On Cryptographic Properties of Boolean Function

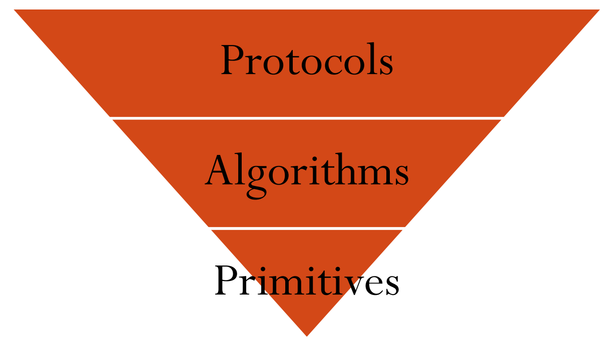
Amr Youssef
Concordia Institute for Information Systems Engineering (CIISE)
Concordia University
Montreal, Canada

Outline

- Motivation
- Boolean functions representations
- Cryptographic properties of Boolean functions
- Construction examples
- Conclusions and open problems

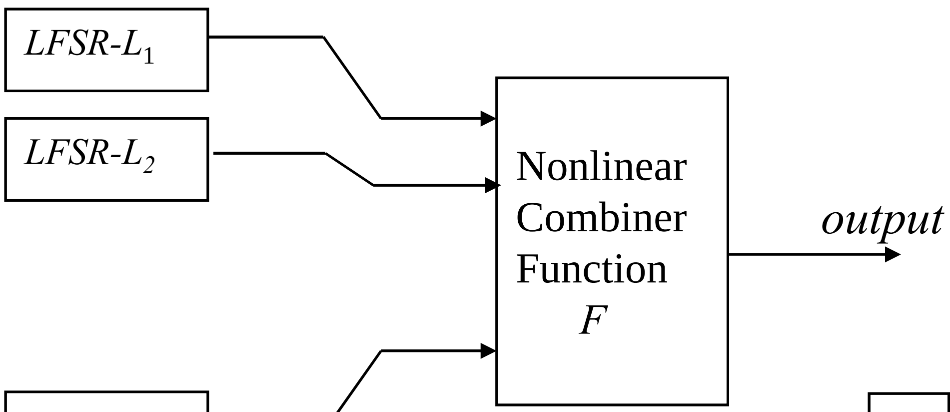
Motivation

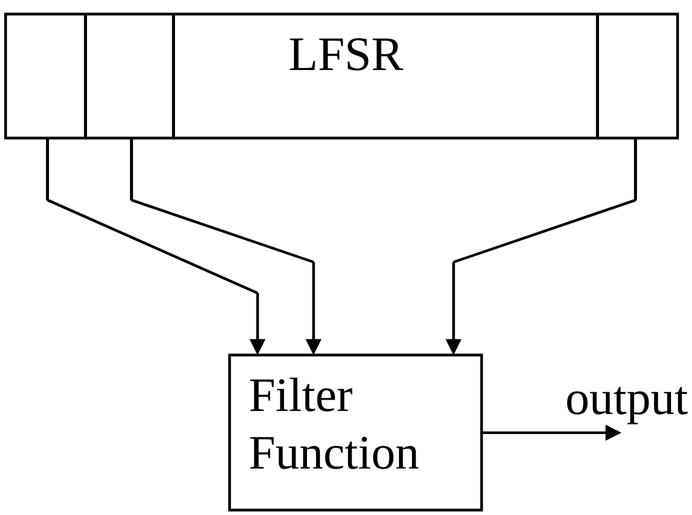
- Hierarchical subdivisions of cryptography
  - Protocols (e.g., Needham Schroeder)
    - Produce solutions for cryptographic problems
  - Algorithms (e.g., AES)
    - Used to construct protocols
  - Primitives
    - Used to construct algorithms



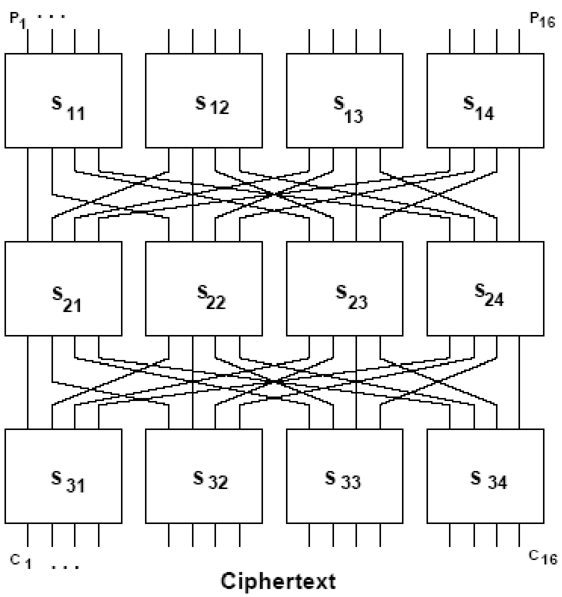
- Boolean functions
    - Constitute one of the basic primitives for symmetric key cryptography
    - Strong connection between cryptanalytic attacks and the properties of the underlying Boolean functions
    - Some attempts for use in public key cryptography

