Pas adaptés pour tous les contextes.

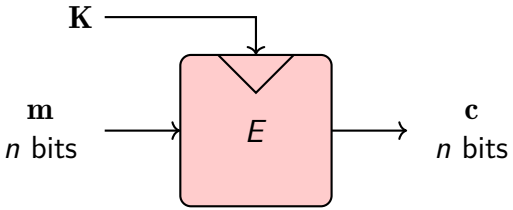
Mais les adversaires ont des ressources *bornées*.

$\rightarrow$ Chiffrement par flot : Clé **pseudo-aléatoire**.

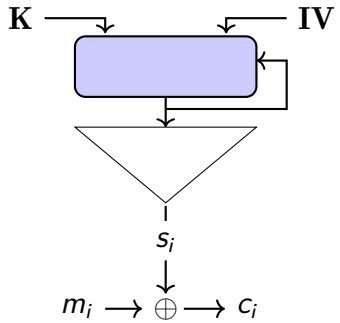
Objectifs des deux prochaines séances

- Générateurs Pseudoaléatoires en Cryptographie
- *Linear Feedback Shift Registers* (LFSR)
- Cryptanalyse
- Exemples de design

Deux Paradigmes de Chiffrement Symétrique



Message découpés en « gros » blocs
Chiffrés indépendamment



Chaque symbole du message
Chiffré par un élément
pseudo-aléatoire et
indépendants du clair

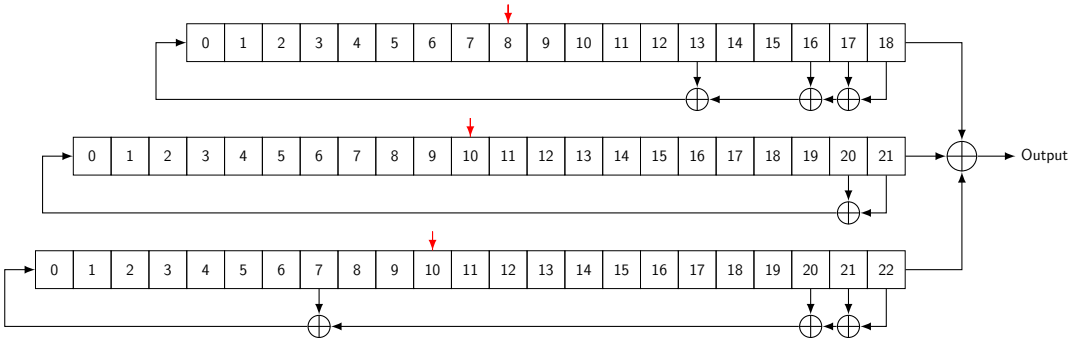
Quelques Propriétés des Chiffrements par Flot

- PRG implantés en **hardware**.
- En général, extrêmement rapides.
- Peu gourmands en énergie.
- Sécurité équivalente à celle du PRG.

$\rightarrow$ Utilisés sur des terminaux aux ressources limitées.

- WEP (Wifi, norme IEEE 802.11, 1999) : Chiffrement RC4.
- Bluetooth : Chiffrement E0.
- 2G/3G : Chiffrement A5/1.
- 4G/5G : Chiffrement SNOW.
- Communications radio par l'armée (utilisation historique).

Chiffrement par Flot A5/1



Attaques extrêmement efficaces (quelques secondes d'écoute du traffic, quelques minutes de temps de calcul).

Déclin du Chiffrement par Flot (1970's – 2000's)

- Télécommunications : Analogique → Découpage par paquets.
- DES (Data Encryption Standard), standardisé en 1976.
- Pas de standards pour les chiffrements par flot.
- Sécurité par l'obscurantisme (contraire à Kirchhoff !).
- Cryptanalyse moderne (linéaire, différentielle, algébrique...) s'applique aussi.
- + Cryptanalyse spécifique des chiffrements par flots.

- RC4 rétroingéniéré en 1994, spécifications sur une *mailing list* : Cipherpunks.
- WEP utilise RC4, mais en plus mal → Surnommé Weak Encryption Protocol (cryptanalyse complète dès 2001).
- `https://github.com/aircrack-ng/aircrack-ng/`

Stream Ciphers: Dead or Alive?

Adi Shamir
Computer Science Dept
The Weizmann Institute
Israel

ASIACRYPT 2004

Regain d'attractivité ?

