

Sequences, Codes, and Cryptography

Tor Helleseeth

Selmersenteret

Department of Informatics

University of Bergen

Outline

- A brief Norwegian **crypto** history
- **Sequences** and their applications
 - Correlation/CDMA/S-boxes/Stream ciphers
- **Coding** theory and applications
 - Authentication codes
 - Stream cipher attacks
 - Correlation attacks/Algebraic attacks
 - Rønjom-Helleseeth attack
- The future worldrecord - **ArcticCrypt**

A Brief Norwegian Crypto History

Norwegian Crypto in the 1930s

- Captain [Alfred Roscher Lund](#)
 - Organized Norwegian cryptology in 1930s
 - Puzzles in [Aftenposten](#), November 23, 1935
 - Recruited a “Crypto Club” when WW II broke out. Members included:
 - Puzzle solvers
 - Mathematicians
 - Bridge players

Premieoppgave.

Hvem kan løse et kryptogram?

Annetsteds i bladet vil det finnes en artikkel om hemmelig skrift som også inneholder en beskrivelse av hvordan et enkelt kryptogram kan knekkes. Som det vil sees er knekningen av et kryptogram ingen heksekunst, men det krever litt tålmodighet og at man prøver sig frem og tar fantasien til hjelp. Benevnelser og arbeidsmetoder kan kanskje virke litt fremmed i første øieblikk, men setter man sig inn i saken, er den i virkeligheten enkel. Særlig i vår tid hvor kryssordoppgavene er så utbredt, vil mange kanskje finne at arbeidet med et kryptogram er en behagelig avveksling. Her er kryptogrammet:

æsxlc oawbw jfcdø fxfla æføvø rjfyx løwbx
lhjeb xlmxy yxoxf øbyxf wøoyb wrfxe hfxlv
æjmxæ fræsx lalyx foawb ølvfx hxbwf xyøvm
xlabx ldrfp lrlv

Til lettelse ved løsningen skal nevnes at kryptogrammet er formet som en spionmelding fra et krigførende land, og at ordet «bomber» forekommer i kryptogrammet. Som en lettelse settes likeledes op i en tabell de forskjellige bokstavers antall og deres forbindelse med E som i kryptogrammet sees å være x.

ETCRRM

- Bjørn Rørholt (1919-1993)
- “One-Time-Pad” using radioactive source
- Constructed by STK in the 50s (now Thales)
- Used to secure the first telex hot line between Kremlin and the White House



PACE - Pocket Automatic Crypto Equipment

- Hand-held device for off-line encryption/decryption
- Produced in 20.000 copies
- Uses an error-correcting code for transmission of data between two devices
- Developed by Lehmkuhl AS in 1970-80s sold to Thales and further to Kongsberg AS

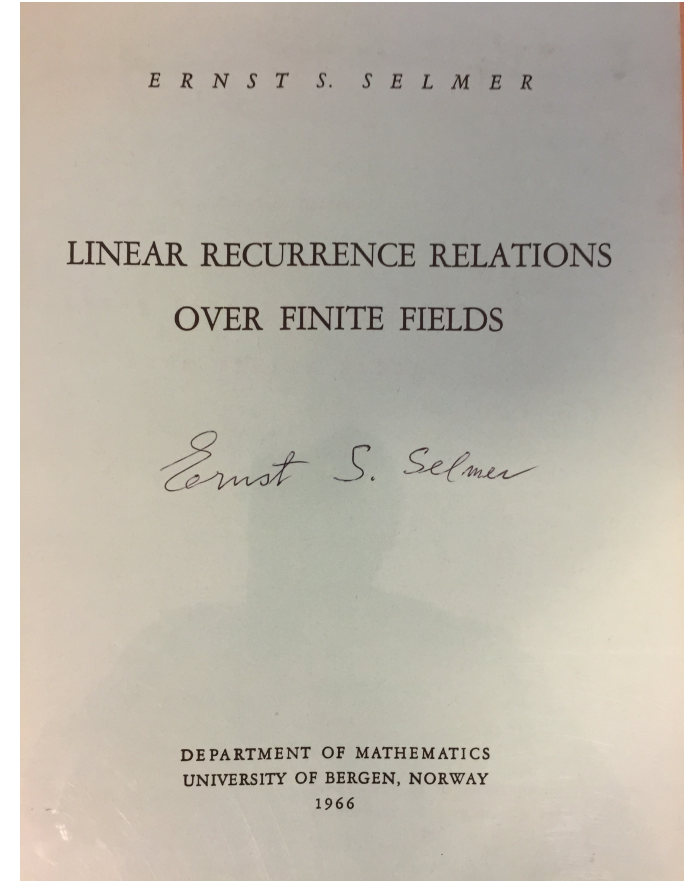


Ernst S. Selmer (1920-2006)



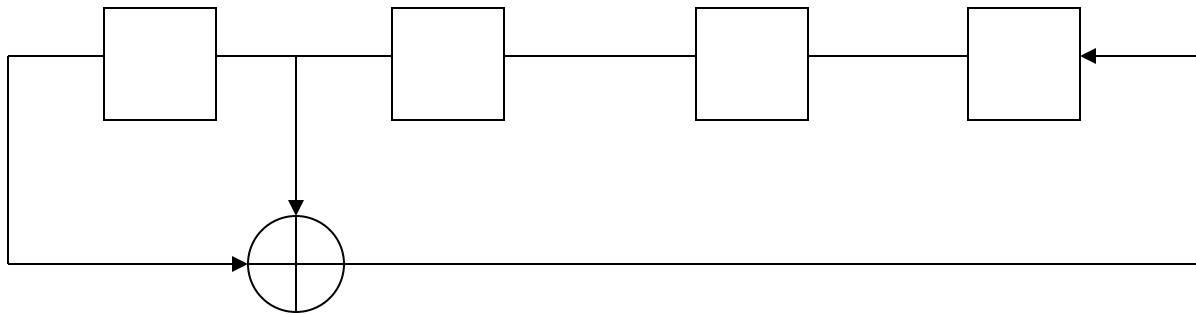
- Professor in Mathematics
University of Bergen
(1957-1990)
- Cryptographer in WW II
- Pioneer in cryptographer
and coding in Norway

- Designed **error control** in Norwegian social security numbers in 1964
- **Monograph (1966)**
 - “Linear Recurrence Relations over Finite Fields”
 - Sold 200 copies in 10 minutes at Eurocrypt 1993
- Selmer groups & Fermat’s theorem



Sequences

m-Sequence (Binary Example)



$$s_{t+4} = s_{t+1} + s_t$$

$$g(x) = x^4 + x + 1$$

$(s_t) : 000100110101111...$

Properties of m-sequences

- Period $\varepsilon = 2^n - 1$
 - Balanced (except for a missing 0)
 - Run properties
 - Shift-and-add property $s_t + s_{t+\tau} = s_{t+\gamma}$
- Decimation property $s_{2d} = s_{t+\gamma}$
- Trace representation



$$s_t = \text{Tr}_n(A\alpha^t) = \sum_j (A\alpha^t)^{2^j} = A_1\alpha^t + A_2\alpha^{2t} + A_3\alpha^{4t} + A_4\alpha^{8t}$$

The Simplex Code

- C is a linear $[N,k,d]$ code if
 - C is a k dimensional subspace of $\{0,1\}^N$
 - $d = \min\{d_H(c_1, c_2) : c_1, c_2 \in C\}$

where d_H denotes the Hamming distance.

