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 encrypts 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

\[ z = z_1, z_2, \ldots \] 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 \( z = z_1, z_2, \ldots, z_N \).

- We assume that an output sequence \( z \) 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 $z = z_1, z_2, \ldots, z_N$ 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 $D(z)$ be an algorithm that takes as input a length $N$ sequence $z$ and as output gives either “X” or “RANDOM”.

With probability $1/2$ the sequence $z$ is produced by generator $X$ and with probability $1/2$ it is a purely random sequence.

The probability that $D(z)$ correctly determines the origin of $z$ is written $1/2 + \epsilon$.

If $\epsilon$ is not very close to zero we say that $D(z)$ is a distinguisher for generator $X$. 

T. Johansson (Lund University)
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 \( c \).

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

\[ s_j^{(1)} \]
\[ s_j^{(2)} \]
\[ s_j^{(n)} \]

\[ f \]

\[ z_i \]
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

$$s_j = -c_1s_{j-1} - c_2s_{j-2} - \cdots - c_Ls_{j-L},$$

for $j = L, L + 1, \ldots$.

- With $c_0 = 1$ we can write

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

The *shift register equation*.

- The first $L$ symbols $s_0, s_1, \ldots, s_{L-1}$ form the *initial state*.
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 $s = s_0, s_1, s_2, \ldots$ 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.
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

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

$\beta$ 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 $\beta$ is multiplied by the fixed element $\alpha$. The result is

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

Reducing $\alpha^L$ 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.
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 $\beta$ by the fixed element $\alpha$ 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 $p_0, p_1, \ldots, p_{L-1}$.
A sequence \( s = \ldots, s_0, s_1, \ldots \) 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 $\beta$.
- 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 $\alpha$.
- If $\alpha$ is a primitive element (its order is $q^L - 1$), then obviously we will go through 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 \( \text{ord}(\alpha) = 5 \). 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 $\text{ord}(\alpha) = 15$. $\pi(x)$ primitive polynomial.
- So starting in any nonzero state $\beta \in \mathbb{F}_{2^4}$, we will jump between all nonzero 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 \( s = [1, 0, 0, \ldots, 0]^{\infty} \) then
  \[
  S(D) = 1 + D^T + D^{2T} + \cdots = \frac{1}{1 - D^T}.
  \]

- If \( s = [0, 1, 0, \ldots, 0]^{\infty} \) then
  \[
  S(D) = D + D^{T+1} + D^{2T+1} + \cdots = \frac{D}{1 - D^T}.
  \]

- In general, if \( s = [s_0, s_1, \ldots, s_{T-1}]^{\infty} \) then
  \[
  S(D) = \frac{s_0}{1 - D^T} + \frac{s_1 D}{1 - D^T} + \cdots = \frac{s_0 + s_1 D + s_2 D^2 + \cdots + s_{T-1} D^{T-1}}{1 - D^T}.
  \] (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 \( s \) with \( D \)-transform

\[
S(D) = \frac{P(D)}{C(D)}
\]

have the same period (the period of \( s \) is the same as the period of the polynomial \( C(D) \)).

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

(example)
Properties of LFSR sequences

**Theorem**

If two sequences, \( s_A \) and \( s_B \), with periods \( T_A \) and \( T_B \) have D-transforms

\[
S_A(D) = \frac{P_A(D)}{C_A(D)}, \quad S_B(D) = \frac{P_B(D)}{C_B(D)},
\]

then the sum of the sequences \( s = s_A + s_B \) has D-transform

\[
S(D) = S_A(D) + S_B(D) \quad \text{and period} \quad \lcm(T_A, T_B), \quad \text{assuming}
\]

\[
\gcd(P_A(D), C_A(D)) = 1, \quad \gcd(P_B(D), C_B(D)) = 1, \quad \gcd(C_A(D), C_B(D)) = 1.
\]

(example)
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.

$n_1(T) \oplus n_2(T) = (n_1 + n_2)(T)$. 
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 $\alpha$ when $\pi(\alpha) = 0$).
If $C(D) = C_1(D)^n$ then the cycle set of $C(D)$ is

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

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

If $C_1(D)$ is irreducible with period $T_1$, then the period of the polynomial $C_1(D)^j$ is $T_j = p^mT_1$ 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

\[
(n_1)T_1 \times (n_2)(T_2) = (n_1 n_2 \cdot \gcd(T_1, T_2)(\text{lcm}(T_1, T_2))
\]

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
  \[
  s' = s_0, s_k, s_{2k}, s_{3k}, \ldots .
  \]

- \( s \) correspond to multiplication of \( \beta \) by the fixed element \( \alpha \). It is clear that \( s' \) corresponds to multiplication of \( \beta \) by the fixed element \( \alpha^k \), 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 \( \text{ord}(\alpha^k) \) and \( \text{ord}(\alpha^k) = q^L - 1 / \gcd(q^L - 1, k) \).
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 \( f_0(x) \mid f(x) \) for all polynomials \( f(x) \) in \( \mathcal{F}(\beta) \).
The polynomial $f_0(x)$ is called the *minimal polynomial* of the element $\beta$.

The minimal polynomial to $\beta$, 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 $q^d \equiv 1 \mod \text{ord}(\beta)$ ($d$ is the number of conjugates of $\beta$).
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.

- $s = s_0, s_1, \ldots$ 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.
Run-distribution properties of $m$-sequences.

- A run of length $r$ in a sequence $s$ 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

<table>
<thead>
<tr>
<th>length</th>
<th>0-runs</th>
<th>1-runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$2^{L-3}$</td>
<td>$2^{L-3}$</td>
</tr>
<tr>
<td>2</td>
<td>$2^{L-4}$</td>
<td>$2^{L-4}$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$L-2$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$L-1$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$L$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>$2^{L-2}$</td>
<td>$2^{L-2}$</td>
</tr>
</tbody>
</table>
The autocorrelation function.

- Let \( x, y \) be two binary sequences of the same length \( n \).
- The correlation \( C(x, y) \) 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 \( x \) and its \( \tau \)th cyclic shift, i.e.,

\[
C(\tau) = \sum_{i=1}^{n} (-1)^{x_i + x_{i+\tau}},
\]

where subscripts are taken modulo \( n \) and addition in the exponent is mod 2 addition.
Theorem

If $s$ is an $m$-sequence of length $2^L - 1$, then

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