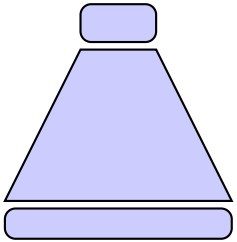
- Compétition eSTREAM (2004-2008) $\rightarrow$ 7 chiffrements, dont Salsa20.
- Renaissance des besoins « ressources limités » et « sobriété »

- ChaCha20 : Variante de Salsa20 avec de meilleures performances.
- Fait partie de TLS1.3. Utilisé notamment dans DNS over TLS.

```
kdig +tls u-bordeaux.fr @80.67.169.12        // DNS de la FDN, FAI Associatif.
```

PseudoRandom (Number) Generators

Caractéristiques d'un PRG



- Algorithme **déterministe** pour générer de l'aléa.
- Sortie entèrement déterminée par la graîne initiale.
- Sortie doit **ressembler** à une chaîne uniformément aléatoire.
- Bonus : Rapide.

- Bonnes propriétés statistiques $\rightarrow$ **pas suffisant**
- Nécessite **Non predictabilité** :

  $(s_n, \ldots, s_{n+m})$ ne doit pas permettre de retrouver s_{n-1} ou s_{n+m+1} efficacement.

Une remarque importante

Un PRG est une **machine à états finis**, dont l'état interne S_t change à chaque tic d'horloge et qui produit un symbole s_t en fonction de S_t.

Soit ℓ la taille de l'état interne. Alors le PRG peut prendre **au plus** 2^ℓ états internes différents, et la suite (s_t) est **ultimement périodique** de période au plus 2^ℓ.

On va chercher à produire des suites chiffrantes de **période maximale**.

Vecteur d'Initialisation (IV)

Réutiliser la même clé secrète résulte en **la même suite chiffrante**.
$\rightarrow$ même problème que le *One-Time Pad*.

- L'état interne initial S_0 est déterminé à partir de la clé privée K et d'une entrée auxiliaire **publique** appelée **vecteur d'initialisation** (IV).
- L'IV est changé régulièrement et permet de réutiliser la même clé K.

Éléments constitutifs d'un PRG

- **Fonction d'initialisation** : Détermine l'état initial S_0 à partir de la clé privée et de l'IV.
- **Fonction de transition** : Modifie la valeur de l'état interne : $S_{t+1} = \Phi(S_t)$ à chaque tic d'horloge. En d'autres termes, $S_t = \Phi^t(S_0)$ (avec $\Phi^{t+1} \stackrel{\text{def}}{=} \Phi \circ \Phi^t$).
- **Fonction de filtrage** : Retourne un symbole s_t en fonction de l'état interne S_t.

Toutes ces fonctions dépendent éventuellement de K et de l'IV.

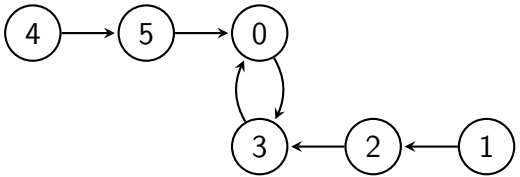
À quoi ressemble une fonction aléatoire ?

Graphe fonctionnel

Définition

Soit $F : \mathcal{E} \to \mathcal{E}$ une fonction définie sur un ensemble fini $\mathcal{E}$. Son **graphe fonctionnel** est le graphe orienté dont les sommets sont les éléments de $\mathcal{E}$ et les arrêtes sont de la forme $(x, F(x))_{x \in \mathcal{E}}$.

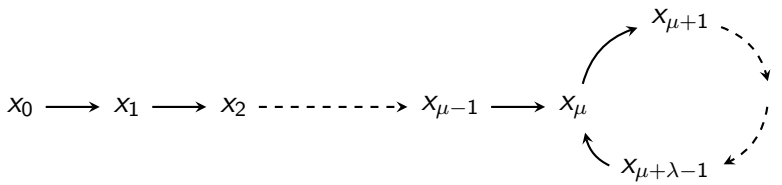
Exemple :



Graphe fonctionnel de $F(x) \stackrel{\text{def}}{=} (x-1)^2 + 2$ définie sur $\mathbb{Z}/6\mathbb{Z}$.