- The m-sequence and all its shifts and (0) form a linear code with parameters

$$[2^n-1, n, 2^{n-1}]$$

Correlation of sequences

- Let (a_t) and (b_t) be binary sequences of period ε over the alphabet $GF(2)$
- The **crosscorrelation** between (a_t) and (b_t) at shift τ is

$$\theta_{a,b}(\tau) = \sum_{t=0}^{\varepsilon-1} (-1)^{a_{t+\tau} - b_t}$$

- The **autocorrelation** of (a_t) at shift τ is

$$\theta_{a,a}(\tau) = \sum_{t=0}^{\varepsilon-1} (-1)^{a_{t+\tau} - a_t}$$

Two-level autocorrelation of m-sequences

- Let (s_t) be an m-sequence of period $\varepsilon=2^n-1$
- Then the autocorrelation of the m-sequence is

$$\begin{aligned}\theta_{s,s}(\tau) &= 2^n - 1 && \text{if } \tau = 0 \pmod{2^n - 1} \\ &= -1 && \text{if } \tau \neq 0 \pmod{2^n - 1}\end{aligned}$$

Proof: Let $\tau \neq 0 \pmod{2^n-1}$. Then

$$\begin{aligned}\theta_{s,s}(\tau) &= \sum_t (-1)^{s_{t+\tau} - s_t} \\ &= \sum_t (-1)^{s_{t+\gamma}} \\ &= -1 \quad (\text{since m-sequence is balanced})\end{aligned}$$

Binary 3-valued crosscorrelation

- $C_d(\tau)$ has exactly 3 different values in the cases:
 - **Gold** : $d = 2^k + 1$ where $n/(n,k)$ is odd
 - **Kasami** : $d = 2^{2k} - 2^k + 1$ where $n/(n,k)$ is odd
 - **Welch's conjecture**: (Canteau, Charpin, Dobbertin 2000)
 $d = 2^m + 3$ where $n=2m+1$ is odd
 - **Niho's conjecture** : (Dobbertin & Hollman and Xiang)
 $d = 2^{(n-1)/2} + 2^{(n-1)/4} - 1$ when $n \equiv 1 \pmod{4}$
 $= 2^{(n-1)/2} + 2^{(3n-1)/4} - 1$ when $n \equiv 3 \pmod{4}$
 - **Cusick and Dobbertin** (Cusick and Dobbertin 1996)
 $d = 2^{n/2} + 2^{(n+2)/2} + 1$ when $n \equiv 2 \pmod{4}$
 $d = 2^{(n+2)/2} + 3$ when $n \equiv 2 \pmod{4}$

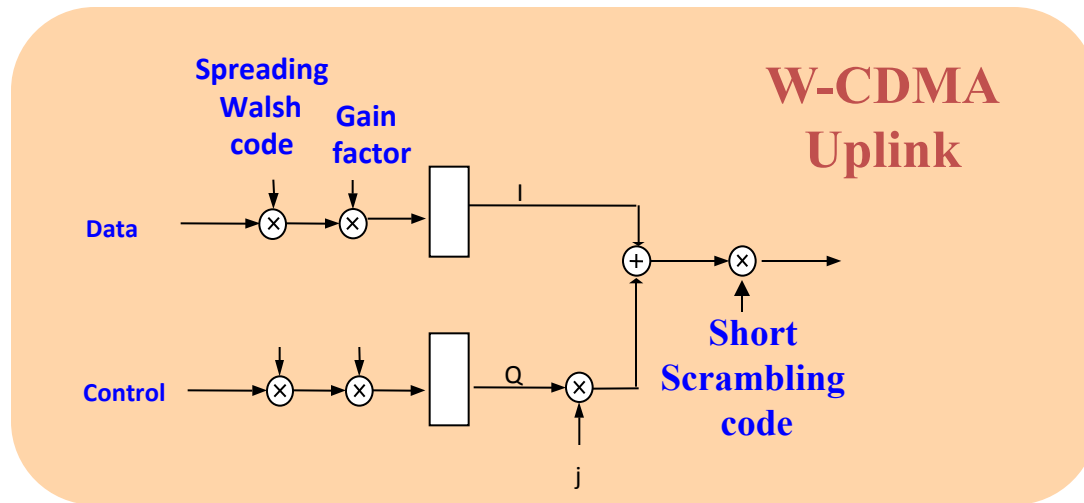
Open problem:

Are these all the cases with 3-valued crosscorrelation?

Applications to CDMA

- In Code-Division Multiple Access(CDMA) one needs large families F of sequences with good correlation properties
- Parameters of a family is $(\varepsilon, M, \theta_{\max})$
 - ε period of the sequences in F
 - M size of family (# of cyclically distinct sequences)
 - θ_{\max} maximal absolute value of the (nontrivial) auto- or cross correlation between any two distinct sequences on the family

