Lecture 8: Stream ciphers - LFSR sequences

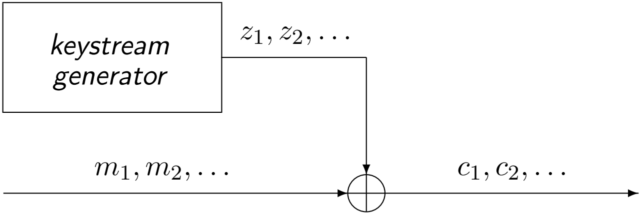
Thomas Johansson

Introduction

- Symmetric encryption algorithms are divided into two main categories, *block ciphers and stream ciphers*.
- Block ciphers tend to encrypt a block of characters of a plaintext message using a fixed encryption transformation
- A stream cipher encrypt individual characters of the plaintext using an encryption transformation that varies with time.

A stream cipher built around LFSRs and producing one bit output on each clock = *classic stream cipher design*.

A stream cipher



- $\mathbf{z} = z_1, z_2, \ldots$ *keystream*
- key K

A stream cipher

- Design goal is to efficiently produce random-looking sequences that are as "indistinguishable" as possible from truly random sequences.
- Recall the unbreakable Vernam cipher.
- For a synchronous stream cipher, a known-plaintext attack (or chosen-plaintext or chosen-ciphertext) is equivalent to having access to the keystream $\mathbf{z} = z_1, z_2, \ldots, z_N$.
- *We assume that an output sequence* $\mathbf{z}$ *of length* N *from the keystream generator is known to Eve.*

Type of attacks

- *Key recovery attack:* Eve tries to recover the secret key K.
- *Distinguishing attack:* Eve tries to determine whether a given sequence $\mathbf{z} = z_1, z_2, \ldots, z_N$ is likely to have been generated from the considered stream cipher or whether it is just a truly random sequence.

Distinguishing attack is a much weaker attack

Distinguishing attack

- Let $D(\mathbf{z})$ be an algorithm that takes as input a length N sequence $\mathbf{z}$ and as output gives either "X" or "RANDOM".
- With probability $1/2$ the sequence $\mathbf{z}$ is produced by generator X and with probability $1/2$ it is a purely random sequence.
- The probability that $D(\mathbf{z})$ correctly determines the origin of $\mathbf{z}$ is written $1/2 + \epsilon$.
- If ϵ is not very close to zero we say that $D(\mathbf{z})$ is a *distinguisher* for generator X.

Distinguishing attack - example

Assume that Alice sends one of N public images $\{I_1, I_2, \ldots, I_N\}$ to Bob.

Eve observes the ciphertext $\mathbf{c}$.

- Guess that the plaintext is the image I_1, i.e., $\mathbf{m} = I_1$.
- Calculate $\hat{\mathbf{z}} = \mathbf{m} + \mathbf{c}$ and compute $D(\hat{\mathbf{z}})$.
- If the guess $\mathbf{m} = I_1$ was correct then $D(\hat{\mathbf{z}}) = X$. If not, $D(\hat{\mathbf{z}}) =$"RANDOM".

More on attacks

- Building a (synchronous) stream cipher reduces to the problem of building a generator that is resistant to all distinguishing attacks.
- There are essentially always both distinguishing attacks and key recovery attacks on a cipher.
- *Exhaustive keysearch*; complexity 2^k
- An attack is considered successful only if the complexity of performing it is considerably lower than 2^k key tests.

Building blocks for stream ciphers

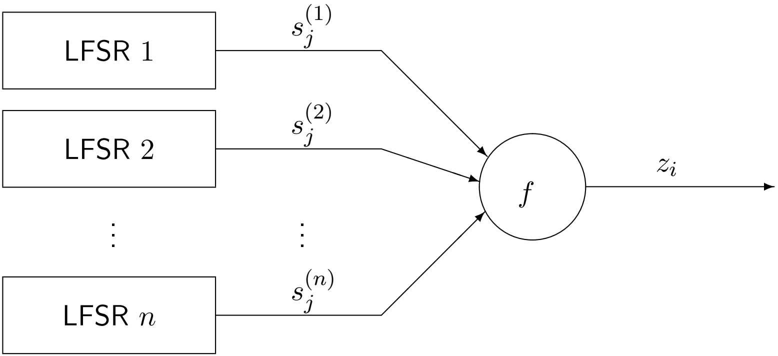
MEMORY

- linear feedback shift registers, or LFSRs for short.
- tables (arrays)

