A survey on construction of Boolean function with optimum algebraic immunity (AI)

Chunlei Li

Selmer Center, Department of Informatics
University of Bergen

November 10, 2011

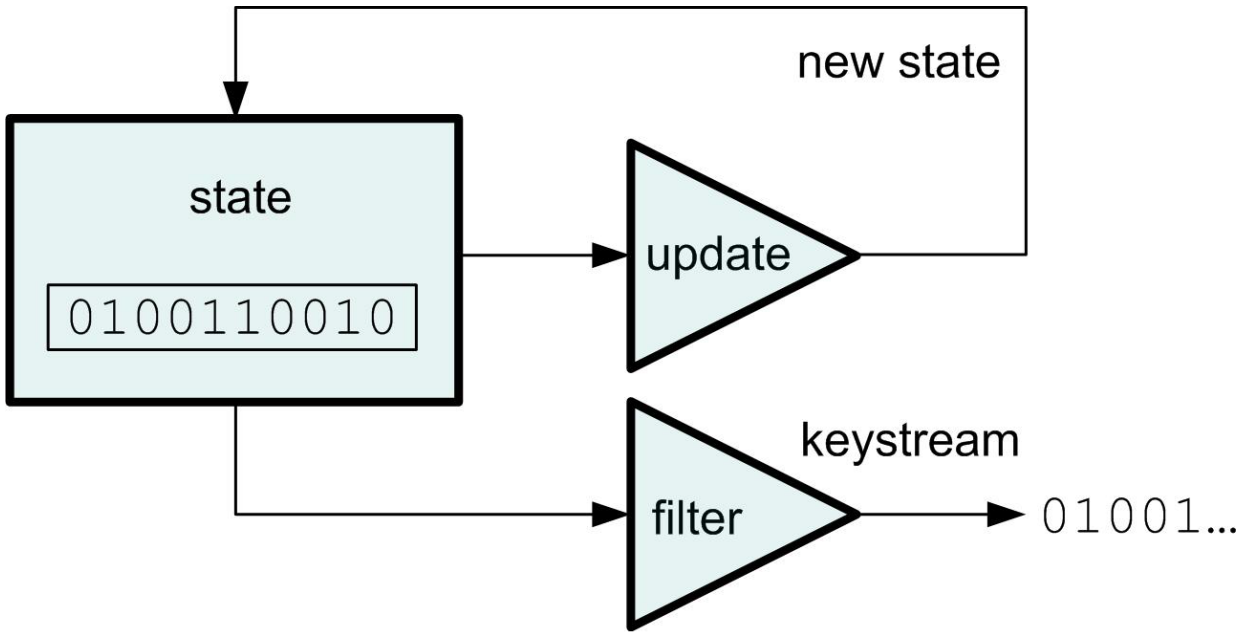
Outline

1. Background on Boolean function (BF)
   - The applications of BF in LFSR-based stream ciphers
   - Cryptographic criteria for BF
2. Algebraic attacks on stream ciphers and algebraic immunity (AI)
   - Algebraic attacks
   - Algebraic immunity
3. A survey for theoretical constructions of BF with optimal AI
   - Constructions over vector space
   - Constructions over finite fields

Outline

Stream Ciphers

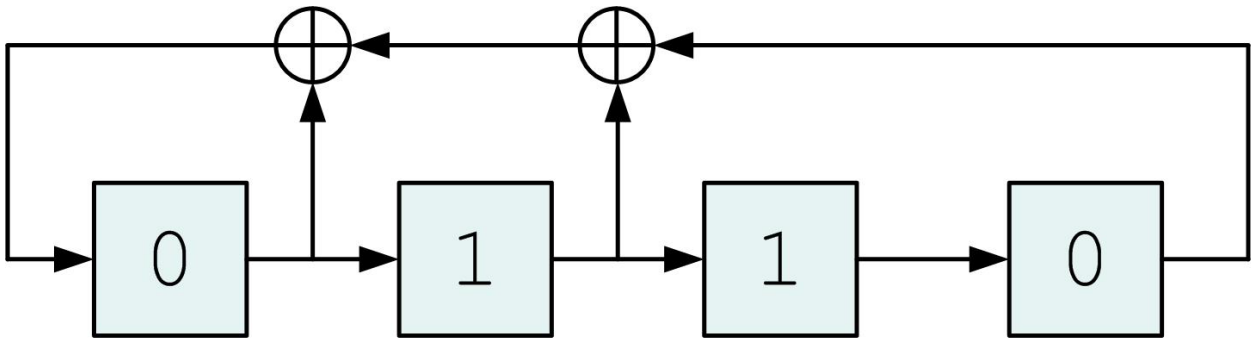
Stream cipher has 3 elements: state/ update function/ filter
State initialised with secret key K, keystream XORed with plaintext.