Scrambling Code Design for 3G Wireless Cellular Communication



P. V. Kumar

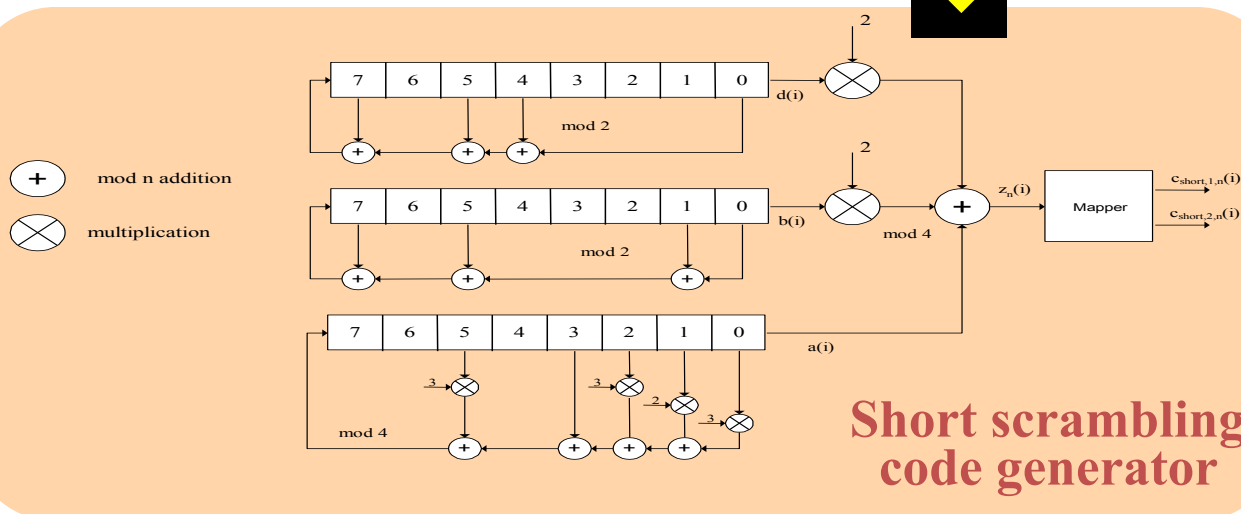
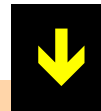
T. Helleseth

A. R. Calderbank

A. R. Hammons Jr.,

"Large Families of Quaternary Sequences with Low Correlation,"

IEEE Trans. Inform. Theory, March 1996.



Short Scrambling Code Family $S(2)$ used in W-CDMA



S-boxes/APN/AB

Sequences and S-boxes in cryptography

- S-box ($n \times n$) is a mapping $f : GF(2^n) \rightarrow GF(2^n)$
- Need good differentiability
 $f(x+a) + f(x) = b$ has “few” solutions for any $a \neq 0, b$
- Need good nonlinearity
 $f(x)$ has “large distance” to all linear functions
- Aiming for
Almost Perfect Nonlinear functions (APN)
Almost Bent functions (AB)

APN and AB functions

- A function $f : GF(2^n) \rightarrow GF(2^n)$ is APN if

$$f(x + a) + f(x) = b$$

has at most two solutions for any $a \neq 0$, b in $GF(2^n)$.

- The Walsh transform of f is defined by

$$\lambda_f(a, b) = \sum_{x \in GF(2^n)} (-1)^{\text{Tr}(af(x) + bx)}$$

- A function $f(x)$ is AB if

$$\{ \lambda_f(a, b) : a, b \in GF(2^n) \} = \{0, \pm 2^{(n+1)/2} \}$$

- $AB \Rightarrow APN$ (Chabaud and Vaudenay 1994)
- Monomial AB functions where $f(x) = x^d$ can be obtained from Gold sequences and several of the decimations with 3-valued cross correlation

Known AB power functions x^d

- Gold: $d=2^k+1$ where $(k,n)=1$
- Kasami: $d=2^{2k}-2^k+1$ where $(k,n)=1$
- Welch: $d=2^m+3$ where $n=2m+1$ is odd
- Niho:
$$d = 2^{(n-1)/2} + 2^{(n-1)/4} - 1 \quad \text{when } n \equiv 1 \pmod{4}$$
$$= 2^{(n-1)/2} + 2^{(3n-1)/4} - 1 \quad \text{when } n \equiv 3 \pmod{4}$$
- Three-valued cross correlation with values $\{-1, -1 \pm 2^{(n+1)/2}\}$
- Open problem(Dobbertin): Is this list complete?

Known APN power functions x^d

- Gold: $d = 2^k + 1$ where $(k, n) = 1$
- Kasami: $d = 2^{2k} - 2^k + 1$ where $(k, n) = 1$
- Welch: $d = 2^m + 3$ where $n = 2m + 1$ is odd
- Niho:
$$d = 2^{(n-1)/2} + 2^{(n-1)/4} - 1 \quad \text{when } n \equiv 1 \pmod{4}$$
$$= 2^{(n-1)/2} + 2^{(3n-1)/4} - 1 \quad \text{when } n \equiv 3 \pmod{4}$$
- Inverse: $d = 2^n - 2 \quad (= -1 \pmod{2^n - 1})$ where n odd
- Dobbertin: $d = 2^{4k} + 2^{3k} + 2^{2k} + 2^k - 1$ where $n = 5k$

Open problem(Dobbertin): Is this list complete?

Sequences and the S-box in AES

- The cross correlation between m-sequence (s_t) and reverse sequence (s_{-t}) corresponds to the famous Kloosterman sum

$$C_{-1}(\tau) = \sum_{x \neq 0} (-1)^{\text{Tr}(ax + bx^{-1})}$$

- Bound for Kloosterman sum

$$|C_{-1}(\tau) + 1| \leq 2 \cdot 2^{n/2}$$

- The AES S-box is based on $f(x) = x^{-1}$ for $n=8$. The correlation between x^{-1} and all linear functions is bounded by $|C_{-1}(\tau)|$
- The S-box is 4-uniform (not APN), the best known for $n=8$.
- The S-box is not AB but the correlation (and nonlinearity) is the best known for $n=8$.

Authentication Codes

Codes Which Detect Deception (1974)

(E.N. Gilbert, F.J. MacWilliam, N.J.A. Sloane)

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 53

March 1974

Number 3

Copyright © 1974, American Telephone and Telegraph Company. Printed in U.S.A.

Codes Which Detect Deception

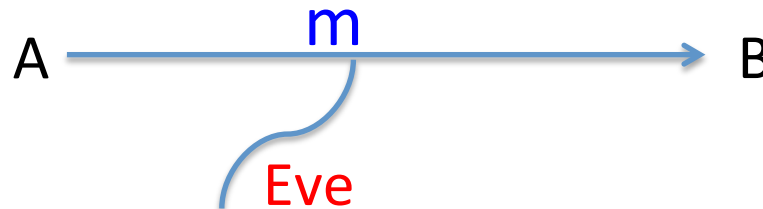
By E. N. GILBERT, Mrs. F. J. MacWILLIAMS, and N. J. A. SLOANE

(Manuscript received May 15, 1973)

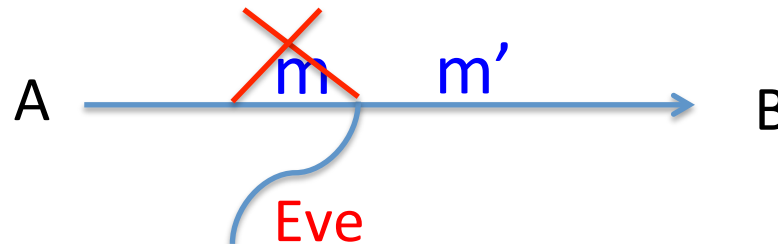
We consider a new kind of coding problem, which has applications in a variety of situations. A message x is to be encoded using a key m to form an encrypted message $y = \Phi(x, m)$, which is then supplied to a user G . G knows m and so can calculate x . It is desired to choose $\Phi(\cdot, \cdot)$ so as to protect G against B , who knows x , y , and $\Phi(\cdot, \cdot)$ (but not m); B may substitute a false message y' for y . It is shown that if the key can take K values, then an optimal strategy for B secures him a probability of an undetected substitution $\geq K^{-1}$. Several encoding functions $\Phi(\cdot, \cdot)$ are given, some of which achieve this bound.