Fonction aléatoire

Prop : Soit $F : \mathcal{E} \to \mathcal{E}$ une fonction définie sur un ensemble fini $\mathcal{E}$ et fixons $x_0 \in \mathcal{E}$. On note $x_n \stackrel{\text{def}}{=} F^n(x_0)$. Alors, le sous-graphe issu de x_0 est toujours de la forme suivante :



(Flajolet, Odlyzko, 1990)

Prop : Pour une fonction aléatoire F, et un point initial x_0 aléatoire, on a

$$\mathbb{E}(\lambda) \sim \mathbb{E}(\mu) \sim \sqrt{\frac{\pi|\mathcal{E}|}{8}}$$

Conséquence Cryptanalytique

Réduction de l'espace des états internes

Si $\mathcal{E}$ désigne l'espace des états à n bits possibles, de taille 2^n, et si on choisit pour Φ une fonction aléatoire, alors après $2^{n/2}$ étapes, on s'attend à ce que la suite produite par le PRG soit périodique de période $2^{n/2}$.

$\rightarrow$ Peut être exploité dans une attaque plus rapide que la recherche exhaustive sur $\mathcal{E}$!

Solution (Flajolet, Sedgewick, 2009)

Choisir Φ **bijective** limite cet effet :

$$\mathbb{P}(\Phi \text{ ait un cycle de taille } \geqslant 2^{n-1}) \approx \ln(2) \approx 0.69$$

Attention à bien analyser les **petits cycles** qui produisent des mauvaises suites chiffrantes...

Pour aller plus loin

La bible en combinatoire analytique :

Philippe Flajolet, Robert Sedgewick - *Analytic Combinatorics* - 2009.

Analyse de PRG

Runs

On considère une suite $(s_n)_{n\in\mathbb{N}}$ binaire, périodique de période T.

Définition

On appelle *run* une succession consécutive de bits identique maximale.

Exemple :
Dans la suite prenant les valeurs (0100011), l'ensemble des *runs* est :

$$\begin{aligned} 0 &: \quad \text{Un run de taille 1,} \\ 1 &: \quad \text{Un run de taille 1,} \\ 000 &: \quad \text{Un run de taille 3,} \\ 11 &: \quad \text{Un run de taille 2.} \end{aligned}$$

Autocorélation

On considère une suite $(s_n)_{n\in\mathbb{N}}$ binaire, périodique de période T.

Définition

La fonction d'autocorrélation de z est

$$C_s(\tau) \stackrel{\text{def}}{=} \sum_{i=0}^{T-1} (-1)^{s_i + s_{i+\tau}}.$$

Remarque :

$$C_s(\tau) = T - 2 \times |\{i \mid s_i \neq s_{i+\tau}\}|$$

Ainsi, $C_s(\tau) = T$ si et seulement si τ est un multiple de la période.

Les critères de Golomb (1982)

Une suite binaire $(s_n)_{n\in\mathbb{N}}$ périodique de période T pseudoaléatoire doit vérifier les critères suivants :

- **Équirépartition des runs** : Soit S l'ensemble des runs de (s), et soit k tel que $2^k \leqslant |S| < 2^{k+1}$. Alors (s) doit avoir $|S|/2^\ell$ runs de taille ℓ pour $1 \leqslant \ell \leqslant k$, et le nombre de runs de 0 doit différer du nombre de runs de 1 d'au plus 1.

- **Suite équilibrée** (Conséquence) : Sur une période, le nombre de 0 et de 1 diffère au plus de 1 :
$$|\sum_{i=0}^{T-1}(-1)^{s_i}| \leqslant 1$$

- **Faible autocorrélation** : C_s est constante sur $\mathbb{Z} \setminus T\mathbb{Z}$,
i.e. (s) est indépendante de tous ses translatés.

Quelques références


Solomon W. Golomb. - *Shift Register Sequences* - Aegean Park Press, 1982.


Tor Helleseth - *Maximal-Length Sequences* - In *Encyclopedia of Cryptography and Security, 2nd Ed.*, 2011.

Un exemple

Exercice

Vérifiez que la suite de période 31 :

$$1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0$$

vérifie les critères de Golomb.

Remarques :

- C'est un exemple de suite produite par un LFSR (*Linear Feedback Shift Register* ou Registre à décalage à rétroaction linéaire) de longueur 5, et dont le polynôme de rétroaction est $X^5 + X^2 + 1$.
- On verra qu'une telle suite n'est **pas imprédictible**, mais est à la base de nombreux stream ciphers.