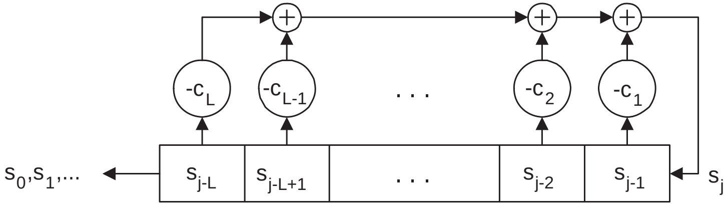
Combinatorial function

- Nonlinear Boolean functions, S-boxes
- XOR, Modular addition, (cyclic) rotations, (multiplications)

Example of a stream cipher design



Linear feedback shift registers



A register of L delay (storage) elements each capable of storing one element from $\mathbb{F}_q$, and a clock signal.

Clocking, the register of delay elements is shifted one step and the new value of the last delay element is calculated as a linear function of the content of the register.

LFSR sequences

- The linear function is described through the coefficients $c_1, c_2, \ldots, c_L \in \mathbb{F}_q$ and the recurrence relation is

$$s_j = -c_1 s_{j-1} - c_2 s_{j-2} - \cdots c_L s_{j-L},$$

  for $j = L, L+1, \ldots$.
- With $c_0 = 1$ we can write

$$\sum_{i=0}^{L} c_i s_{j-i} = 0, \text{ for } j = L, L+1, \ldots.$$

  The *shift register equation*.
- The first L symbols $s_0, s_1, \ldots, s_{L-1}$ form the *initial state*.

LFSR sequences

- The coefficients $c_0, c_1, \ldots, c_L$ are summarized in the *connection polynomial* $C(D)$ defined by

$$C(D) = 1 + c_1 D + c_2 D^2 + \cdots + c_L D^L.$$

- Write $< C(D), L >$ to denote the LFSR with connection polynomial $C(D)$ and length L.
- *D-transform* of a sequence $\mathbf{s} = s_0, s_1, s_2 \ldots$ as

$$S(D) = s_0 + s_1 D + s_2 D^2 + \cdots ,$$

  assuming $s_i \in \mathbb{F}_q$.
- The indeterminate D is the "delay" and its exponent indicate time.

LFSR sequences

- We assume $s_i = 0$ for $i < 0$. The set of all such sequences having the form

$$f(D) = \sum_{i=0}^{\infty} f_i D^i,$$

$f_i \in \mathbb{F}_q$, is denoted $\mathbb{F}_q[[D]]$ and called the *ring of formal power series*.

Theorem

The set of sequences generated by the LFSR with connection polynomial $C(D)$ is the set of sequences that have D-transform

$$S(D) = \frac{P(D)}{C(D)},$$

where $P(D)$ is an arbitrary polynomial of degree at most $L-1$,

$$P(D) = p_0 + p_1 D + \ldots + p_{L-1} D^{L-1}.$$

Furthermore, the relation between the initial state of the LFSR and the $P(D)$ polynomial is given by the linear relation

$$\begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_{L-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ c_1 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ c_{L-1} & c_{L-2} & \ldots & 1 \end{pmatrix} \begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{L-1} \end{pmatrix}.$$

LFSR sequences and extension fields

- Let $\pi(x)$ be an irreducible polynomial over $\mathbb{F}_q$ and assume that its coefficients are

$$\pi(x) = x^L + c_1 x^{L-1} + \cdots + c_L.$$

This means that $\pi(x)$ is the *reciprocal* polynomial of $C(D)$.