Two attacks

- Impersonation attack



- Substitution attack



- Authentication codes
 - Compute $t = h_K(m)$ where K is authentication key (Send message and tag (m, t) to B)
 - Ideally $P_I \approx P_S \approx 2^{-r}$, probability of successful attack
 - Coding theory is behind many good constructions

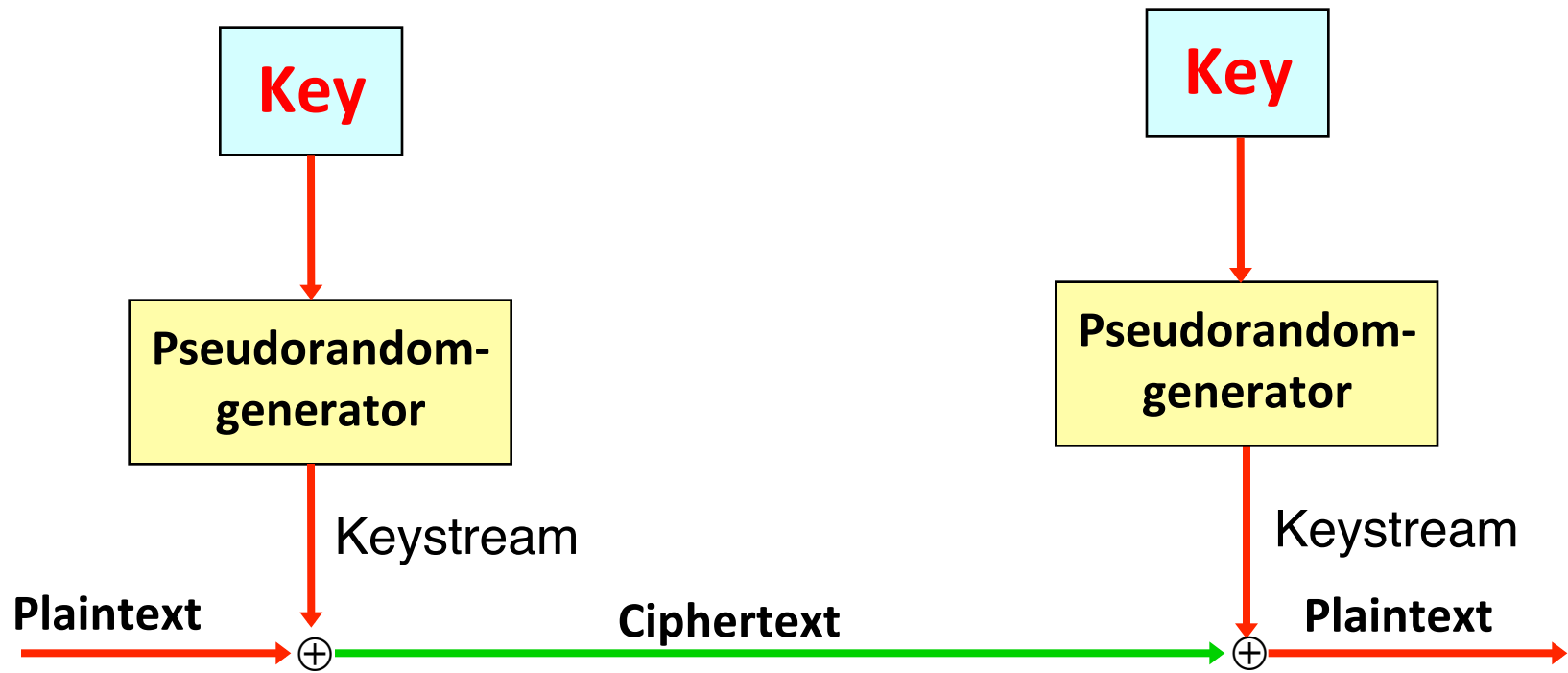
Authentication Codes

$K \backslash m$	m_1	m_2	...	m_j	...	m_b
k_1						
k_2						
...						
k_i				t_{ij}		
...						
k_q						

- Tag $t = h_k(m)$
- Need balanced columns for good P_i
- Need good balance between pairs of columns for good P_s
- Can make A-codes based on (non-binary) codes and sequences
- GMAC based on Reed-Solomon codes

Stream Ciphers

Stream Cipher



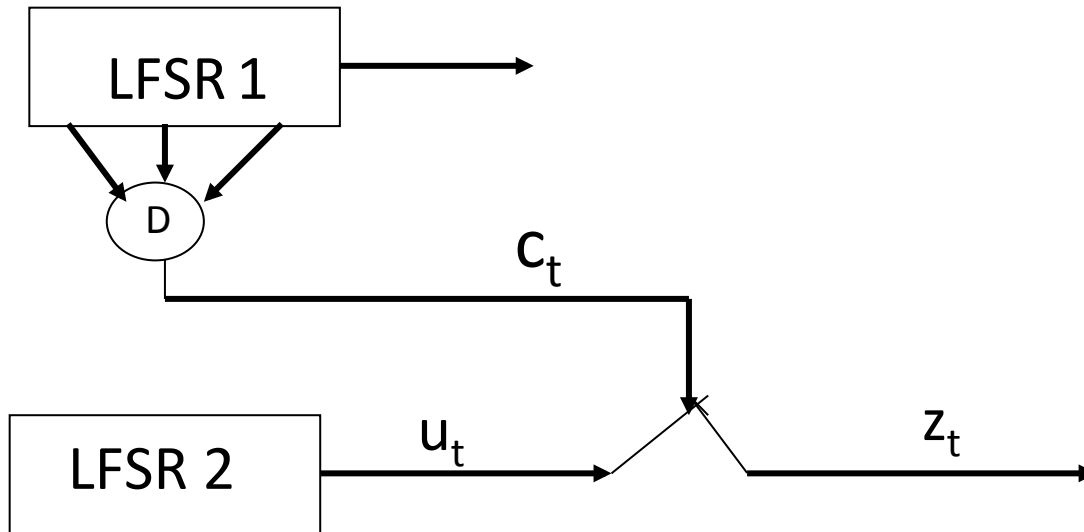
Requirements for a good keystream

- **Good randomness distribution**
- **Long period**
- **High complexity**

Nonlinear Components in Stream Cipher

- Techniques to get higher linear complexity
 - The LFSRs are **clocked irregularly**
 - The LFSR bits are sent through a **nonlinear function**
 - **Nonlinear combiner** (several shift registers)
 - Attacks are using correlation attacks
(based on coding theory)
 - **Filter generator** (one shift register)
 - Algebraic attacks
(solving nonlinear equations)