Générateurs Pseudoaléatoires à la Sécurité Prouvée ?

Ouverture : Cryptographie Avancée - **Sécurité réductioniste**.

- Pour le moment : Sécurité assurée par la cryptanalyse *ad-hoc*.
- Il existe des PRG dont la sécurité repose sur des problèmes bien connus :
  - Le PRG de Blum, Blum et Shub (1986) est **impredictible** si le problème de *résidu quadratique* est difficile (*a priori* plus simple que la factorisation).
  - Le PRG de Blum et Micali (1984) est **imprédictible** si le problème du logarithme discret dans $(\mathbb{Z}/p\mathbb{Z})^\times$ est difficile. En particulier, p doit être grand !

Ces deux PRG sont très lents.

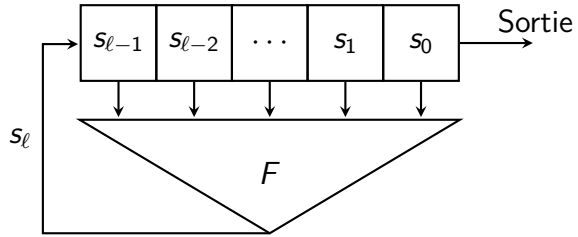
Linear Feedback Shift Registers

Feedback Shift Registers

Définition

$(s_n)_{n\in\mathbb{N}} \in \mathbb{F}_q^{\mathbb{N}}$ est dite produite par une fonction de réatroaction (*Feedback function*) s'il existe $\ell \in \mathbb{N}$ et $F : \mathbb{F}_q^{\ell} \to \mathbb{F}_q$ tels que

$$s_{n+\ell} = F(s_{n+\ell-1}, s_{n+\ell-2}, \ldots, s_{n+1}, s_n), \quad \forall n \in \mathbb{N}$$



Linear Feedback Shift Registers (LFSR)

Définition

$(s_n)_{n\in\mathbb{N}} \in \mathbb{F}_q^{\mathbb{N}}$ est dite produite par une fonction de réatroaction **Linéaire** (*Linear Feedback function*) s'il existe $n \in \mathbb{N}$ et $F : \mathbb{F}_q^n \to \mathbb{F}_q$ une **forme linéaire** tels que

$$s_{n+\ell} = F(s_{n+\ell-1}, s_{n+\ell-2}, \ldots, s_{n+1}, s_n), \quad \forall n \in \mathbb{N}$$

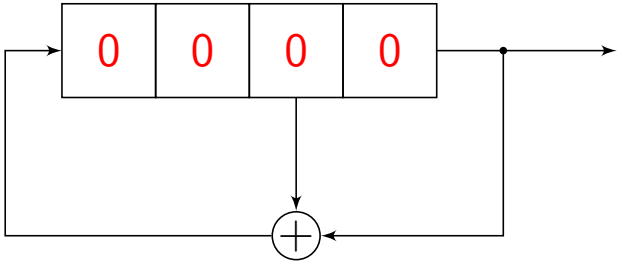
En d'autres termes, (s_n) est produite par un LFSR s'il existe un vecteur constant $(c_1, \ldots, c_\ell) \in \mathbb{F}_q^\ell$ tel que

$$s_{n+\ell} = c_1 s_{n+\ell-1} + \cdots + c_\ell s_n$$

Proposition

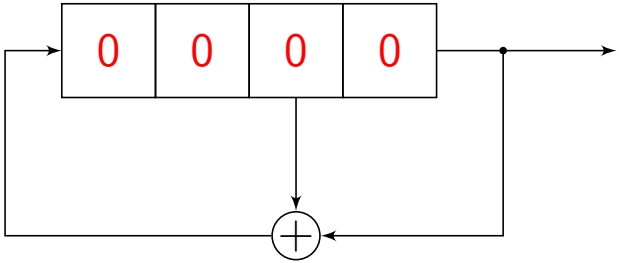
Toute suite récurrente linéaire d'ordre ℓ est produite par un LFSR de longueur ℓ.

