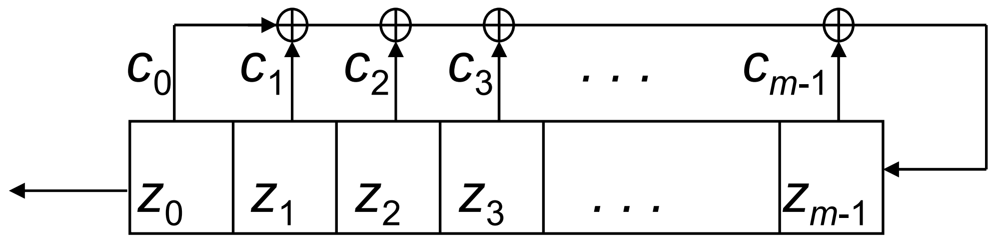
T-79.5501
Cryptology

Lecture 4 (Feb 19, 2008):

- Linear feedback shift registers
- Polynomials over $\mathbf{Z}_2$
- Linear complexity of sequences

Linear Feedback Shift Registers

A binary linear feedback shift register (LFSR) is the following device



where the i^{th} tap constant $c_i = 1$, if the switch connected, and $c_i = 0$ if it is open. The contents of the register $z_0, z_1, z_2, z_3, \ldots, z_{m-1}$ are binary values. Given this state of the device the output is z_0 and the new contents are $z_1, z_2, z_3, \ldots, z_{m-1}, z_m$, where z_m is computed using the recursion equation

$$z_m = c_0 z_0 + c_1 z_1 + c_2 z_2 + c_3 z_3 + \ldots + c_{m-1} z_{m-1}$$

The sum is computed *modulo* 2. As this process is iterated, the LFSR outputs a binary sequence $z_0, z_1, z_2, z_3, \ldots, z_{m-1}, z_m, \ldots$ Then the terms of this sequence satisfy the linear recursion relation

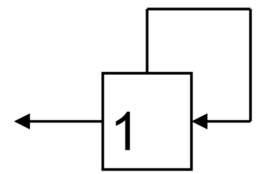
LFSR: The first examples

$$z_{k+m} = c_0 z_k + c_1 z_{k+1} + c_2 z_{k+2} + c_3 z_{k+3} + \ldots + c_{m-1} z_{k+m-1}$$

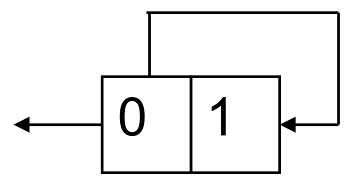
for all $k = 0,1,2,\ldots$

<u>Examples 1.</u>

a) $z_i = 0$, $i = 0,1,2,\ldots$ shortest LFSR: (no contents, length = 0)

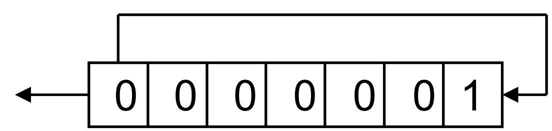
b) $z_i = 1$, $i = 0,1,2,\ldots$ shortest LFSR: 

(length $m = 1$)

c) sequence 010101… ; shortest LFSR: 

(length $m = 2$)

$z_0 = 0$, $z_1 = 1$, $z_{k+2} = z_k$, $k = 0,1,2,\ldots$

d) sequence 000000100000010… LFSR: 

LFSR: Connection polynomial

The polynomial over $\mathbf{Z_2}$

$$f(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \ldots + c_{m-1} x^{m-1} + x^m$$

is called the connection polynomial of the LFSR with taps $c_0\ c_1 \ldots c_{m-1}$.

Given $f(x) = c_0 + c_1 x + \ldots + c_{m-1} x^{m-1} + x^m$, of degree m, we denote by $f^*(x)$ the reciprocal polynomial of f, defined as follows:

$$f^*(x) = x^m f(x^{-1}) = c_0 x^m + c_1 x^{m-1} + c_2 x^{m-2} + \ldots + c_{m-1} x + 1.$$

It has the following properties:

1. $\deg f^*(x) \leq \deg f(x)$, and $\deg f^*(x) = \deg f(x)$ if and only if $c_0 = 1$.
2. Let $h(x) = f(x)g(x)$. Then $h^*(x) = f^*(x)g^*(x)$.