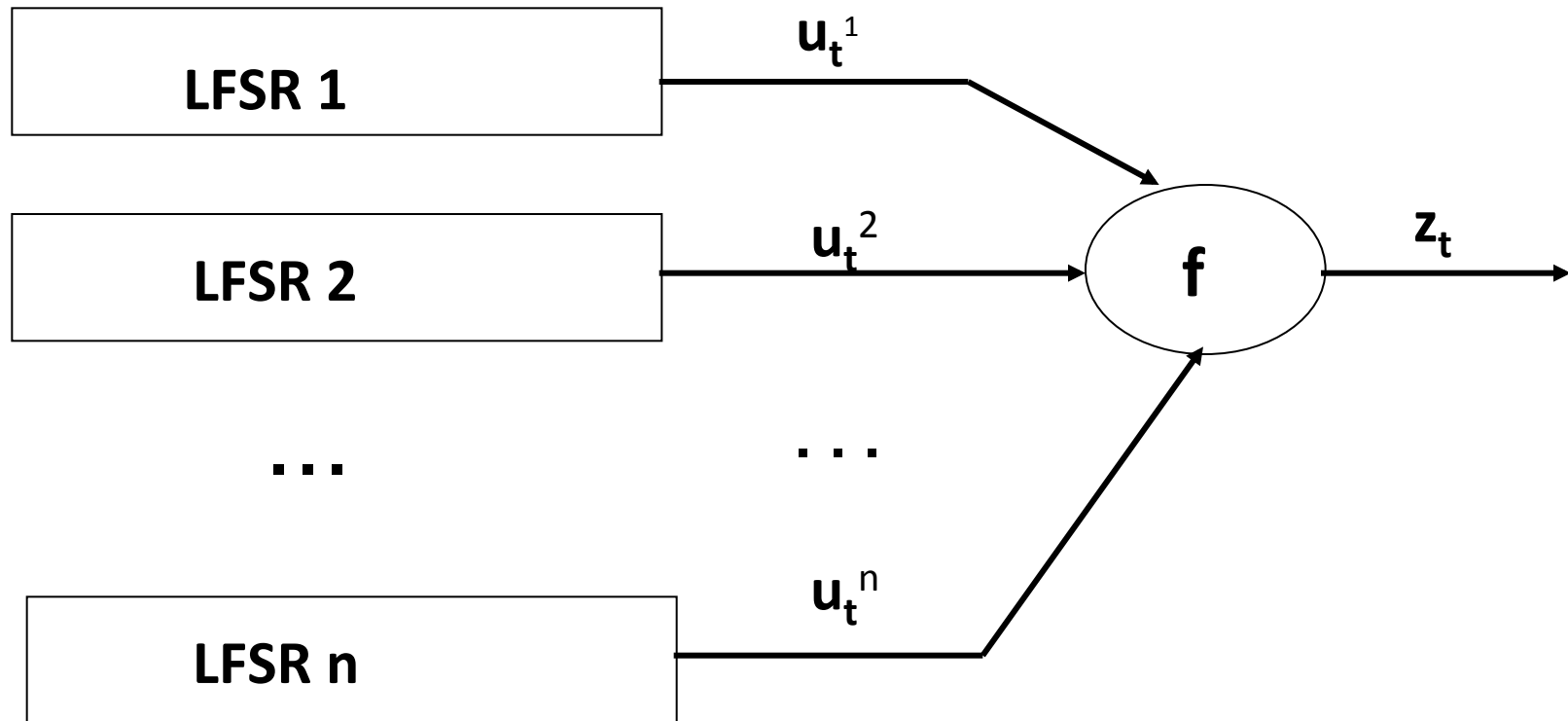
Clock Controlled LFSRs



- **LFSR 1** generates an m-sequence mapped by D to an integer clock sequence c_t used to select the bits in u_t generated by **LFSR 2** to be the output bit z_t

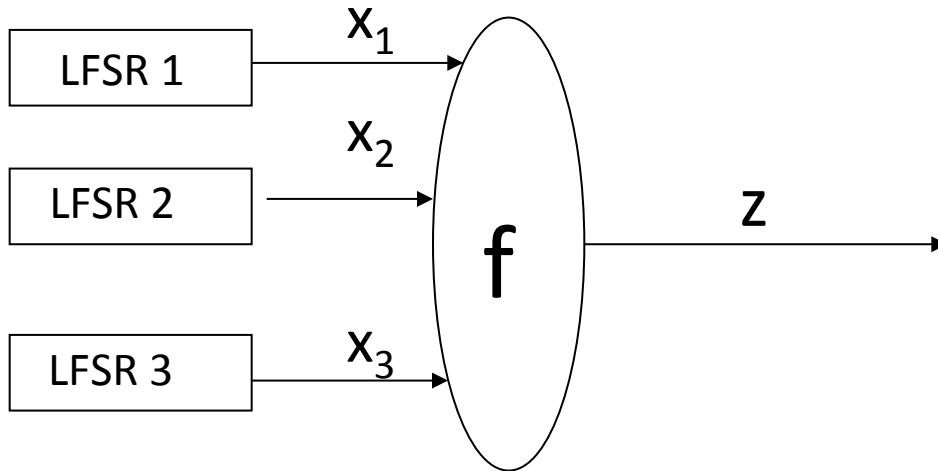
Nonlinear Combining LFSRs

Using several LFSRs



$$f(x_1, x_2, \dots, x_n) = \sum a_{i_1 i_2 \dots i_n} x_{i_1} x_{i_2} \dots x_{i_n}$$