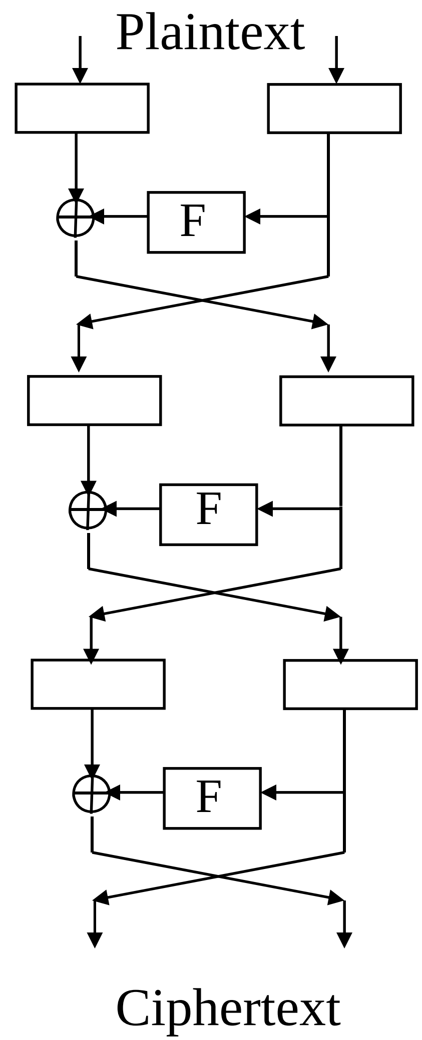
Classical examples for stream ciphers





Classical examples for Block ciphers





Boolean Functions

- A Boolean function in n variables $f : F_2^n \to F_2$
- Multiple-output Boolean functions $f : F_2^n \to F_2^m$
  - Also known as
    - S-Boxes
    - Vectorial Boolean functions
- $B_{n,m}$: the set of "Boolean" functions $f : F_2^n \to F_2^m$
  - $|B_{n,m}| = 2^{m2^n}$
  - Exhaustive search is not an option

Boolean function Representation

Truth Table

x_1	x_2	$f(x_1, x_2)$
0	0	1
0	0	1
0	1	1
0	1	0

$$f(x_1, x_2) = 1 + x_1 x_2$$

Algebraic Normal Form (ANF)

$$f(x_1, x_2, \cdots, x_n) = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n + a_{12} x_1 x_2 + a_{13} x_1 x_3 + \cdots + \cdots + a_{12\cdots n} x_1 x_2 \cdots x_n$$

$$f(x) = \bigoplus_{I \subseteq \{1,\ldots,n\}} a_I \left(\prod_{i \in I} x_i \right)$$

- Exists and unique
- The ANF degree is affine invariant
- Evaluation requires $O(n2^n)$ operations

Walsh-Hadamard Transform

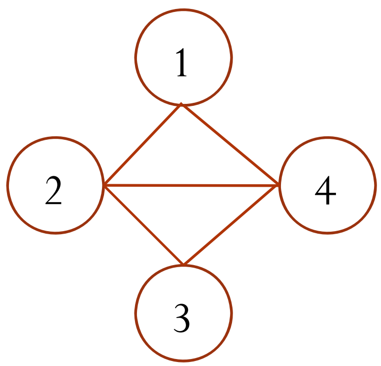
- $$F(w) = \sum_{x \in F_2^n} (-1)^{f(x) + w \cdot x} \quad \text{where} \quad w \cdot x = w_1 x_1 + \cdots + w_n x_n$$
- Almost all cryptographic properties can be expressed in terms of the WHT
- Can be evaluated in $O(n2^n)$ operations

- What is the best representation?
  - TT, WHT, or ANF
  - Example:
    - ANFD
    - $w_H(f) = \#\{x \in F_2^n \mid f(x) \neq 0\}$

Graph Representation: Quadratic functions

- Boolean functions with *only* quadratic terms
  - can be represented by an undirected graph with n nodes
  - An edge between node i and j exists iff $a_{ij} = 1$ in the ANF of f
- Boolean functions corresponding to isomorphic graphs belong to the same affine class

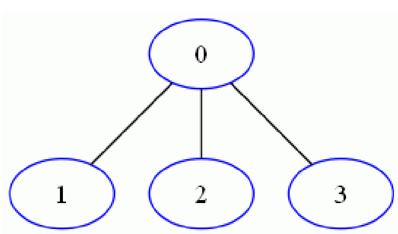
- Example $f(x_1, x_2, x_3, x_4) = x_1x_2 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4$

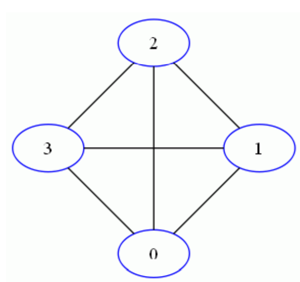


Definitions

Connected Graphs

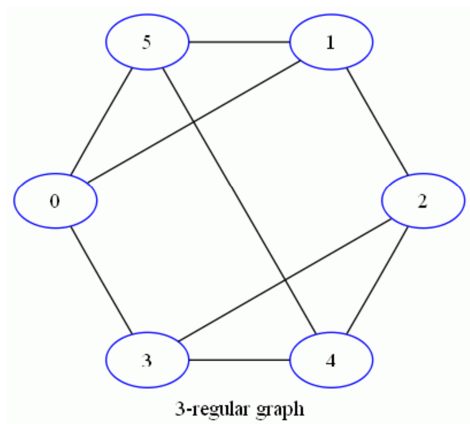
- A graph in which any two vertices are connected by a path is called a connected graph.