Only few operations: Fast and very low HW requirements.

Linear Feedback Shift Register (LFSR)

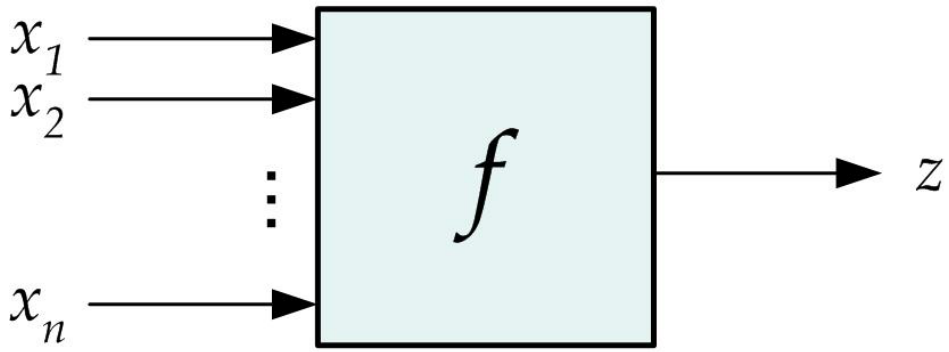
- State has l bits
- Linear update: new state=L(state)



Advantages: Very efficient, good statistical properties
Limitations: Easily predicted, requires good filter...

Boolean Function

Boolean function f with n input variables



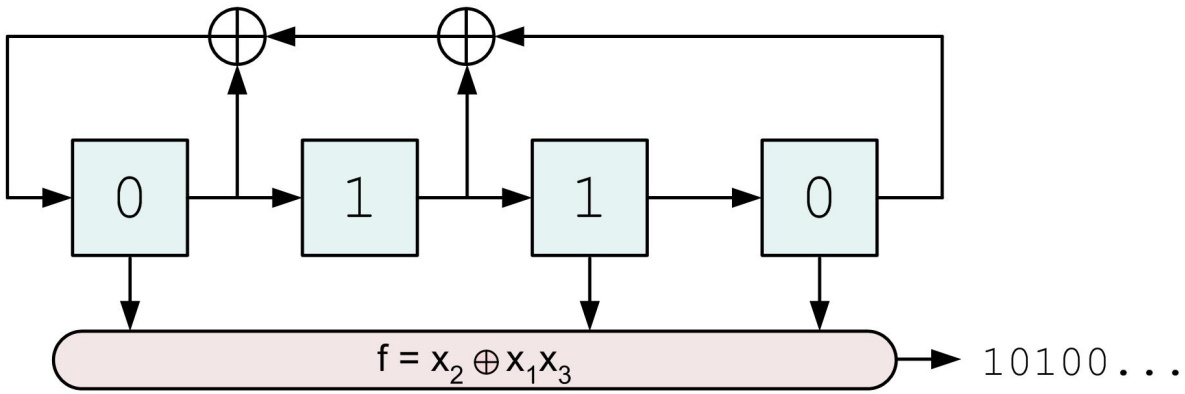
Computation in finite field: $x \oplus x = 0$ and $x^2 = x$.

A toy example: $f(x_1, x_2, x_3) = x_1 \oplus x_2 \oplus x_1x_3$

$x_1x_2x_3$	$f(x)$	$x_1x_2x_3$	$f(x)$
000	0	100	1
001	0	101	0
010	1	110	0
011	1	111	1

Filter Generator

Combination of LFSR and Boolean function $\rightarrow$ filter generator:



- State of l bits, initialized with secret key K (Example: $l = 128$ bits)
- Filter takes n bits of state (Example: $n = 20$ bits)
- Output of keystream bit $z^t = f(L^t(K))$

Related designs: Combiner, clock-controlled generators,...