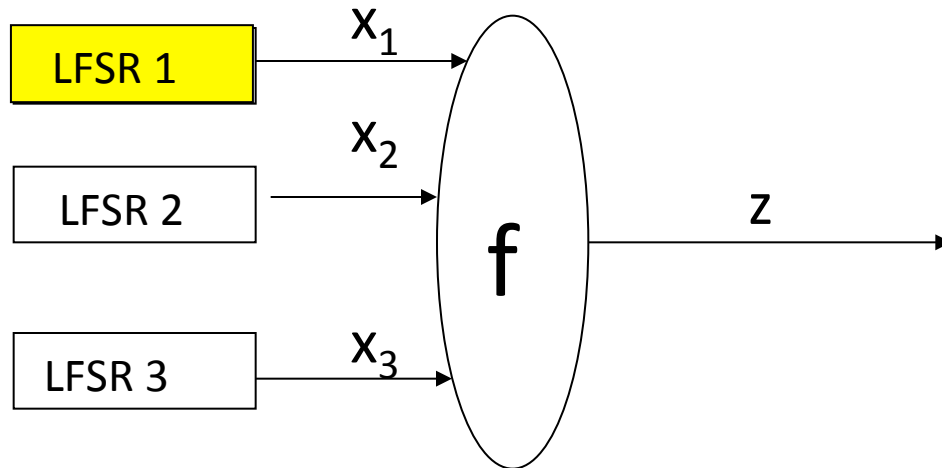
Geffe generator



The LFSRs generate m-sequence of period $2^{n_i} - 1$, $\gcd(n_i, n_j) = 1$

- $z = f(x_1, x_2, \dots, x_n) = x_1x_2 + x_2x_3 + x_3$
- $x_2 = 1 \rightarrow f = x_1$
- $x_2 = 0 \rightarrow f = x_3$
- Period = $(2^{n_1} - 1)(2^{n_2} - 1)(2^{n_3} - 1)$
- Linear complexity = $n_1n_2 + n_2n_3 + n_3$

Correlation attack - Geffe generator



Correlation attack of Geffe generator

(NB! $\text{Prob}(z = x_1) = \frac{3}{4}$)

- Guess the initial state of LFSR 1
- Compare x_1 and z
 - If agreement $\frac{3}{4}$, guess is likely to be correct
 - If agreement $\frac{1}{2}$, guess is likely to be wrong

Fast correlation attacks

- Need a correlation $p \neq 0.5$ between keystream and register
- Do **not** need to guess a full register
- Construct a new linear code where bits are linear combinations of a subset of bits in initial state of register.
- Each code position estimated by few $w \leq 4$ keystream bits
- Ideas from coding theory are used to construct the closest codeword i.e., bits in the subset
- Efficient implementations of Viterbi decoder with rate $R = 10^{-10}$ and error probability $p = 0.49$

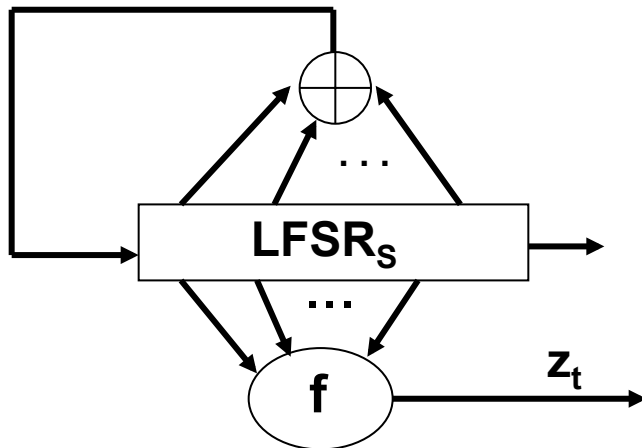
Filter Generator

Filter Generator

LFSR of length n generating an m-sequence

(s_t) of period $2^n - 1$ determined by initial state $(s_0, s_1, \dots, s_{n-1})$

Nonlinear Boolean function $f(x_0, x_1, \dots, x_{n-1})$ of degree d

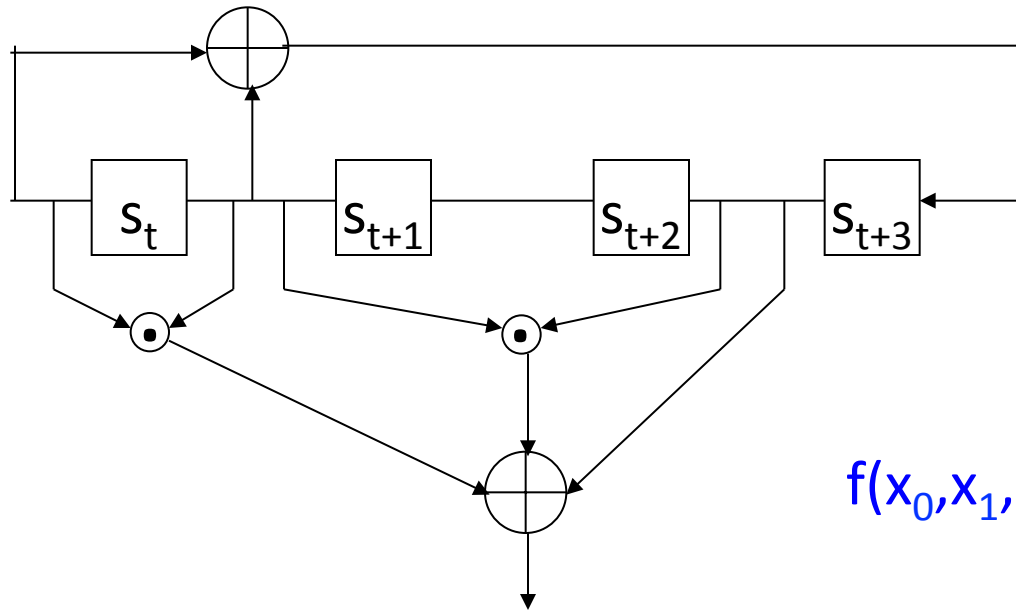


Keystream

$$\begin{aligned} z_t &= f(s_t, s_{t+1}, \dots, s_{t+n-1}) \\ &= f_t(s_0, s_1, \dots, s_{n-1}) \end{aligned}$$

$$f(x_0, x_1, \dots, x_{n-1}) = \sum c_{i_0 i_1 \dots i_{r-1}} x_{i_0} x_{i_1} \dots x_{i_{r-1}} = \sum_l c_l x_l$$

Example – Filter Generator



$$g(x) = x^4 + x + 1$$

$$s_{t+4} = s_{t+1} + s_t$$

$$f(x_0, x_1, x_2, x_3) = x_0x_1 + x_1x_3 + x_3$$

$$z_t = s_t s_{t+1} + s_{t+1} s_{t+3} + s_{t+3}$$

$$z_0 = f(s_0, s_1, s_2, s_3) = s_0 s_1 + s_1 s_3 + s_3 \quad (= f_0)$$

$$z_1 = f(s_1, s_2, s_3, s_4) = f(s_1, s_2, s_3, s_0 + s_1) = s_0 + s_1 + s_0 s_2 \quad (= f_1)$$

$$z_2 = f(s_2, s_3, s_4, s_5) = f(s_2, s_3, s_0 + s_1, s_1 + s_2) = s_1 + s_2 + s_1 s_3 \quad (= f_2)$$

.....