- Construct the extension field $\mathbb{F}_{q^L}$ through $\pi(\alpha) = 0$.
- β from $\mathbb{F}_{q^L}$ can be expressed in a polynomial basis as

$$\beta = \beta_0 + \beta_1 \alpha + \cdots + \beta_{L-1} \alpha^{L-1},$$

where $\beta_0, \beta_1, \ldots \beta_{L-1} \in \mathbb{F}_q$.

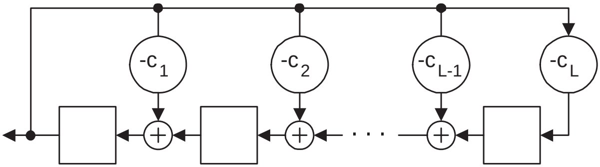
LFSR sequences and extension fields

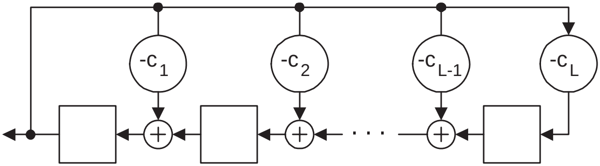
Assume that the (unknown) element β is multiplied by the fixed element α. The result is

$$\alpha\beta = \beta_0\alpha + \beta_1\alpha^2 + \cdots + \beta_{L-1}\alpha^L.$$

Reducing α^L using $\pi(\alpha) = 0$ gives

$$\alpha\beta = -c_L\beta_{L-1} + (\beta_0 - c_{L-1}\beta_{L-1})\alpha + \cdots + (\beta_{L-2} - c_1\beta_{L-1})\alpha^{L-1}.$$





- It is quickly checked that

$$s_j = -c_1 s_{j-1} - c_2 s_{j-2} - \cdots c_L s_{j-L},$$

  when $j \geq L$.
- $p_0 = s_0$, $p_1 = s_1 + c_1 s_0$, etc, where $p_0, p_1, \ldots, p_{L-1}$ is the initial state
- The sequence fulfills the shift register equation, but uses $p_0, p_1, \ldots p_{L-1}$ as initial state.

LFSR sequences and extension fields

- **The set of LFSR sequences, when $C(D)$ is irreducible, is exactly the set of sequences possible to produce by the implementation of multiplication of an element β by the fixed element α in $\mathbb{F}_{q^L}$.**
- For a specific sequence specified as $S(D) = P(D)/C(D)$ the initial state is the first L symbols whereas the same sequence is produced in the figure if the initial state is $p_0, p_1, \ldots, p_{L-1}$.

Properties of LFSR sequences

- A sequence $\mathbf{s} = \ldots, s_0, s_1, \ldots$ is called *periodic* if there is a positive integer T such that $s_i = s_{i+T}$, for all $i \geq 0$.
- The *period* is the least such positive integer T for which $s_i = s_{i+T}$, for all $i \geq 0$.
- The LFSR state runs through different values. The initial state will appear again after visiting a number of states. If $\deg C(D) = L$, the period of a sequence is the same as the number of different states visited, before returning to the initial state.

Properties of LFSR sequences

- $C(D)$ irreducible: the state corresponds to an element in $\mathbb{F}_{q^L}$, say β.
- The sequence of different states that we are entering is then

$$\beta, \alpha\beta, \alpha^2\beta, \ldots, \alpha^{T-1}\beta, \alpha^T\beta = \beta,$$

  where T is the order or α.
- If α is a primitive element (its order is $q^L - 1$), then obviously we will go trough all $q^L - 1$ different states and the sequence will have period $q^L - 1$. Such sequences are called m-*sequences* and they appear if and only if the polynomial $\pi(x)$ is a primitive polynomial.