Exemple : LFSR binaire



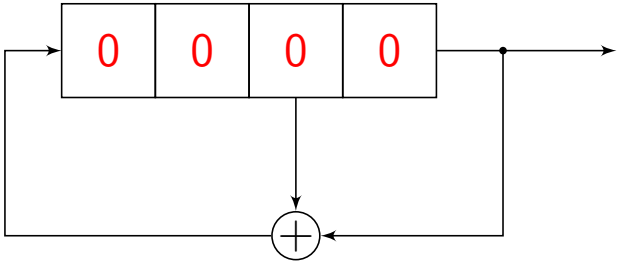
0

Exemple : LFSR binaire



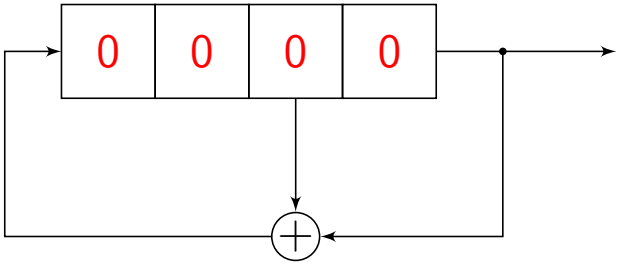
00

Exemple : LFSR binaire



000

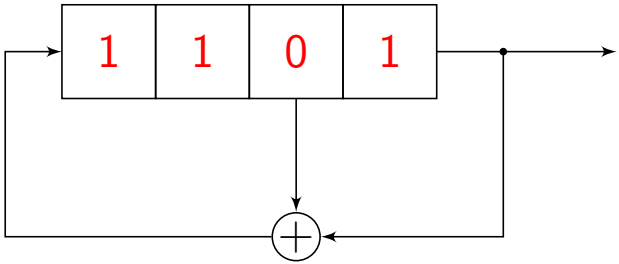
Exemple : LFSR binaire



$0000\cdots$

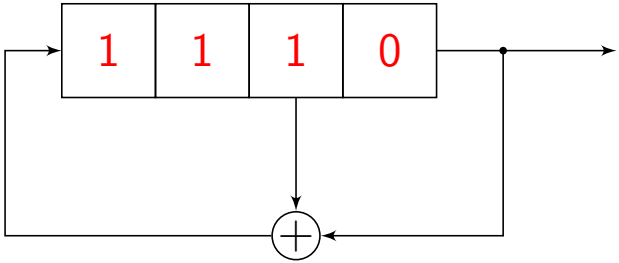
Il faut éliminer l'état 0.