Outline

Basics for Boolean functions (BF)

The Algebraic Normal Form (ANF) for $f : F_2^n \to F_2$:

$$f(x_1, \cdots, x_n) = \sum_{I \subseteq \{1,\ldots,n\}} a_I \left(\prod_{i \in I} x_i \right), \ a_I \in F_2.$$

The algebraic degree: $\deg(f)$ is the degree of the ANF.
Affine functions: degree ≤ 1 with the form

$$a(x) = a_1 x_1 + \cdots + a_n x_n + a_0.$$

The **support set** of a function: $supp(f) = \{x \in {F_2}^n \mid f(x) = 1\}$.
The **Hamming weight** of a function f: $wt(f) = \#supp(f)$, f is **balanced** if $wt(f) = 2^{n-1}$.

Basics for Boolean functions (BF)

The **Hamming distance** between two functions:

$$d_H(f,g) = wt(f+g) = \#\{x \in F_2^n \mid f(x) \neq g(x)\}.$$

The **nonlinearity** of f is the minimum Hamming distance to affine functions.

$$nl(f) = \min\{d(f,l) : \deg(l(x)) \leq 1\} = 2^{n-1} - \frac{1}{2} \max_{a \in {F_2}^n} |\widehat{\mathrm{f}}(a)|,$$

where the **Discrete Fourier Transform**:

$$\widehat{\mathrm{f}}(a) = \sum_{x \in {F_2}^n} (-1)^{f(x)+a \cdot x}.$$

The nonlinearity is upper bounded by $2^{n-1} - 2^{n/2-1}$ and *must be close to this maximum* to prevent the system from linear and correlation attacks.

Important criteria for BF design

To resist against known attacks, an n-variable BF used in stream ciphers should:

- be balanced
- have high algebraic degree close to n
- have large nonlinearity close to $2^{n-1} - 2^{n/2-1}$

Outline

Algebraic Equations

Any stream cipher is defined by a system of algebraic equations.
A solution of this system gives the secret key.

Our system of equations:

$$\begin{aligned} f(K) &= z^0 \\ f(L(K)) &= z^1 \\ f(L^2(K)) &= z^2 \\ &\vdots \end{aligned}$$

Example:

$$\begin{aligned} k_1 \oplus k_2 \oplus k_1 k_2 &= 1 \\ k_3 \oplus k_2 k_3 &= 0 \\ k_2 \oplus k_1 k_2 \oplus k_2 k_3 &= 0 \\ &\vdots \end{aligned}$$

Properties: l unknowns, many equations, nonlinear (degree k)
Solution: How to solve this system?

Linearization

Linearization: Introduce new variables for each monomials.

Example

1. Equation: $k_1 \oplus k_2 \oplus k_1 k_2 = 1$
2. New variables: $v_1 = k_1$, $v_2 = k_2$, $v_3 = k_1 k_2$
3. New equation: $v_1 \oplus v_2 \oplus v_3 = 1$

$$\#monomials \leq \binom{l}{0} + \binom{l}{1} + \cdots + \binom{l}{k} \approx \binom{l}{k}$$

Complexity for solving this linear system with $\binom{l}{k}$ variables:

- Data complexity is $\binom{l}{k}$ keystream bits — l : size of state
- Time complexity is $\binom{l}{k}^3$ for Gaussian elimination — k : degree of f

Algebraic Attacks

Problem: Simple linearization is not efficient
Idea: Reduce degree of equations $\rightarrow$ algebraic attacks
How?

1. find g of degree $d \leq k$, such that $f \cdot g = 0$, or
2. find g of degree $d \leq k$, such that $(f+1) \cdot g = 0$

Example 1:

$$\left.\begin{aligned} f(x) &= x_1x_2x_3x_4x_5 \\ g(x) &= x_1 + x_4 \end{aligned}\right\} \Rightarrow fg = 0$$

Example 2:

$$\left.\begin{aligned} f(x) &= x_1x_2x_3x_4 + 1 \\ g(x) &= x_1 + x_2 \end{aligned}\right\} \Rightarrow fg = g$$

And then? New equation of degree d from $z^t = f(L^t(K))$:

1. if $z^t = 1$ and $f \cdot g = 0$, then $g(L^t(k)) = 0$, or
2. if $z^t = 0$ and $(f+1) \cdot g = 0$, then $g(L^t(k)) = 0$

Outline

New criterion: Algebraic Immunity

Algebraic Immunity (AI)

Given f, find g of minimum degree d, such that $f \cdot g = 0$ or $(f+1) \cdot g = 0$. The algebraic immunity of f is d.