Multivariate Equations

$$z_0 = s_0 s_1 + s_1 s_3 + s_3$$

$$z_1 = s_0 s_2 + s_0 + s_1$$

$$z_2 = s_1 s_3 + s_1 + s_2$$

$$z_3 = s_0 s_2 + s_1 s_2 + s_2 + s_3$$

$$z_4 = s_1 s_3 + s_2 s_3 + s_0 + s_1 + s_3$$

$$z_5 = s_0 s_2 + s_0 s_3 + s_1 s_2 + s_1 s_3 + s_0 + s_1 + s_2 \quad \dots$$

Linearization gives a linear system with $\binom{4}{2} + \binom{4}{1} = 10$ unknowns

$$z_0 = a_4 + a_8 + a_3$$

$$z_1 = a_5 + a_0 + a_1$$

$$z_2 = a_8 + a_1 + a_2$$

$$z_3 = a_5 + a_7 + a_2 + a_3$$

$$z_4 = a_8 + a_9 + a_0 + a_1 + a_3$$

$$z_5 = a_5 + a_6 + a_7 + a_8 + a_0 + a_1 + a_2 \quad \dots$$

Solve by using Gaussian elimination

Standard Linearization Attack

- Shift register **m-sequence** (s_t) of period $2^n - 1$
- Boolean function $f(x_0, x_1, \dots, x_{n-1})$ of **degree d**
$$z_t = f(s_t, s_{t+1}, \dots, s_{t+n-1}) = f_t(s_0, s_1, \dots, s_{n-1})$$
- **Nonlinear equation system** of degree **d** in **n** unknowns s_0, \dots, s_{n-1}
- Reduce to linear system: **D** unknown monomials
- $D = \binom{n}{d} + \binom{n}{d-1} + \dots + \binom{n}{1}$
- Need about **D** keystream bits
- **Complexity** D^ω , $\omega = \log_2 7 \approx 2.807$

Example - Coefficient Sequences

- Let $s_{t+4}=s_{t+1}+s_t$ i.e., $s_4=s_1+s_0$
- $f(x_0,x_1,x_2,x_3) = x_2+x_0x_1+x_1x_2x_3+x_0x_1x_2x_3$
- $z_t = f(s_t,s_{t+1},s_{t+2},s_{t+3}) = s_{t+2} + s_t s_{t+1} + s_{t+1} s_{t+2} s_{t+3} + s_t s_{t+1} s_{t+2} s_{t+3}$

$$\begin{aligned}
 z_0 &= f_0(s_0,s_1,s_2,s_3) = s_2 + s_0s_1 + s_1s_2s_3 + s_0s_1s_2s_3 \\
 z_1 &= f_1(s_0,s_1,s_2,s_3) = s_3 + s_1s_2 + s_0s_2s_3 + s_0s_1s_2s_3 \\
 z_2 &= f_2(s_0,s_1,s_2,s_3) = s_0 + s_1 + s_1s_3 + s_2s_3 + s_0s_1s_3 + s_1s_2s_3 + s_0s_1s_2s_3 \\
 z_3 &= f_3(s_0,s_1,s_2,s_3) = s_1 + s_2 + s_0s_2 + s_0s_3 + s_1s_3 + s_0s_1s_2 + s_0s_2s_3 + s_0s_1s_2s_3 \\
 z_4 &= f_4(s_0,s_1,s_2,s_3) = s_1 + s_2 + s_3 + s_0s_1 + s_0s_2 + s_1s_2 + s_0s_1s_3 + s_0s_1s_2s_3 \\
 z_5 &= f_5(s_0,s_1,s_2,s_3) = s_0 + s_1 + s_2 + s_3 + s_1s_3 + s_2s_3 + s_0s_1s_2 + s_0s_1s_3 + s_0s_1s_2s_3
 \end{aligned}$$

Some coefficient sequences

$$l=\{0,1,2,3\} \quad K_{l,t} = 1 \ 1 \ 1 \ 1 \ 1 \ 1 \dots$$

$$l=\{0,2,3\} \quad K_{l,t} = 0 \ 1 \ 0 \ 1 \ 0 \ 0 \dots$$

$$l=\{1,3\} \quad K_{l,t} = 0 \ 0 \ 1 \ 1 \ 0 \ 1 \dots$$

Rønjom-Helleseeth Algebraic Attack

- Recovering **initial state** of the binary filter generator in complexity
 - Pre-computation $O(D (\log_2 D)^3)$
 - Attack $O(D)$
 - Need D keystream bits
- **Main idea** - **Coefficient sequences of $I=\{i_0, i_1, \dots, i_{r-1}\}$**
 - Consider (binary) coefficient $K_{I,t}$ in $f_t(s_0, s_1, \dots, s_{n-1})$ of the monomial $s_I = s_{i_0} s_{i_1} \dots s_{i_{r-1}}$ at time t
 - $K_{I,t}$ obeys some nice recursions

Multivariate - Univariate

Let $x = \sum_i x_i \alpha_i$ where $\alpha_0, \dots, \alpha_{n-1}$ basis $\text{GF}(2^n)$, $x_i = \{0, 1\}$

- 1-1 correspondence $\text{GF}(2)^n \leftrightarrow \text{GF}(2^n)$
- $(x_0, \dots, x_{n-1}) \leftrightarrow x$
- Then Boolean function "becomes univariate"

$$f(x_0, \dots, x_{n-1}) = f(x)$$

for some polynomial $f(x)$ in $\text{GF}(2^n)[x]$ of degree at most $2^n - 2$ (if we do not care for the value at 0)

- The degree d of $f(x_0, \dots, x_{n-1})$ is the largest $\text{wt}(j)$ such that coefficient in $f(x)$ of x^j is nonzero

Rønjom-Helleseth Attack - Univariate

- Let L be the shift operator of the LFSR

$$L(s_t, \dots, s_{t+n-1}) = (s_{t+1}, \dots, s_{t+n})$$

- Define $f(\alpha^t) = f(L^t(s_0, \dots, s_{n-1}))$

- Let x denote the unknown initial state, then

$$z_t = f(x\alpha^t) \text{ where we want to find } x$$

- Univariate equation system in x ($q=2^n$)

$$z_0 = f_0(x) = c_0 + c_1 x + \dots + c_{q-2} x^{q-2}$$

$$z_1 = f_1(x) = c_0 + c_1 \alpha x + \dots + c_{q-2} \alpha^{q-2} x^{q-2}$$

$$z_2 = f_2(x) = c_0 + c_1 \alpha^2 x + \dots + c_{q-2} \alpha^{2(q-2)} x^{q-2}$$

.....