Exemple : LFSR binaire non trivial



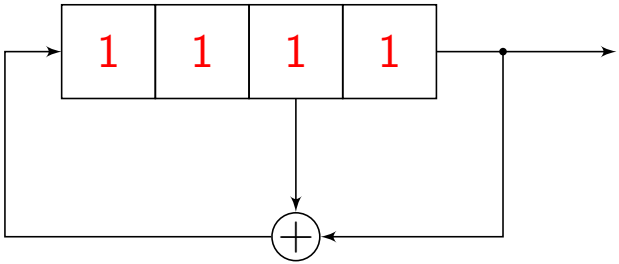
1

Exemple : LFSR binaire non trivial



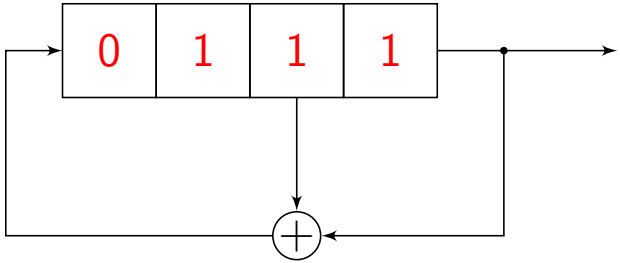
10

Exemple : LFSR binaire non trivial



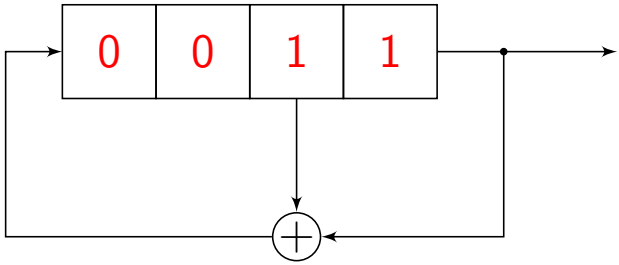
101

Exemple : LFSR binaire non trivial



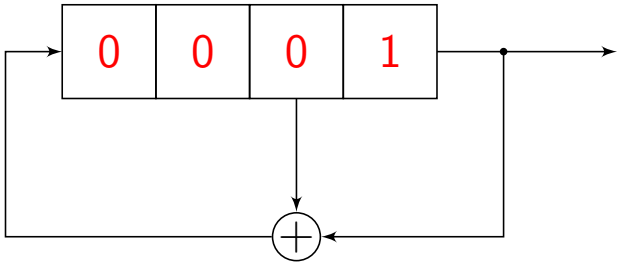
1011

Exemple : LFSR binaire non trivial



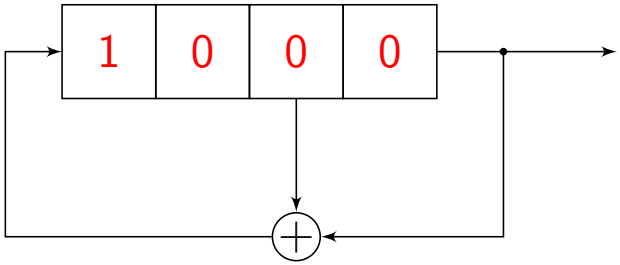
10111

Exemple : LFSR binaire non trivial



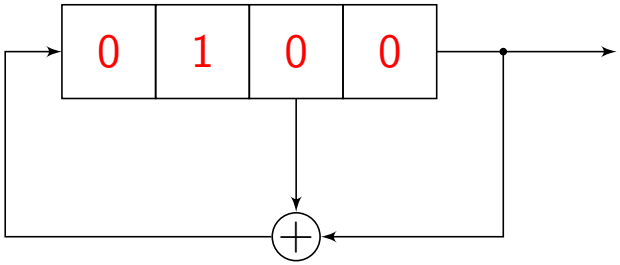
101111

Exemple : LFSR binaire non trivial



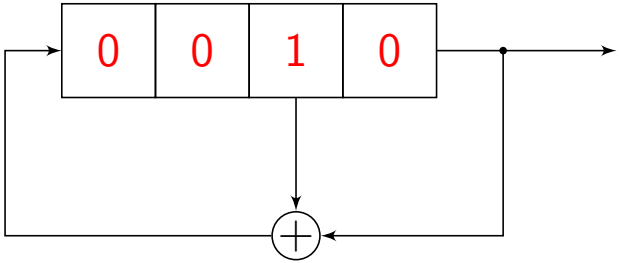
1011110

Exemple : LFSR binaire non trivial



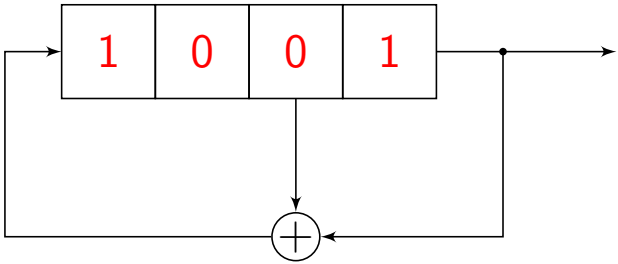
10111100

Exemple : LFSR binaire non trivial



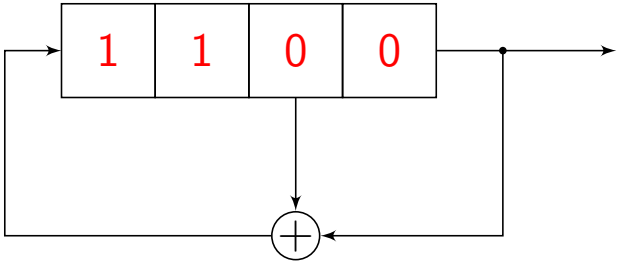
101111000

Exemple : LFSR binaire non trivial



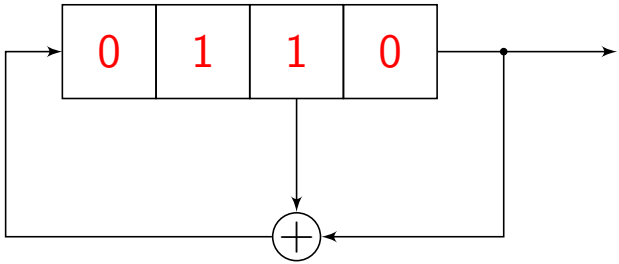
1011110001

Exemple : LFSR binaire non trivial



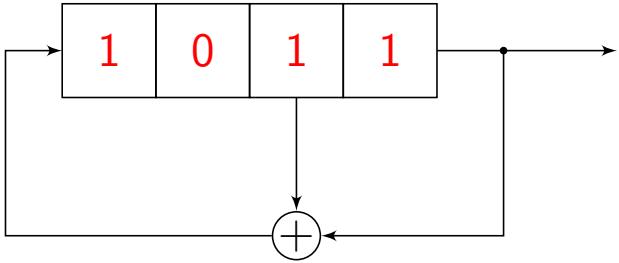
10111100010

Exemple : LFSR binaire non trivial



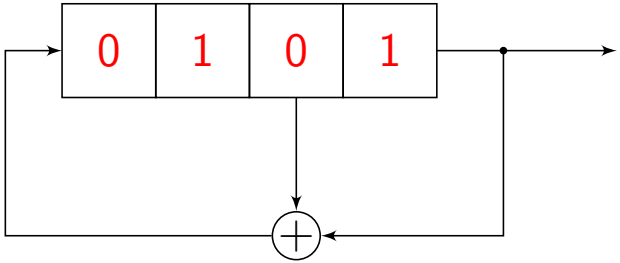
101111000100

Exemple : LFSR binaire non trivial



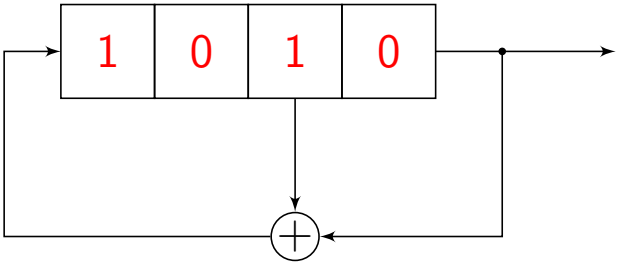
1011110001001

Exemple : LFSR binaire non trivial



10111100010011

Exemple : LFSR binaire non trivial



$$101111000100110\cdots$$

- La période est $15 = 2^4 - 1$
- Elle est **maximale** !

Polynôme de Rétroaction

Définition

Soit un LFSR de longueur ℓ et de coefficients $(c_1, \ldots, c_\ell) \in \mathbb{F}_q^\ell$. On lui associe son **polynôme de rétroaction** défini par