Example

- Length 4 LFSR with connection polynomial $C(D) = 1 + D + D^2 + D^3 + D^4$ in $\mathbb{F}_2$.
- Starting in (0001), we return after 5 clockings of the LFSR.
- There are three cycles of length 5 and one of length one.
- Explanation: $\mathbb{F}_{2^4}$, we get through $\pi(x) = x^L C(x^{-1}) = x^4 + x^3 + x^2 + x + 1$ and $\pi(\alpha) = 0$.
- $\alpha^5 = 1$ and $\mathrm{ord}(\alpha) = 5$. So starting in any nonzero state $\beta \in \mathbb{F}_{2^4}$, we will jump between the states

$$\beta, \alpha\beta, \alpha^2\beta, \alpha^3\beta, \alpha^4\beta, \alpha^5\beta = \beta.$$

Example

- Length 4 LFSR with connection polynomial $C(D) = 1 + D + D^4$ in $\mathbb{F}_2$.
- Starting in (0001), we return after 15 clockings of the LFSR.
- Explanation: $\mathbb{F}_{2^4}$, we get through $\pi(x) = x^L C(x^{-1}) = x^4 + x^3 + 1$ and $\pi(\alpha) = 0$.
- $\alpha^{15} = 1$ and $\mathrm{ord}(\alpha) = 15$. $\pi(x)$ primitive polynomial.
- So starting in any nonzero state $\beta \in \mathbb{F}_{2^4}$, we will jump between all nnzero states before returning.

Properties of LFSR sequences

The different state cycles that will appear for an arbitrary LFSR.

- $[s_0, s_1, \ldots, s_{T-1}]^\infty$ denote the periodic and causal sequence

  $$s_0, s_1, \ldots, s_{T-1}, s_0, s_1, \ldots, s_{T-1}, s_0, \ldots,$$

  where $s_i \in \mathbb{F}_q$, $i = 0, 1, \ldots, T-1$.
- $(s_0, s_1, \ldots, s_{N-1})$ denote a sequence where the first N symbols are $s_0, s_1, \ldots, s_{N-1}$ (and the upcoming symbols are not defined), where $s_i \in \mathbb{F}_q$, $i = 0, 1, \ldots, N-1$.

Properties of LFSR sequences

- If $\mathbf{s} = [1, 0, 0, \ldots, 0]^\infty$ then

$$S(D) = 1 + D^T + D^{2T} + \cdots = \frac{1}{1 - D^T}.$$

- il $\mathbf{s} = [0, 1, 0, \ldots, 0]^\infty$ then

$$S(D) = D + D^{T+1} + D^{2T+1} + \cdots = \frac{D}{1 - D^T}$$

- In general, if $\mathbf{s} = [s_0, s_1, \ldots, s_{T-1}]^\infty$ then

$$S(D) = \frac{s_0}{1 - D^T} + \frac{s_1 D}{1 - D^T} + \ldots = \frac{s_0 + s_1 D + s_2 D^2 + \ldots s_{T-1} D^{T-1}}{1 - D^T}. \quad (1)$$

Properties of LFSR sequences

Definition

The *period of a polynomial* $C(D)$ is the least positive number T such that $C(D)|(1 - D^T)$.

- Calculated by division of 1 by $C(D)$ and continuing until the we receive the first remainder of the form $1 \cdot D^N$. Then the period is $T = N$.

(example)

Properties of LFSR sequences

Theorem

If $\gcd(C(D), P(D)) = 1$ *then the connection polynomial* $C(D)$ *and the sequence* $\mathbf{s}$ *with D-transform*

$$S(D) = \frac{P(D)}{C(D)}$$

have the same period (the period of $\mathbf{s}$ *is the same as the period of the polynomial* $C(D)$*).*

- Note: This $C(D)$ gives the shortest LFSR generating $\mathbf{s}$. Any other connection polynomial generating $\mathbf{s}$ must be a multiple of $C(D)$.

(example)

Properties of LFSR sequences

Theorem

If two sequences, $\mathbf{s}_A$ *and* $\mathbf{s}_B$*, with periods* T_A *and* T_B *have D-transforms*

$$S_A(D) = \frac{P_A(D)}{C_A(D)}, S_B(D) = \frac{P_B(D)}{C_B(D)},$$

then the sum of the sequences $\mathbf{s} = \mathbf{s}_A + \mathbf{s}_B$ *has D-transform* $S(D) = S_A(D) + S_B(D)$ *and period* $\mathrm{lcm}(T_A, T_B)$*, assuming* $\gcd(P_A(D), C_A(D)) = 1$, $\gcd(P_B(D), C_B(D)) = 1$, $\gcd(C_A(D), C_B(D)) = 1$.