The set of sequences generated by the LFSR with connection polynomial $f(x)$ is denoted by $\Omega(f)$:

$$\Omega(f) = \{S = (z_i) \mid z_i \in \mathbf{Z_2};\ z_{k+m} = c_0 z_k + c_1 z_{k+1} + \ldots + c_{m-1} z_{k+m-1},\ k = 0,1,\ldots\}.$$

LFSR: Generating function

$\Omega(f)$ is a linear space over $\mathbf{Z_2}$ of dimension *m.* Its elements *S* can also be expressed using the formal power series notation:

$$S = S(x) = z_0 + z_1 x + z_2 x^2 + z_3 x^3 + . . . = \sum_{i=0...\infty} z_i x^i$$

Theorem 1. If $S(x) \in \Omega(f)$, where $\deg f(x) = m$, then there is a polynomial $P(x)$ of degree less than m such that $S(x) = P(x)/f^*(x)$.

Proof. $f^*(x) = \sum_{i=0...m} c_{m-i} x^i = \sum_{i=0...\infty} c_{m-i} x^i$, where $c_m = 1$, and $c_{m-i} = 0$, unless $0 \le i \le m$. Then

$S(x)\, f^*(x) = (\sum_{i=0...\infty} z_i x^i)(\sum_{i=0...\infty} c_{m-i} x^i) = \sum_{i=0...\infty} (\sum_{t=0...i} z_{i-t} c_{m-t})\, x^i .$

For $i \ge m$, denote $r = i - m$, and consider the i^{th} term in the sum above:

$\sum_{t=0...i} z_{i-t} c_{m-t} = \sum_{t=0...r+m} z_{r+m-t} c_{m-t} = \sum_{k=0...m} z_{r+k} c_k = 0$, as $S(x) \in \Omega(f)$.

Then $S(x)f^*(x) = \sum_{i=0...m-1} (\sum_{t=0...i} z_{i-t} c_{m-t})\, x^i = P(x)$, where $\deg P(x) < m$.

$\square$

Generating function, example

In Theorem 1, $P(x)$ =

$$z_0 + (z_1 + c_{m-1}z_0)x + (z_2 + c_{m-1}z_1 + c_{m-2}z_0)x^2 + \ldots + (z_{m-1} + c_{m-1}z_{m-2} + \ldots + c_1z_0)\, x^{m-1}$$

Hence m first terms of the sequence determine P(x) uniquely.

Example 2. 0010111 0010111 001… is generated by LFSR with polynomial f(x) = 1 + x + x^3 . Then f*(x) = $x^3 + x^2 + 1$

Generating function

$$S(x) = \underbrace{x^2 + x^4 + x^5 + x^6} + \underbrace{x^9 + x^{11} + x^{12} + x^{13}} + \underbrace{x^{16} + \ldots}$$

What is $P(x)$? $m = 3$, $z_0 = 0$, $z_1 = 0$, $z_2 = 1$, and we get

$$P(x) = z_0 + (z_1 + c_{m-1}z_0)x + (z_2 + c_{m-1}z_1 + c_{m-2}z_0)x^2 + \ldots$$
$$+ (z_{m-1} + c_{m-1}z_{m-2} + \ldots + c_1z_0)\, x^{m-1} = x^2$$

Check: $S(x) =$ P(x)/ f*(x) $= x^2/(x^3 + x^2 + 1)$

$$= x^2 + x^4 + x^5 + x^6 + x^9 + x^{11} + x^{12} + x^{13} + x^{16} + \ldots$$

LFSR: Sum sequence

Corollary 1. $\Omega(f) = \{ S(x) = P(x)/f^*(x) \mid \deg P(x) < \deg f(x) \}$.

Proof. Both sets are linear spaces over $\mathbf{Z_2}$ of the same dimension (deg $f(x)$). By Thm 1, $\Omega(f)$ is contained in the space on the right hand side. Therefore, the sets are equal.

Theorem 2. Let $h(x) = \mathrm{lcm}\,(f(x), g(x))$, and let $S_1(x) \in \Omega(f)$ and $S_2(x) \in \Omega(g)$. Then $S_1(x)+S_2(x) \in \Omega(h)$.

Proof. $h(x) = f(x)q_1(x) = g(x)q_2(x)$, where $\deg q_1(x) = \deg h(x) - \deg f(x)$ and $\deg q_2(x) = \deg h(x) - \deg g(x)$. Then by Thm 1:

$$S_1(x) + S_2(x) = (P_1(x)/f^*(x)) + (P_2(x)/g^*(x))$$

$$= (P_1(x)q_1^*(x) + P_2(x)q_2^*(x))/h^*(x)$$

where $\deg(P_1(x)q_1^*(x) + P_2(x)q_2^*(x)) \leq$

$$\max\{\deg P_1(x) + \deg q_1^*(x), \deg P_2(x) + \deg q_2^*(x)\} < \deg h(x).$$