Regular Graphs

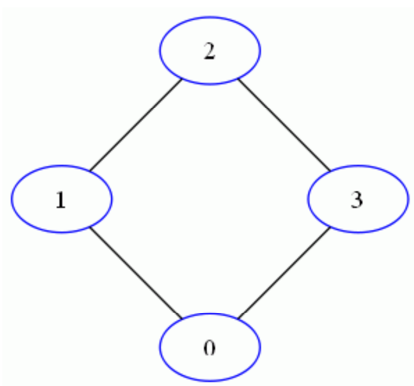
- A graph in which every vertex has the same degree is called a regular graph



Strongly Regular Graph

A graph G is strongly regular if there exist nonnegative integers e and d such that, for all vertices μ, ν, the number of vertices adjacent to both μ and ν, $\delta(\mu,\nu)$ is given by

$$\delta(\mu,\nu)=\begin{cases} e, \text{if } \mu \text{ and } \nu \text{ are adjacent} \\ d, \qquad \text{otherwise} \end{cases}$$



node 0 and 1 are adjacent and have 0 common neighbours $\Rightarrow e=0$

node 0 and 2 are not adjacent and have 2 common neighbours $\Rightarrow d=2$

Graph Spectrum

- Given a graph G and its adjacency matrix A, the spectrum of G is the set of the eigenvalues of A, which are also called eigenvalues of G.
- Isomorphic graphs have the same spectrum

Graph Representation: General case

- A general Boolean function can be associated with a Cayley graph

$$V_f = F_2^n$$

$$E_f = \{(w,u) \in F_2^n \times F_2^n \mid f(w \oplus u) = 1\}$$

- There is a 1-1 relationship between the graph eigenvalues and the Walsh coefficients: $\lambda_i = 2^n F(i)$

Example:

- Truth Table:

  f(x) = [0 0 1 1]

- Walsh Transform: $F(\omega) = 2^{-n} \sum_x f(x)(-1)^{\omega \cdot x}$

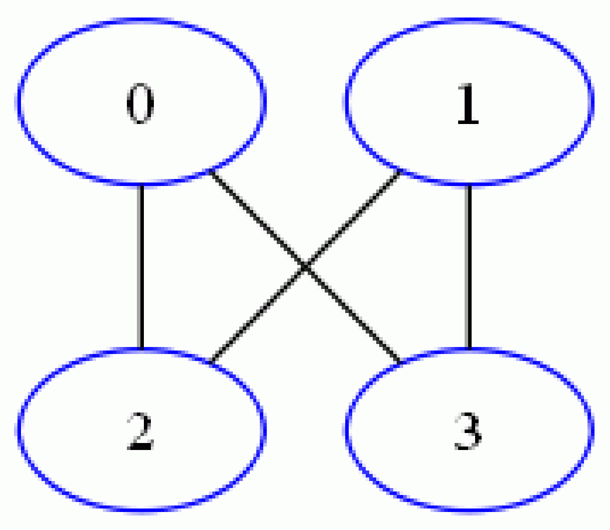
  F(ω) = [2 0 -2 0]

- Adjacency Matrix:

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

- Eigenvalues:

  λ = [-2 0 0 2]



Associated Cayley Graph

Example

- Truth Table:

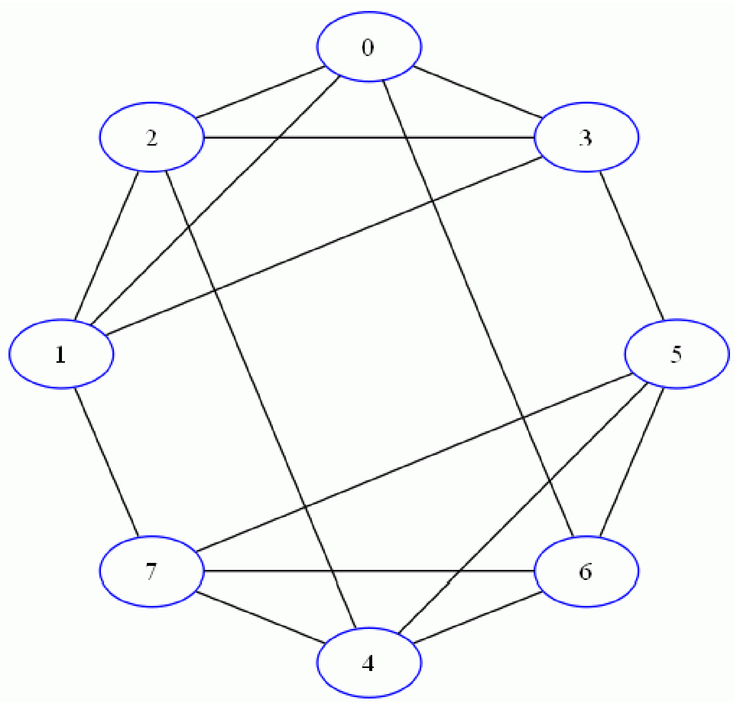
  f(x) = [0 1 1 1 0 0 1 0]

- Walsh Transform:

  F(ω) = [4 0 -2 -2 2 -2 0 0]

- Adjacency Matrix:

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

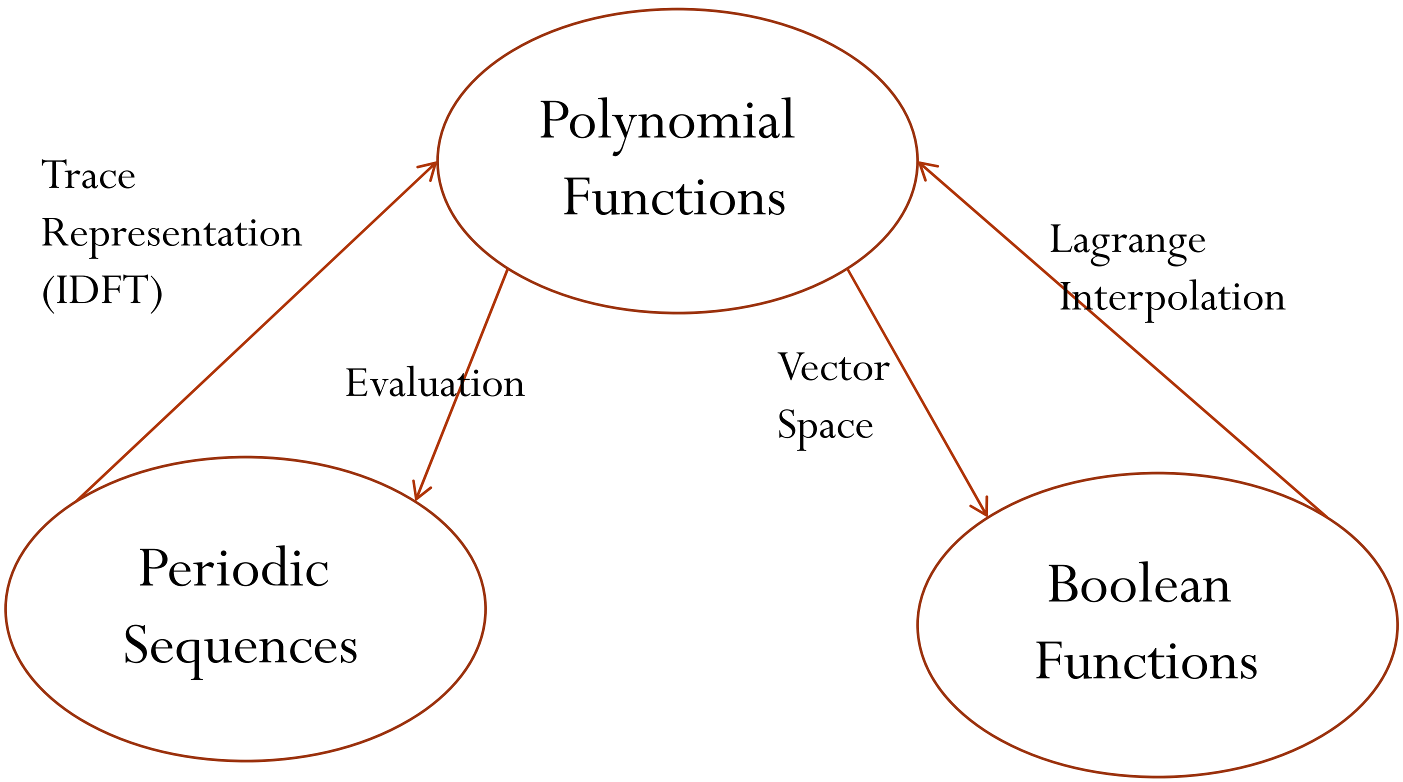


Associated Cayley Graph

- Eigenvalues:

  λ = [-2 -2 -2 0 0 0 2 4]

1-1 Correspondences with Polynomial Functions and Periodic Sequences



Example

Truth Table

x	0	1	2	3	4	5	5	7	8	9	10	11	12	13	14	15
S(x)	0	1	8	15	12	10	1	1	10	15	15	12	8	10	8	12

Interpolation

Evaluation

Corresponding Polynomial Function

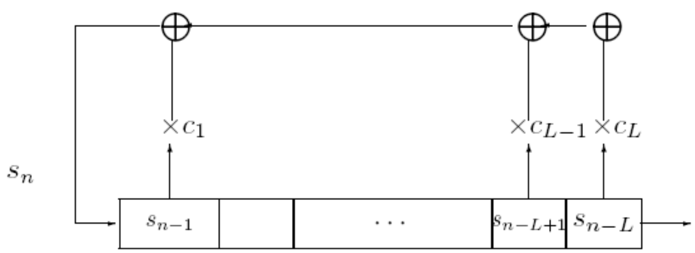
$$GF(2^4) \text{ defined by } f(x) = x^4 + x + 1 \Rightarrow s(x) = x^3$$

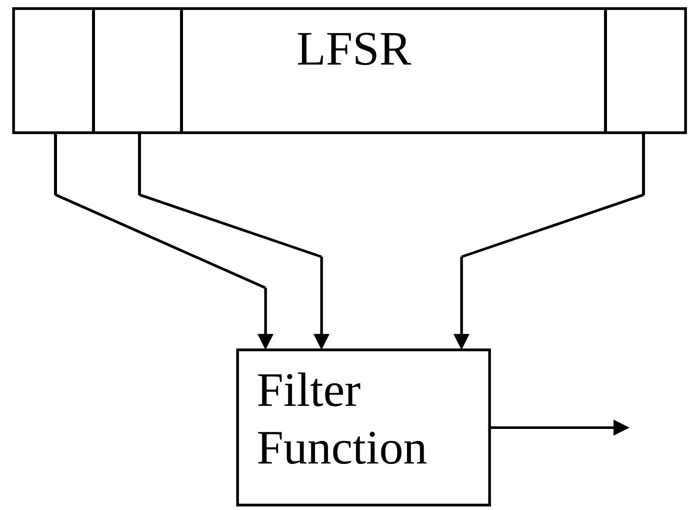
$$GF(2^4) \text{ defined by } f(x) = x^4 + x^3 + 1 \Rightarrow$$
$$s(x) = x + x^2 + 7x^3 + 15x^4 + 5x^5 + 14x^6 + 14x^8 + 2x^9 + 7x^{10} + 9x^{12}$$

Cryptographic properties of Boolean functions

- Balance
- Correlation immunity
- Resiliency
- Nonlinearity
- Algebraic normal form degree
- Algebraic immunity degree

ANFD





Using Berlekamp Massey algorithm,
the initial value and the connection Polynomial of the LFSR can be deduced using 2L consecutive bits

$$f(x) = \bigoplus_{I \subseteq \{1,\ldots,n\}} a_I \left(\prod_{i \in I} x_i \right)$$

Output will have an equivalent length

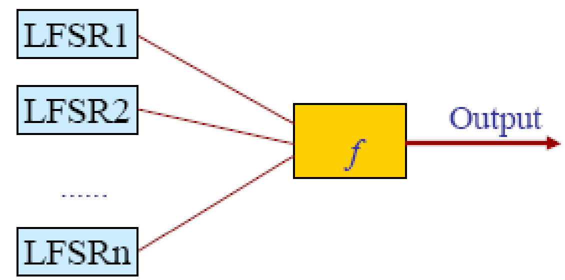
$$L = \sum_{I \subseteq \{1,\ldots,n\}} a_I \left(\prod_{i \in I} L_i \right)$$