$$P(X) \stackrel{\text{def}}{=} 1 - \sum_{i=1}^{\ell} c_i X^i \in \mathbb{F}_q[X]$$

Exemple

Le LFSR précédent a pour polynôme de rétroaction

$$P(X) = 1 + X^3 + X^4 \in \mathbb{F}_2[X]$$

Interlude Mathématique

Rappel : Séries Formelles

L'espace des suites $\mathbb{F}_q^{\mathbb{N}}$ est naturellement muni d'une structure de $\mathbb{F}_q$-espace vectoriel (de dimension infinie). On peut aussi lui munir d'une **structure d'algèbre** en définissant le produit

$$(a_n) \odot (b_n) \stackrel{\text{def}}{=} (c_n)$$

où

$$c_n \stackrel{\text{def}}{=} \sum_{k=0}^{n} a_k b_{n-k}.$$

Cette structure est appelée algèbre des **séries formelles** à une indéterminée sur $\mathbb{F}_q$ et est notée $\mathbb{F}_q[[X]]$, et une suite (a_n) vue comme élément de $\mathbb{F}_q[[X]]$ est notée

$$\sum_{n \geqslant 0} a_n X^n.$$

Quelques Propriétés et Exemples

- $\mathbb{F}_q[[X]]$ est une algèbre commutative, intègre, de dimension infinie sur $\mathbb{F}_q$.
- $\mathbb{F}_q[[X]]$ contient l'algèbre des polynômes $\mathbb{F}_q[X]$.
- $A = \sum_{n \geqslant 0} a_n X^n \in \mathbb{F}_q[[X]]$ est inversible si et seulement si $a_0 \neq 0$.
- $\mathbb{F}_q[[X]]$ n'est pas un corps.

- $1 + X \in \mathbb{F}_q[X] \subset \mathbb{F}_q[[X]]$ est inversible, d'inverse

$$\frac{1}{1+X} \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} (-1)^n X^n$$

- X n'est pas inversible dans $\mathbb{F}_q[[X]]$.

Pour l'intuition : Analogie $\mathbb{Z}$ et $\mathbb{F}_q[X]$

Au tableau, si on a le temps.