Algebraic attacks of $f(x_0, \dots, x_{n-1})$

Definition

The Boolean function $g(x_0, \dots, x_{n-1})$ is an **annihilator** of $f(x_0, \dots, x_{n-1})$ if

$$f(x_0, \dots, x_{n-1}) g(x_0, \dots, x_{n-1}) = 0 \text{ for all } x_0, \dots, x_{n-1} \in \{0, 1\}$$

Definition

The **algebraic immunity (AI)** of f

$$AI(f) = \min\{\deg(g) \mid f g = 0 \text{ or } (1+f) g = 0\}$$

Hence if $z_t=1$ then

$$\begin{aligned} f(s_t, \dots, s_{t+n-1}) g(s_t, \dots, s_{t+n-1}) &= z_t g(s_t, \dots, s_{t+n-1}) \\ &= g_t(s_0, \dots, s_{n-1}) = 0 \end{aligned}$$

Coding theory – Cyclic Codes

Definition – Linear $[N,k,d]_q$ code

C is an $[N,k,d]_q$ code iff

- 1) C subset of dimension k over $GF(q)^N$
- 2) $d = \min\{d_H(c_1, c_2) \mid c_1 \neq c_2 \in C\}$

Definition – Cyclic code

$$C = (G(x)) \pmod{x^n-1}$$

(= Ideal generated by $G(x)$)

Coding and algebraic attack

Theorem

Let $f(x)$ be a Boolean function in **univariate** form, $q=2^n$.
Then any **annihilator** $g(x)$ of $f(x)$ belongs to the **q -ary** cyclic code C_f with generator polynomial

$$G_f(x) = \gcd(f(x)+1, x^{q-1}+1)$$

Proof: Let $g(x)$ be annihilator of $f(x)$, then $f(x)g(x)=0$ for all x in $GF(2^n)$. Then $f(x)g(x) = 0 \pmod{x^{q-1}+1}$.

Hence, $g(x) = 0 \pmod{\gcd(f(x)+1, x^{q-1}+1)}$.

”Need to find special codewords ”

- $g(x)$ in C_f (and C_{f+1}) of smallest **$\max\{\text{wt}(j) \mid g_j \neq 0\}$** .
- $g(x) \in \{0,1\}$ for all x in $GF(2^n)$

Spectral Immunity

Definition

The spectral immunity of (z_t) is the smallest linear complexity(LC) of a sequence (u_t) over $GF(2^n)$ such that $z_t u_t = 0$ or $(1+z_t) u_t = 0$ for all t

Let $z_t = f(x\alpha^t)$ and $u_t = g(x\alpha^t)$ where (u_t) annihilates (z_t)
Then if $z_t = 1$ we obtain

$$g(x\alpha^t) = 0 \rightarrow \sum g_i \alpha^{ti} x^i = 0 \quad (\text{Note: } wt(g) = LC(u_t))$$

- Linear system in the LC unknowns $x^{i1}, x^{i2}, \dots, x^{iLC}$
- Knowing $2 \cdot LC(u_t)$ bits finds x^{i1}, \dots and hence x

Spectral Immunity and Cyclic Codes (I)

Theorem

Let $z_t = f(x\alpha^t)$ and $u_t = g(x\alpha^t)$ be such that

$$f(x) g(x) = 0 \text{ for all } x \text{ in } GF(2^n)$$

Then $g(x)$ is a codeword in the cyclic code C_f with symbols from $GF(2^n)$ and generator polynomial

$$G_f(x) = \gcd(f(x)+1, x^{q-1}+1)$$

Proof:

Follows since $f(x)$ is Boolean and **only takes on the values 0 and 1**. Therefore the elements in $GF(2^n)$ are zeros of either $f(x)$ or $f(x)+1$

Spectral immunity and cyclic codes (II)

Theorem

The **spectral immunity (SI)** of (z_t) is the smallest **weight of a codeword** in the codes over $GF(2^n)$ with generator polynomials

$$G_f = \gcd(f(x)+1, x^{q-1}+1)$$

$$G_{f+1} = \gcd(f(x), x^{q-1}+1)$$

Corollary

$$SI \leq D = \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{AI}$$

SI versus AI

Corollary

$$SI \leq D = \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{AI}$$

- $SI \text{ large} \Rightarrow AI \text{ large}$
- $AI \text{ Large} \not\Rightarrow SI \text{ large}$

Can use codes G_f and G_{f+1} to evaluate AI

$$AI = \min_c \max_i \{wt(i) \mid c_i \neq 0 \text{ for } c(x) \text{ in } C_f \text{ or } C_{f+1}\}$$

Open problems

- What are minimum distance of the codes C_f ?
- How much better is the spectral Immunity (SI) compared to Algebraic immunity (AI)?
- How to use the spectral immunity in an optimal way. This may be a challenge since (SI) is based on the univariate representation while (AI) depends on the multivariate representation.
- The method works well to attack many variants of the WG cipher family (Rønjom 2015)

The Future

ArcticCrypt

A New Northern World Record

Northernmost Crypto Conferences

(Top 5 ranking)

- Espoo (1998) 60' N
- Lofthus (1993) 60' N
- St. Petersburg (2006) 59' N
- Tallin (2011) 59' N
- Aarhus (2005) 56' N

Arctic Crypt




July 17-22, 2016, Longyearbyen, Svalbard, (Norway)

<http://www.selmer.uib.no/ArcticCrypt/>

- Location: 78 degrees north
 - Latitude: $78^{\circ}13'11''$ N
 - Longitude: $15^{\circ}39'00''$ E
 - Elevation above sea level: 1 m = 03 ft

Why go to Arctic Crypt?

- Fantastic scenery – Glaciers, Wildlife, Mountains
 - Midnight sun the whole week
 - One Full Day of sightseeing
- 
- Hotel gives you a gun (and ammunition) as protection against polar bears if you hike outside Longyearbyen
- “Probability of being shot by a tourist is higher than probability of being killed by a polar bear”**

Scientific Program

Program co-chair **Bart Preneel**

- 4 days of lecturers
- 10 Invited lectures
- Contributed talks by submissions
- Support for younger researchers
- “Midnight lecture”
- **Speakers should not hike outside Longyearbyen before giving their talk**