C1. The ANFD, d, should be as high as possible

Resiliency

- Combining functions must be balanced
- If f remains balanced if we fixed up to m of its input coordinates, then f is called m-resilient
- In terms of WHT

$$F(w) = 0$$
$$\text{for all } w \in F_2^n \text{ such that } w_H(w) \le m$$



C2. The resiliency degree m should be as high as possible

- Siegenthaler bound (c1 & c2) :

$$m + d \le n,$$
$$m + d \le n - 1 \text{ for balanced functions}$$

Nonlinearity

- The nonlinearity of f is the minimum hamming distance between f and the set of Affine functions
- In terms of WT

$$NL_f = 2^{n-1} - \frac{1}{2}\max_{w \in F_2^n} |F(w)|$$

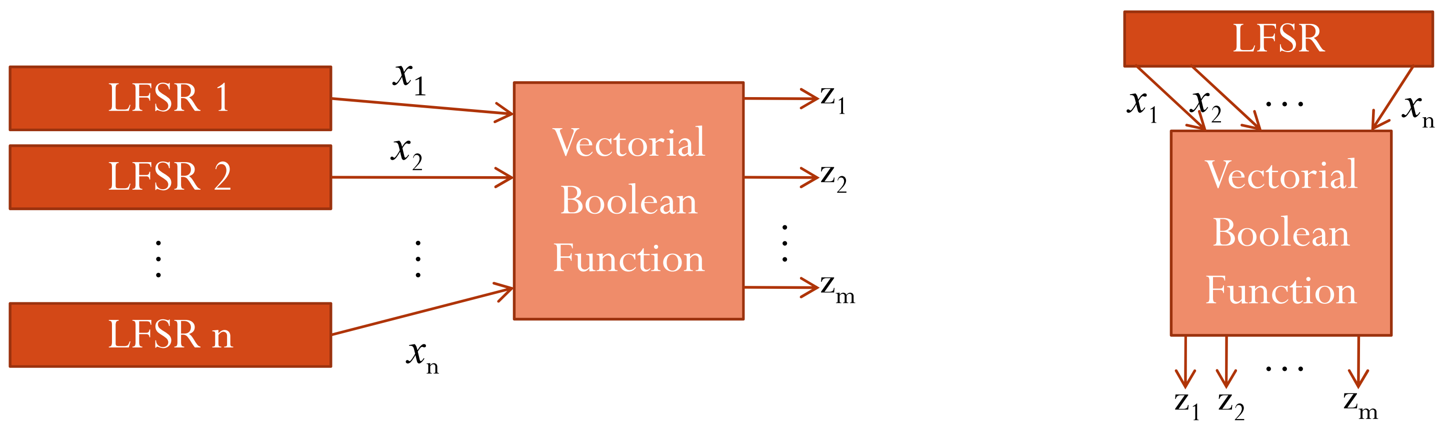
C3. NL should be as high as possible

- Sarkar-Maita Bound (C2 & C3): $NL \leq 2^{n-1} - 2^{m+1}$

Bent functions

- Bent functions are functions that
  - have flat WHT spectrum
  - achieve the maximum possible nonlinearity
- Let f be a bent function and G its associated graph. Then, G is strongly regular graph and has the additional property e=d.
- Different generalizations
  - Carlet Hyper-bent functions
  - Youssf and Gong Hyper-bent functions

Correlation Attack of Vectorial Stream Ciphers

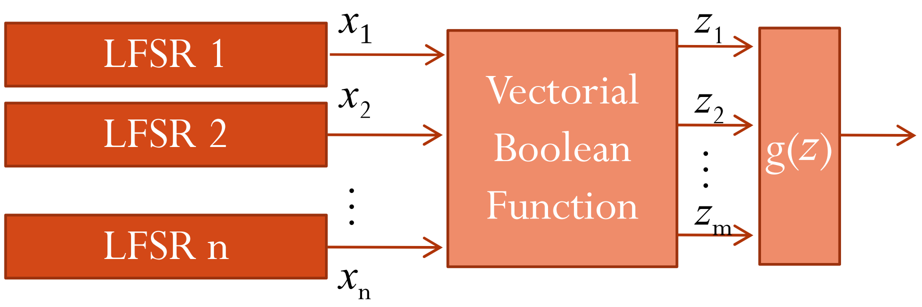


$$\Pr(b \cdot z = w \cdot x) = \Pr(b_1 z_1 \oplus \ldots \oplus b_m z_m = w_1 x_1 \oplus \ldots \oplus w_n x_n).$$

- For correlation attack to succeed, we require $Bias = |\Pr(b \cdot z = w \cdot x) - \frac{1}{2}|$ to be high where $z=f(x)$ is the output. i.e. probability is far away from ½.
- Thus the *nonlinearity*:

$$N_f = 2^{n-1} - \frac{1}{2} \max_{w \neq 0, b} \left| \sum_{x \in F_2^n} (-1)^{b \cdot f(x) \oplus w \cdot x} \right| \text{ should be as high as possible}$$

Unrestricted Nonlinearity



- Since z is known, the attacker can consider

$$\Pr(g(z) = w_1 x_1 \oplus \ldots \oplus w_n x_n) = \Pr(g(z) = w \cdot x).$$

which is linear in x for any Boolean function g(·).