System of equations of degree d.

Complexity for linearization

- Data complexity is $\binom{l}{d}$
- Time complexity is $\binom{l}{d}^3$

l : size of state

$d : AI(f)$

Computation of AI

$AI(f)$ must be large to resist against Algebraic Attacks.

Upper bound?

- we have $AI(f) \leq \deg(f)$ since $f+1$ itself is an annihilator of f
- Courtois-Meier (2003): $AI(f) \leq \lceil \frac{n}{2} \rceil$.

In practical situation, $AI(f)$ *must be greater than or equal to 7*.
The best known algorithm for computing AI of f can be efficient only when $n \leq 20$. Besides, randomly generating Boolean functions with high algebraic immunity in 20 variables is too slow.

We need theoretical construction for BF with high AI.

More requirements?...

Remark: *a high value of $AI(f)$ is not sufficient*:

- If g of small degree and h, of degree not too large, exist such that $fg = h$, then a fast algebraic attack is possible.
- Note that if $fg = h$ and $h \neq 0$, then the degree of h is at least $AI(f)$. Hence, a high AI is a necessary condition for a resistance to fast algebraic attacks as well.
- A very efficient algebraic-like attack on the filter generator was found by Sondre Rønjom and Tor Helleseth. Its time complexity is roughly $O(\mathcal{D})$, where $\mathcal{D} = \sum_{i=0}^{\deg(f)} \binom{l}{i}$ and its data complexity is also $O(\mathcal{D})$. So $\deg(f)$ must be close to n.

Security requirements

A cryptographically secure Boolean function used in stream ciphers should

- be balanced
- have high algebraic degree *close to* n
- have large nonlinearity *close to* $2^{n-1} - 2^{n/2-1}$
- have high algebraic immunity *close to* $\lceil n/2 \rceil$
- have good resistance against the fast algebraic attack

Not a easy work!

Outline

Generalized Majority Function (Dalai-Maitra-Sarkar, 2005)

The majority function (Braeken et al.) is the symmetric function defined by:

$$f(x) = 1 \text{ iff. } wt(x) \leq n/2.$$

Generalized construction: for any subset $T \subset \{x : wt(x) = n/2\}$,

$$f(x) = 1 \text{ iff. } wt(x) > n/2 \text{ or } x \in T.$$

Properties:

1. $AI(f)$ is optimal
2. balancedness depends on T
3. weak against fast algebraic attacks
4. nonlinearity at most $2^{n-1} - \binom{n-1}{\lfloor n/2 \rfloor}$.

   Note that $\binom{n-1}{\lfloor n/2 \rfloor} \approx \sqrt{\frac{2}{\pi n}} 2^{n-1} \gg 2^{n/2-1}$.

Iterative Construction (Dalai-Gupta-Maitra, 2005)

A $2k+2$-variable function is concatenated by four $2k$-variables functions:

$$\phi_{2k+2} = \phi_{2k}||\phi_{2k}||\phi_{2k}||\phi_{2k}^1,$$

where $\phi_{2j}^i\ (i > 0)$ is defined by:

$$\phi_{2j}^i = \phi_{2j-2}^{i-1}||\phi_{2j-2}^i||\phi_{2j-2}^i||\phi_{2j-2}^{i+1},$$

with base step $\phi_{2j}^0 = \phi_{2j}$ for $j \geq 0, \phi_0^i = \phi_0^{i \text{ [mod 2]}}$ for $i \geq 0$ and ϕ_0^1 is the complement function of ϕ_0.

A toy example: $\phi_0 = x_1x_2 : 0001$, $n = 2k = 6$

Step 1: $\phi_0 = 0001, \phi_0^1 = 1110$

Step 2: $\phi_2 = \phi_0||\phi_0||\phi_0||\phi_0^1 = 0001||0001||0001||1110,$
$\phi_2^1 = \phi_0^0||\phi_0^1||\phi_0^1||\phi_0^2 = 0001||1110||1110||0001$

Step 3: $\phi_4 = \phi_2\phi_2\phi_2\phi_2^1$

Iterative Construction

Properties:

1. $AI(f)$ is optimal
2. balancedness depends on the initial function
3. little better behavior against fast algebraic attacks
4. has same nonlinearity as the majority function

A General Construction (Claude Carlet, 2006)

Theorem

Assume that there exists a sequence of flats (i.e. of affine subspaces of F_2^n) $(A_i)_{1\leq i\leq r}$ of dimensions at least $\lceil n/2\rceil$, such that:

$$\forall i \leq r,\ card\big(A_i \setminus \big[\bigcup_{j<i} A_j \cup supp(f)\big]\big) \leq 1$$
$$F_2^n \setminus supp(f) \subseteq \bigcup_{i\leq r} A_i.$$

Then any nonzero annihilator of f has degree $\geq \lceil n/2\rceil$.