(example)

LFSR cycle sets

- Introduce the *cycle set* for $C(D)$ (assuming $L = \deg C(D)$).
- Written in the form $n_1(T_1) \oplus n_2(T_2) \oplus \ldots$.
- $1(1) \oplus 3(5)$, one cycle of length one and three cycles of length 5.
- $n_1(T) \oplus n_2(T) = (n_1 + n_2)(T)$.

LFSR cycle sets

Already established facts:

- If $C(D)$ is a primitive polynomial of degree L over $\mathbb{F}_q$ then the cycle set is
$$1(1) \oplus (1)(q^L - 1).$$
- If $C(D)$ is an irreducible polynomial then the cycle set is
$$1(1) \oplus \frac{(q^L - 1)}{T}(T),$$
where T is the period of the polynomial $C(D)$ (or the order of α when $\pi(\alpha) = 0$).

LFSR cycle sets - remaining cases

Theorem

If $C(D) = C_1(D)^n$ *then the cycle set of* $C(D)$ *is*

$$1(1) \oplus \frac{(q^{L_1} - 1)}{T_1}(T_1) \oplus \frac{q^{L_1}(q^{L_1} - 1)}{T_2}(T_2) \oplus \cdots \frac{q^{(n-1)L_1}(q^{L_1} - 1)}{T_n}(T_n),$$

where $\deg C(D) = L$ *and* T_j *is the period of the polynomial* $C_1(D)^j$.

Theorem

If $C_1(D)$ *is irreducible with period* T_1, *then the period of the polynomial* $C_1(D)^j$ *is* $T_j = p^m T_1$ *where* p *is the characteristic of the field and* m *the integer satisfying* $p^{m-1} < j \leq p^m$.

(example)

LFSR cycle sets - remaining cases

Theorem

For a connection polynomial $C(D)$ factoring like

$$C(D) = C_1(D)^{n_1} C_2(D)^{n_2} \cdots C_m(D)^{n_m},$$

$C_i(D)$ irreducible, has cycle set $S_1 \times S_2 \times \cdots S_m$, where S_i is the cycle set for $C_i^{n_i}$, and

$$(n_1)T_1 \times (n_2)(T_2) = (n_1 n_2 \cdot \gcd(T_1, T_2)(\mathrm{lcm}(T_1, T_2))$$

and the distributive law holds for $\times$ and $\oplus$.

(example)

Decimation

An m-sequence $s = s_0, s_1, s_2, \ldots$

- Define the sequence s' obtained through *decimation* by k, defined as the sequence
$$\mathrm{s}' = s_0, s_k, s_{2k}, s_{3k}, \ldots .$$
- s correspond to multiplication of β by the fixed element α. It is clear that s' corresponds to multiplication of β by the fixed element α^k, i.e, the cycle of different states correspond to the sequence
$$\beta, \alpha^k\beta, \alpha^{2k}\beta, \ldots, \alpha^{(T-1)k}\beta, \alpha^{Tk}\beta = \beta.$$
- the period of s' is $\mathrm{ord}(\alpha^k)$ and $\mathrm{ord}(\alpha^k) = q^L - 1/\gcd(q^L - 1, k)$.

Decimation - advanced

$\mathbb{F}_{q^L}$ through a degree L polynomial $\pi(x) \in \mathbb{F}_q[x]$ with $\pi(\alpha) = 0$.

- Let $\beta \in \mathbb{F}_q$ and consider the set of polynomials

$$\mathcal{F}(\beta) = \{f(x) \in \mathbb{F}_q[x] : f(\beta) = 0\}.$$

- The set will contain at least one polynomial of degree $\leq L$.
- Let $f_0(x)$ be the polynomial in $\mathcal{F}(\beta)$ of lowest degree. Any other polynomial $f(x)$ in $\mathcal{F}(\beta)$ can be written as $f(x) = q(x)f_0(x) + r(x)$, $\deg r(x) < \deg f_0(x)$ and

$$0 = f(\beta) = q(\beta)f_0(\beta) + r(\beta) = r(\beta).$$

- So $r(\beta) = 0$ and this means that $f_0(x)|f(x)$ for all polynomials $f(x)$ in $\mathcal{F}(\beta)$.