- For the attack to succeed, we require

$$Bias = |\Pr(g(z) = w \cdot x) - \tfrac{1}{2}| \text{ to be high}$$

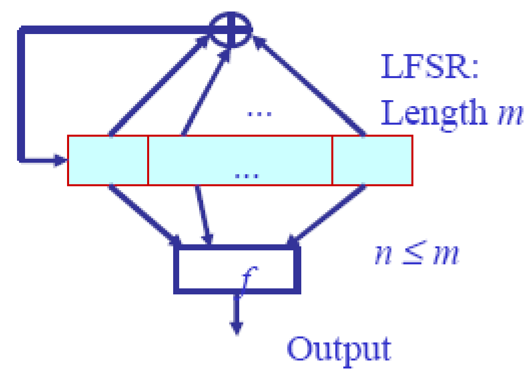
- Thus, the *unrestricted nonlinearity*

$$UN_f = 2^{n-1} - \frac{1}{2} \max_{w \neq 0, g(\cdot)} \left| \sum_{x \in F_2^n} (-1)^{g(f(x)) \oplus w \cdot x} \right| \text{ should be as high as possible}$$

Algebraic Attacks

- Initial state $s = (s_0, s_1, \cdots s_{n-1})$
- The output stream is given by

$$
\begin{aligned}
o_0 &= f(s_0, s_1, \cdots, s_{n-1}), \\
o_1 &= f(L(s_0, s_1, \cdots, s_{n-1})), \\
&\vdots \\
o_k &= f(L^k(s_0, s_1, \cdots, s_{n-1}))
\end{aligned}
$$



- Algebraic attacks try to efficiently recover s from the output sequence O

Algebraic Attacks

- In general, solving the system of multivariate equations is NP complete (even if all the equations are quadratic)
  - Linearization
  - Gröbner Basis
- If f has ANFD d, then $f(L^k(s_0, s_1, \cdots, s_{n-1}))$ would roughly have $\binom{n}{d}$ monomials
- Using a simple Linearization approach, S can be recovered by solving a system with $\binom{n}{d}$ variables; complexity $\approx \binom{n}{d}^3$

Linearization Example

System of nonlinear equations:

$$\begin{aligned} x \oplus y \oplus xy &= 1 \\ y \oplus xy &= 1 \end{aligned}$$

New Variables: $M_1 := x$, $M_2 := y$ and $M_3 := xy$
New system of linear equations:

$$\begin{aligned} M_1 \oplus M_2 \oplus M_3 &= 1 \\ M_2 \oplus M_3 &= 1 \end{aligned}$$

Applying Gauss reveals:

$$\begin{aligned} M_1 &= 0 \\ M_2 \oplus M_3 &= 1 \end{aligned}$$

$\Rightarrow$ Two solutions:

$$M_1 = 0,\ M_2 = 0,\ M_3 = 1$$

$$M_1 = 0,\ M_2 = 1,\ M_3 = 0$$

$$\begin{aligned} 0 &= M_1 = x \\ 0 &= M_2 = y \\ 1 &= M_3 = xy \end{aligned}$$

Solution doesn't make sense!

$$\begin{aligned} 0 &= M_1 = x \\ 1 &= M_2 = y \\ 0 &= M_3 = xy \end{aligned}$$

Solution correct!

Algebraic Attacks

- If one can find a (non zero) function g of degree $d_g < d_f$ such that

$$g * f = 0 \text{ or } g * (1 + f) = 0$$

then the number of unknowns can be reduced to $\binom{n}{d_g} < \binom{n}{d_f}$

- eXtended Linearization (XL algorithm)

Algebraic Immunity

- AI(g) is the lowest degree of any non zero g such that

$$g * f = 0 \text{ or } g*(1+f) = 0$$

- Some argues that it should be called annihilator immunity
- $AI(f) \le \lceil \frac{n}{2} \rceil$
- For even n, AI is almost always $\approx \frac{n}{2}$
- For odd n, AI is almost always $\approx \frac{n-1}{2}$
- AI implies a lower bound on nonlinearity $NL \ge 2\sum_{i=0}^{AI-2} \binom{n-1}{i}$

Complexity of finding AI

- Compute the annihilator space of degree ≤ d
- Number of coefficients in g $k = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d}$
- $\forall x$ such that $f(x) = 1 \rightarrow$ linear equation $g(x) = 0$
- Number of equations: $w_H(f)$
- Gaussian elimination requires $O(2^{2^n} k)$
- Meier, Pasalic and Carlet: $O(k^3)$
- State of the art (Armknecht et. al): $O(k^2)$

Examples for well known constructions

- Maiorana-McFarland's (MM) constructions (concatenation of affine functions)

$$f(x,y) = x \cdot \phi(y) + g(y),$$

$$\text{where } \phi : F_2^{n/2} \to F_2^{n/2}, g : F_2^{n/2} \to F_2$$

$$f \text{ is bent iff } \phi \text{ is a permutation}$$

- Similar constructions for resilient functions

$$f(x,y) = x \cdot \phi(y) + g(y),$$

$$\text{where } n = r + s, \phi : F_2^{n/2} \to F_2^{r}, g : F_2^{s} \to F_2,$$

$$w(\phi(y)) > k \Rightarrow f \text{ is } m \geq k \text{ resilien with}$$

$$2^{\text{n-1}} - 2^{r-1} A \leq NL \leq 2^{\text{n-1}} - 2^{r-1} \left\lceil \sqrt{A} \right\rceil,$$