Retour à nos LFSR

Longueur d'une suite LFSR

Pour $(s_n) \in \mathbb{F}_q^{\mathbb{N}}$, on note $S(X) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} s_n X^n$ sa serie formelle associée.

Théorème

- $(s_n)_{n\in\mathbb{N}}$ est produite par un LFSR de longueur ℓ si et seulement si $S(X) \in \mathbb{F}_q(X)$.
- Dans ce cas, si $P(X)$ désigne son polynôme de rétroaction, alors il existe $Q(X) \in \mathbb{F}_q[X]$ de degré $< \ell$ tel que

$$S(X) = \frac{Q(X)}{P(X)}.$$

Preuve au tableau.

Conséquence 1 : Sparsification

Proposition

Toute suite produite par un LFSR de rétroaction P peut aussi être produite par tout LFSR de rétroaction un multiple quelconque de P.

Exemple : Soit (s_n) une suite binaire vérifiant

$$s_{n+6} = s_{n+4} + s_{n+3} + s_{n+1} + s_n, \quad \forall n \geqslant 6.$$

Son polynôme de rétroaction est $P(X) = 1 + X^2 + X^3 + X^5 + X^6$. On vérifie que cette suite vérifie aussi $s_{n+8} = s_{n+7} + s_n$, puisque

$$1 + X + X^8 = (1 + X + X^2)P(X).$$

Utile pour des implémentations optimales, et peut être exploité pour des distingueurs.

Conséquence 2 : Minimisation

Soit $\mathbf{s} \stackrel{\text{def}}{=} (s_n)_{n\in\mathbb{N}} \in \mathbb{F}_q^{\mathbb{N}}$ une suite récurrente linéaire.

Théorème

Parmi tous les polynômes de rétroaction pour $\mathbf{s}$, il en existe 1 de **degré minimal**.

Exercice

Comment calculer (algorithmiquement) le polynôme minimal associé à une suite LFSR ?

Profile de Complexité Linéaire d'une Suite

Définition

La **complexité linéaire** $\Lambda(s)$ d'une suite $(s_n)_{n\in\mathbb{N}}$ est :

- 0 si s est la suite nulle.
- le degré de son polynôme de rétroaction minimal s'il existe.
- ∞ sinon.

Le **profile de complexité linéaire** d'une suite $\mathbf{s} \stackrel{\text{def}}{=} (s_n)_{n\in\mathbb{N}}$ est l'ensemble des $\Lambda(\mathbf{s}^{(n)})$ où

$$\mathbf{s}^{(n)} \stackrel{\text{def}}{=} (s_0, \ldots, s_{n-1}, s_0, \ldots).$$

Rueppel 1986 - *Analysis and Design of Stream Ciphers*

Soit $\mathbf{s}^{(n)} = (\mathbf{s}_0, \ldots, \mathbf{s}_{n-1})$ formée de n variables indépendantes et uniformes sur $\mathbb{F}_2$. Alors

$$\mathbb{E}(\Lambda(\mathbf{s}^{(n)})) = \frac{n}{2} + \frac{4 + \varepsilon(n)}{18} + 2^{-n}\left(\frac{n}{3} + \frac{2}{9}\right)$$

où $\varepsilon(n) \stackrel{\text{def}}{=} n \mod 2$ est la parité de n.

Comparer le profile de complexité linéaire d'une suite avec ce résultat peut parfois fournir un distingueur.

Période d'un LFSR

Soit $\mathbf{s} \stackrel{\text{def}}{=} (s_n)_{n\in\mathbb{N}} \in \mathbb{F}_q^{\mathbb{N}}$ une suite récurrente linéaire, et soit P son polynôme de rétroaction **minimal**. On note ℓ son degré.

Prop : La période de $\mathbf{s}$ est $q^\ell - 1$ si et seulement si P est **primitif**.

Exemple

Le polynôme de rétroaction du LFSR du début était $P(X) = 1 + X^3 + X^4$ qui est bien primitif. On retrouve que sa période est bien de $2^4 - 1 = 15$.

Algorithme de Berlekamp-Massey

En TD

- Un LFSR primitif produit des suites aux bonnes propriétés statistiques.
- Un LFSR seul ne peut pas être utilisé pour construire un bon chiffrement par flot sécurisé.
- En pratique, on va **combiner** des LFSR, ou bien rajouter une **fonction de filtrage** en sorte.