Decimation - minimal polynomial

- The polynomial $f_0(x)$ is called the *minimal polynomial* of the element β.
- The minimal polynomial to β, now denoted $\pi_\beta(x)$, can be calculated as

$$\pi_\beta(x) = (x - \beta)(x - \beta^q)(x - \beta^{q^2}) \cdots (x - \beta^{q^{d-1}}),$$

  where d is the smallest integer such that $q^d \equiv 1 \bmod \mathrm{ord}(\beta)$ (d is the number of conjugates of β).

- The reciprocal of the minimal polynomial $\pi_\beta(x)$ gives the connection polynomial for a minimal LFSR producing a sequence corresponding to the state sequence

$$\beta, \alpha^k\beta, \alpha^{2k}\beta, \ldots, \alpha^{(T-1)k}\beta, \alpha^{Tk}\beta = \beta.$$

- The decimated sequence s' can be generated by an LFSR with a connection polynomial being the reciprocal of $\pi_{\alpha^k}(x)$.

(example)

Statistical properties of LFSR sequences

The importance of LFSR sequences in general and m-sequences in particular is due to their pseudo randomness properties.

- $\mathbf{s} = s_0, s_1, \ldots$ is an m-sequence, recall that an r-gram is a subsequence of length r,

$$(s_t, s_t + 1, \ldots, s_{t+r-1}),$$

for $t = 0, 1, \ldots$.

Theorem

Among the $q^L - 1$ L-grams that can be constructed for $t = 0, 1, \ldots, q^L - 2$, every nonzero vector appears exactly once.

Statistical properties of LFSR sequences

Run-distribution properties of m-sequences.

- A *run of length* r in a sequence $\mathbf{s}$ is a subsequence of *exactly* r zeros (or ones). This means that the r zeros must have a one before.

Statistical properties of LFSR sequences

Theorem

The run distribution of any m-sequence of length $2^L - 1$ is given as

length	0-*runs*	1-*runs*
1	2^{L-3}	2^{L-3}
2	2^{L-4}	2^{L-4}
$\vdots$	$\vdots$	$\vdots$
$L-2$	1	1
$L-1$	1	0
L	0	1
Total	2^{L-2}	2^{L-2}

Statistical properties of LFSR sequences

The *autocorrelation function*.

- Let $\mathbf{x}, \mathbf{y}$ be two binary sequences of the same length n.
- The correlation $C(\mathbf{x}, \mathbf{y})$ between the two sequences is defined as the number of positions of agreements minus the number of disagreements.
- The autocorrelation function $C(\tau)$ is defined to be the correlation between a sequence $\mathbf{x}$ and its τth cyclic shift, i.e.,

$$C(\tau) = \sum_{i=1}^{n} (-1)^{x_i + x_{i+\tau}}, \tag{2}$$

  where subscripts are taken modulo n and addition in the exponent is mod 2 addition.

Statistical properties of LFSR sequences

Theorem

If $\mathbf{s}$ *is an* m*-sequence of length* $2^L - 1$*, then*

$$C(\tau) = \begin{cases} 2^L - 1 & \textit{if } \tau \equiv 0 \pmod{n} \\ -1 & \textit{otherwise} \end{cases}$$

Statistical properties of LFSR sequences

More comments:

- The decimation of an m-sequence or the sum of two different m-sequences are (under some assumptions) again m-sequences.
- One property is completely away from random sequences. Let the binary m-sequence be generated by the recursion $s_j = \sum_{i=1}^{L} c_i s_{j-i}$. By forming a set of random variables $X_j = \sum_{i=0}^{L} c_i s_{j-i}, j \leq L$ we see that $P(X_j = 0) = 1$. An extreme point of nonrandomness.