The claim follows using Corollary 1. □

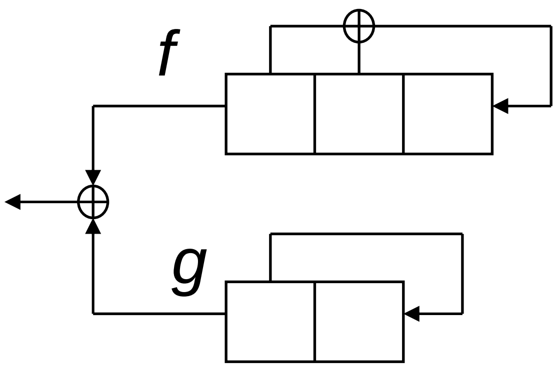
LFSR: sum sequence example

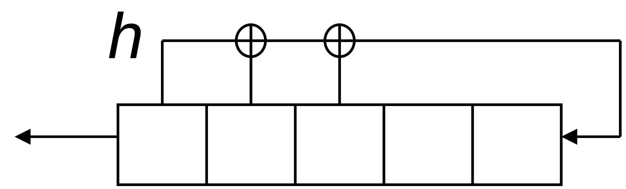
Corollary 2. If $f(x)$ divides $h(x)$, then $\Omega(f) \subset \Omega(h)$.

<u>Example 3.</u> $f(x) = 1 + x + x^3$; $g(x) = 1 + x^2$;

$$h(x) = \text{lcm}\,(f(x), g(x)) = 1 + x + x^2 + x^5 .$$

All sequences generated by the combination of the two LFSRs on the left hand side can be generated using a single LFSR of length 5:





Further, if f-LFSR is initialized with 011, g-LFSR with 00, and the h-LFSR with 01110, then the two systems generate the same sequence: 011100101110010… Indeed, take the five first bits of any sequence generated by the f register and use them to initialize the h register. Then the h register generates the same sequence as f register.

Two equivalent orderings for polynomials

Theorem 2*. Suppose that $f(x)$ and $h(x)$ are connection polynomials with constant term equal to 1. Then $\Omega(f) \subset \Omega(h)$ if and only if $f(x)$ divides $h(x)$.

Proof. By Corollary 2 it remains to show that if $\Omega(f) \subset \Omega(h)$ then $f(x)$ divides $h(x)$. Take the sequence $S(x) = 1/f^*(x) \in \Omega(f)$. Then, by the assumption $1/f^*(x) \in \Omega(h)$. By Theorem 1, there is a polynomial $P(x)$ of degree less than $\deg(h)$ such that $1/f^*(x) = P(x)/h^*(x)$. It follows that $f^*(x)P(x) = h^*(x)$. By taking reciprocals of both sides, and observing that $f^{**}(x) = f(x)$ and $h^{**}(x) = h(x)$ because the constant terms of $f(x)$ and $h(x)$ are equal to 1, we get that $h(x) = f(x)P^*(x)$, which proves the claim. $\square$

LFSR: State space

In the example above the LFSR with connection polynomial $f(x)$ runs through all seven possible non-zero states.

Whereas, the state space of the LFSR with polynomial $h(x)$ splits into five Separate sets of states as follows:

00000	11111	01010	01110	10001	00001
		10101	11100	00011	00010
			11001	00110	00100
			10010	01101	01001
			00101	11010	10011
			01011	10100	00111
			10111	01000	01111
					11110
					11101
					11011
					10110
					01100
					11000
					10000

$$1 + 1 + 2 + 7 + 7 + 14 = 32 = 2^5$$

Polynomials: Exponent

FACT 1. For all binary polynomials $f(x)$ there is a polynomial of the form $1 + x^e$, where $e \geq 1$, such that $f(x)$ divides $1 + x^e$. The smallest of such non-negative integers e is called the exponent of $f(x)$. The exponent of $f(x)$ divides all other numbers e such that $f(x)$ divides $1 + x^e$.

If $S = (z_i) \in \Omega(1 + x^n)$, then clearly $z_i = z_{i+n}$, for all $i = 0,1,\ldots$ Then it must be that the period of the sequence $S = (z_i)$ divides n.

We have the following theorem:

Theorem 3. If $S = (z_i) \in \Omega(f(x))$, then the period of S divides the exponent of $f(x)$.

FACT 2. There exist polynomials $f(x)$ for which all non-zero sequences in $\Omega(f)$ have a period equal to the exponent of $f(x)$. The polynomials with this property are exactly the irreducible polynomials.

Polynomials: Primitive polynomials

FACT 3. For all positive integers m, the largest possible value of the exponent of a polynomial of degree m is $2^m - 1$, and there exist polynomials with exponent equal to $2^m - 1$. Such polynomials are called primitive. Primitive polynomials are irreducible.