$$where\ A = \max_{a \in F_2^r} \# \phi^{-1}(a)$$

Other Algebraic constructions

- Power functions x^d over $GF(2^n)$

n	d	weight	degree	nonlinearity	alg. immunity
8	31	128	5	112	4
8	39 (Kasami)	128*	6	114	4
9	57 (Kasami)	256	4	224	4
9	59	256	5	240	5
9	115	256	5	240	5
10	241 (Kasami)	512	5	480	5
10	362	512	5	480	5
10	31 (Dillon)	512*	9	486	5
10	339 (Dobbertin)	512*	9	480	5
11	315	1024	6	992	6
12	993 (Kasami)	2048*	11	2000	6
12	63 (Dillon)	2048*	11	2000	6
12	636	2048*	11	2000	6
13	993 (Kasami)	4096	6	4032	6
13	939	4096*	12	4030	7
14	4033 (Kasami)	8192	7	8064	7
14	127 (Dillon)	8192*	13	8088	7

n	d	weight	degree	nonlinearity	alg. immunity
6	-1	32	5	24	3
7	-1	64	6	54	4
8	-1	128	7	112	4
9	-1	256	8	234	4
10	-1	512	9	480	5
11	-1	1024	10	980	5
12	-1	2048	11	1984	5
13	-1	4096	12	4006	6
14	-1	8192	13	8064	6

Heuristic optimization based constructions

- Previous algebraic approaches may not always allow the system designer to achieve optimal constructions
- Exhaustive search is not an option for $n > 8$
- Cryptographically rich Boolean function classes
  - Limited search space but rich in cryptographically good functions
- Spectral Inversion
  - Possible cost functions $\sum_{s \in Z_n 2^n} | \sum_{w \in Z_2^n} F(w)F(w \oplus s))|.$

Cryptographically rich classes

- Symmetric functions (too restrictive)

$$f(x_1,\ldots,x_n) = f(x_{\sigma(1)},\ldots,x_{\sigma(n)}) \text{ for all permutations } \sigma$$

- Rotation symmetric functions

$$f(\rho^k(x_1,\ldots,x_n)) = f(x_1,\ldots,x_n) \text{ for all cyclic shifts } \rho^k$$

- Dihedral Symmetric Boolean
  - Functions invariant under the action of Dihedral group D_n
  - In addition to the cyclic shift, D_n includes a reflection operator

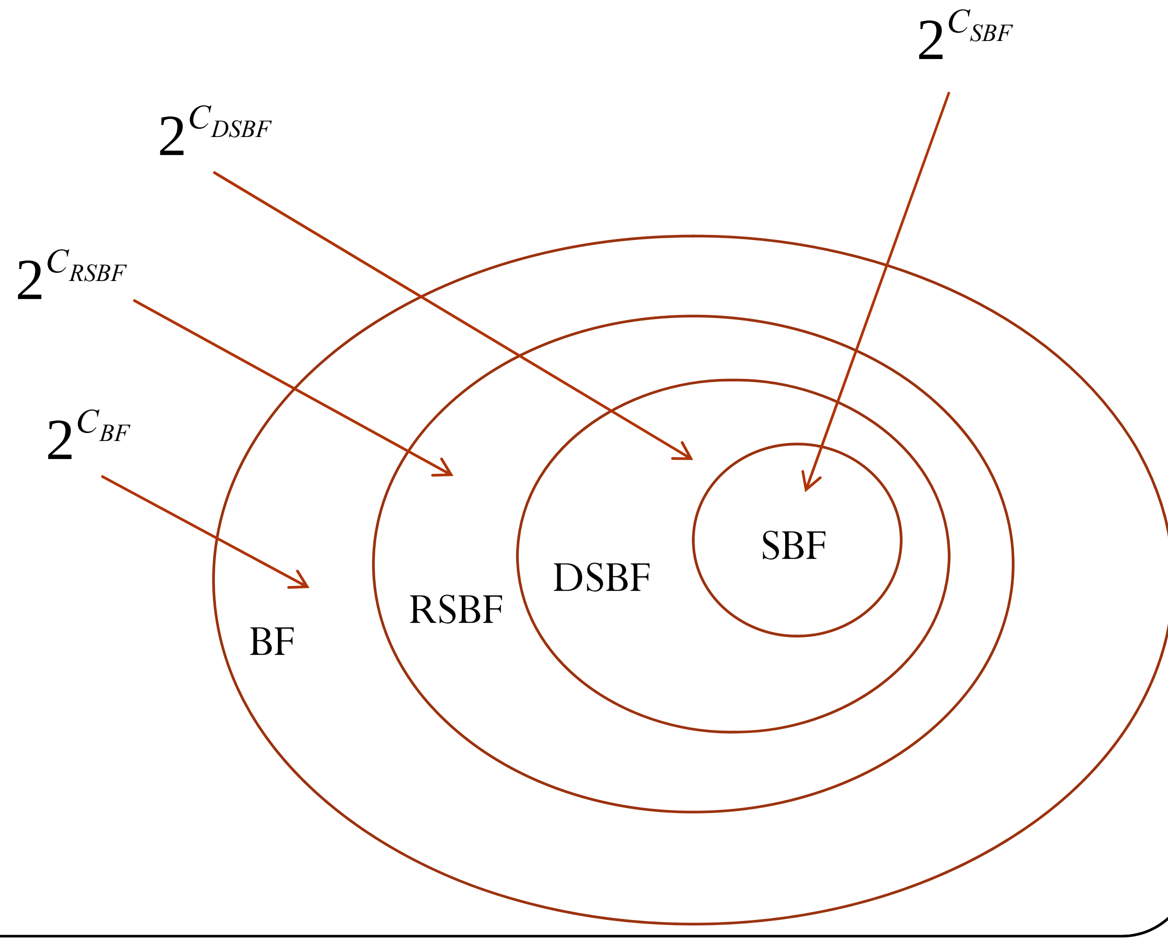
$$\tau_n(x_1,x_2,\ldots,x_n) = (x_n,\ldots,x_2,x_1)$$

n	3	4	5	6	7	8	9	10
C_{BF}	8	16	32	64	128	256	512	1024
C_{RSBF}	4	6	8	14	20	36	60	108
C_{DSBF}	4	6	8	13	18	30	46	78
C_{SBF}	4	5	6	7	8	9	10	11

$$C_{BF} = 2^{n}$$

$$C_{RSBF} = \frac{1}{n}\sum_{k|n}\phi(k)2^{\frac{n}{k}}$$

$$C_{SBF} = n+1$$



Solving two open problems

- Let *(n,m,d,nl)* denote
  - n-variable
  - m-resilient
  - ANF degree, d
  - Nonlinearity nl
- The existence of *(9,3,5,240)* and *(10,2,7,488)* has been an open problem.
- Using a heuristic search, we are able to construct several examples for such resilient functions.