Corollary 3. Let $f(x)$ be a primitive polynomial of degree m. Then all sequences generated by an LFSR with polynomial $f(x)$ have period $2^m - 1$.

Example 4. Binary polynomials of degree 4 with non-zero constant term :

	exponent		exponent
$x^4 + 1 = (x + 1)^4$	4	$x^4 + x^2 + x + 1 = (x^3 + x^2 + 1)(x + 1)$	7
$x^4 + x + 1$ (primitive)	15	$x^4 + x^3 + x + 1 = (x + 1)^2(x^2 + x + 1)$	6
$x^4 + x^2 + 1 = (x^2 + x + 1)^2$	6	$x^4 + x^3 + x^2 + 1 = (x^3 + x + 1)(x + 1)$	7
$x^4 + x^3 + 1$ (primitive)	15	$x^4 + x^3 + x^2 + x + 1$ irreducible	5

Sanastoa

LFSR = lineaarinen siirtorekisteri

connection polynomial = kytkentäpolynomi

feedback polynomial = takaisinsyöttöpolynomi

tap (switch) constant = hana (kytkin) vakio

state = tila

power series = potenssisarja

generating function = generoiva funktio

initialize = alustaa

irreducible = jaoton

recursion = rekursio, palautuvuus

Linear complexity

Let $S = z_0, z_1, z_2, z_3, \ldots$ be a finite or infinite sequence. We say that the linear complexity LC(S) of S is the length of the shortest LFSR which generates it.

Linear complexity of a finite sequence does not decrease if new terms are added to the sequence, but it may remain the same.

<u>Examples 5.</u>

a) S = 000…01 (with n - 1 zeroes); LC(S) = n; one feedback polynomial of the LFSR is $1 + x^n$; indeed, any polynomial of degree n can be taken as feedback polynomial.

b) S = 111..10 (with n ones); LC(S) = n; one feedback polynomial of the LFSR is $1 + x + x^n$; indeed, any polynomial of degree n with odd number of terms can be taken as feedback polynomial.

c) By example 3, the linear complexity of 0111001011 is less than or equal to 3, since the polynomial f has degree 3. From b) above it follows that the linear complexity is exactly 3.

Linear complexity

Theorem 4. Let LC(S) = L . Consider the LFSR of length L which generates the sequence S of length n (where n can be infinite). Then

a) the L subsequent states of the the LFSR are linearly independent.

b) the L + 1 subsequent states are linearly dependent.

c) If moreover, at least $2L$ terms of the sequence are given, that is, $n \geq 2L$, then the connection polynomial of the generating LFSR is uniquely determined (see also Stinson: Section 1.2.5).

Proof. Let the connection coefficients be $c_0\, c_1\; c_2\; c_3\; . . . c_{L\text{-}1}$. Writing the recursion equation

$$z_{k+L} = c_0\, z_k + c_1\, z_{k+1} + c_2\, z_{k+2} + . . . + c_{L-1}\, z_{k+L-1}$$

in vector form we get

$$(z_L\; z_{L+1}\; z_{L+2}\; z_{L+3}\; . . .\; z_{2L-1}) = (c_0\; c_1\; c_2\; c_3\; . . . c_{L-1})\, Z \qquad (*)$$

Linear Complexity

where the rows (and columns) of the matrix Z are vectors $(z_k\ z_{k+1}\ z_{k+2}\ z_{k+3}\ .\ .\ .\ z_{k+L-1})$, for $k = 0,1,\ldots,L - 1$. Claim b) follows immediately from this representation. Further, if L subsequent states are linearly dependent, the sequence satisfies a linear recursion relation of length (at most) L -1, and can be generated using a LFSR of length less than L . This gives a).

Finally, if at least 2L terms of the sequence are given, then the L vectors

$$(z_k\ z_{k+1}\ z_{k+2}\ z_{k+3}\ .\ .\ .\ z_{k+L-1}),\ \ k = 0,1,\ldots,L$$

that determine the columns of the matrix Z in equation (*) are known. By a), the matrix Z is invertible. This gives a unique solution for the tap constants $(c_0\ c_1\ c_2\ c_3\ .\ .\ .c_{L-1})$. □

Linear Complexity

Now we know:

1. Any finite or periodic sequence has a finite linear complexity. Linear complexity is less than or equal to the length and the period of the sequence.
2. If we know the linear complexity of the sequence we can compute the feedback polynomial. The feedback polynomial is unique if the length of the available sequence is at least twice as much as the linear complexity.

Question:

How can we determine the linear complexity for a sequence?

Answer:

Using Berlekamp-Massey Algorithm