Construction of a (9,3,5,240) function

- Consideration of the Search Space
  - BF search space is too large (2^{512})
  - RSBF space is moderate (2^{60}) but it was proved that no such RSBF function exists
  - Spectral inversion: $res(f) = m \Rightarrow |F(\omega)| = 0 \bmod 2^{m+2}$
    - The spectrum of any $(n, m, -, 2^{n-1} - 2^{m+1})$ function is necessarily a three-valued function (Plateaued) $(0, \pm 2^{m+2})$, $m > \left\lfloor \frac{n}{2} - 2 \right\rfloor$
    - Direct spectral inversion

$$|F(\omega)| = \begin{cases} 0, & \text{if } wt(\omega) \le 3, \\ 0 \text{ or } 32, & \text{if } wt(\omega) > 3 \end{cases}$$

did not prove to be useful

(9,3,5,240)

- Concatenation idea

Let $f : \mathbf{F}_2^{n+2} \to \mathbf{F}_2$ and $f_1, f_2, f_3, f_4 : \mathbf{F}_2^n \to \mathbf{F}_2$.

$$f = [f_1 \mid f_2 \mid f_3 \mid f_4]$$

From the Hadamard matrix $H_0 = 1, H_n = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \oplus H_{n-1}, H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$

The Walsh transform $F(w)$ of f is given by

$$F = [F_1 + F_2 + F_3 + F_4 \mid F_1 - F_2 + F_3 - F_4 \mid F_1 + F_2 - F_3 - F_4 \mid F_1 - F_2 - F_3 + F_4]$$

(9,3,5,240)

- It is possible to construct an $\left(n, m, n-m-1, 2^{n-1}-2^{m+1}\right)$ function where $m > \left\lfloor \frac{n}{2} - 2 \right\rfloor$ from the concatenation of four $\left(n-2, m, n-m-3, 2^{n-3}-2^{m+1}\right)$ functions with nonoverlapping Walsh coefficients, if such four functions exist.

- Thus, the search for $(9,3,5,240)$ functions is reduced to finding four $(7,3,3,48)$ functions with nonoverlapping spectrum coefficients. This helps us in reducing the search space dramatically compared to the direct search for $(9,3,5,240)$ functions

- The algebraic degree of such functions is always maximum (n-m-1)

- Several examples were obtained using PSO optimization

Construction of a (10,2,7,488) function

- We can't specify the distribution of the Walsh spectrum for f.

- We only know that the Walsh spectrum of $(10;2;7;488)$ Boolean function satisfy the following constraints:

$$\left|F(\omega)\right| = \begin{cases} 0, & \text{if } wt(\omega) \le 2, \\ 0,16,32 \text{ or } 48, & \text{if } wt(\omega) > 2 \end{cases}$$

But we can't determine their distribution.

(10,2,7,488)

- Direct construction is ineffective because of the super-exponential increase in the search space which grows as $2^{2^n} = 2^{1024}$.
- Even if the search space is constrained to the set of RSBFs, the search space is still relatively large (2^{108}).

(10,2,7,488) – Back to concatenation

- Our main observation is that the search space can be reduced dramatically by noting that a $(10,2,7,488)$ function f may be constructed by concatenating $f_1 : Z_2^{n-1} \to Z_2$ and $f_2 : Z_2^{n-1} \to Z_2$ that satisfy the following constraints:

$$|F_i(\omega)| = \begin{cases} 0, & \text{if } wt(\omega) \le 1, \\ \le 24, & \text{if } wt(\omega) = 2, \\ \le 48, & \text{if } wt(\omega) > 2 \end{cases}$$

$$i = 1,2 \; .$$

(10,2,7,488) – our search procedure

- Obtain a 9-bit RSBF f_1 that satisfies the above constraints using the following cost function.

$$\cos t_1(f_1) = \sum_{\omega | wt(\omega) \le 1} |F_1(\omega)|^2 + \sum_{\substack{\omega | wt(\omega)=2, \\ |F_1(\omega)| \notin \{8,16,24\}}} |F_1(\omega)|^2 + \max_{F_1(\omega)} |F_1(\omega) - 32|^2$$

where $\omega \in Z_2^9$.

- Once f_1 is found, Obtain a 9-bit RSBF f_1 that minimizes the following cost function.

$$\cos t_2(f_2) = \sum_{\omega | wt(\omega) \le 1} |F_2(\omega)|^2 + \sum_{\omega | wt(\omega)=2,} |F_1(\omega) + F_2(\omega)|^2 + \max_{F_2(\omega)} |F_2(\omega) - 32|^2$$

where $\omega \in Z_2^9$.

- Test if $f = [f_1 \mid f_2]$ is a function, if the search for f_2 under certain f_1 failed after certain number, go to step 1 and find another f_1.

Conclusion and open problems

- There is no such thing as a secure Boolean function.
  - There may be functions that are appropriate to be used in particular contexts to give secure system.
- Almost every Boolean function paper has a list of open problems
  - Some are very specific
    - e.g., find (8,0,7, 118)
- More work is needed
  - at the interface between symmetric algorithms and Boolean function layers
  - constructions of Boolean functions with implementation constraints