Applying this result to f and $f+1$, one obtains construction of functions with optimum AI.

This general method may contain many classes of Boolean functions with optimum AI (E.g. the *majority function*), yet it is difficult to find other classes satisfying the above conditions.

Outline

Pioneer work (Carlet-Feng, 2008)

Define the function f with support set

$$supp(f) = \{0, 1, \alpha, \alpha^2, \cdots, \alpha^{2^{n-1}-2}\},$$

where α is a primitive element of F_{2^n}.

Properties:

1. $AI(f)$ is optimal
2. balanced and with degree $n-1$
3. nonlinearity lower bounded by $2^{n-1} - \frac{2\ln 2}{\pi} n 2^{n/2}$, which is better than the previous ones, and the exact value of nonlinearity is much better than this bound for small values of n
4. resist all the main attacks ((the Berlekamp-Massey and Rønjom-Helleseth attacks, fast correlation attacks, standard and fast algebraic attacks)

Construction from $\mathcal{PS}$ Bent function (Tu-Deng, 2009)

$\mathcal{PS}_{ap}$ Bent function (Dillon, 1974)

Given any balanced Boolean function g over $\mathbb{F}_{2^k}$, the function

$$f(x,y) = g(xy^{2^k-2})$$

is a **Bent** function over $\mathbb{F}_{2^{2k}}$.

If we take g has support set $supp(g) = \{1, \alpha, \alpha^2, \cdots, \alpha^{2^{k-1}-1}\}$, then the function $f(x,y) = g(xy^{2^k-2})$ is a Bent function. Furthermore, $f(x,y)$ has $AI = k$ if the following conjecture is true.

Conjecture 1

For any integer $0 < t < 2^k - 1$, the set

$$S_t = \{(a,b) : 0 \leq a, b < 2^k,\ a + b \equiv t \pmod{2^k - 1},\ wt(\bar{a}) + wt(\bar{b}) < k\}$$

has at most 2^{k-1} elements.

Construction from $\mathcal{PS}$ Bent function (Tu-Deng, 2009)

Suitably modifying the truth table of f yields a balanced function f_1 with optimum AI and high nonlinearity:

$$f_1(x,y) = \begin{cases} g(xy^{2^k-2}), & \text{if } x \cdot y \neq 0 \\ 1, & \text{if } x = 0, y \in \Delta \\ 0, & \text{otherwise} \end{cases}$$

where $\Delta = \{\alpha^{2^{k-1}-1}, \alpha^{2^{k-1}}, \cdots, \alpha^{2^k-2}\}$.

Properties:

1. $AI(f_1) = k$
2. balanced and with degree $2k-1$
3. better lower bound for nonlinearity
4. however, bad resistance to FAA since $l(x,y)f_1(x,y) \leq k+1$. This weekness was repaired by Carlet with slight modifications on the truth table.

A construction with similar conjecture (Tang-Carlet, 2010)

Conjecture 2

For any integer $0 < t < 2^k - 1$, the set

$$S_t = \{(a, b) : 0 \leq a, b < 2^k,\ a - b \equiv t \pmod{2^k - 1},\ wt(\bar{a}) + wt(\bar{b}) < k\}$$

has at most 2^{k-1} elements.

Suppose the above conjecture is true, let g a function on $\mathbb{F}_{2^k}$ with $supp(g) = \{\alpha^s, \alpha^{s+1}, \cdots, \alpha^{s+2^{k-1}-1}\}$. Then the functions

$$f_1(x, y) = g(xy), \text{ and } f_2(x, y) = \begin{cases} g(xy) & \text{if } x \neq 0, \\ h(y) & \text{if } x = 0, \end{cases}$$

where $h(y)$ is a balanced function over $\mathbb{F}_{2^k}$ with $\deg(h) = k - 1$, have maximum $AI = k$.

A construction with similar conjecture (Tang-Carlet, 2010)

Properties:

1. $AI(f_1) = AI(f_2) = k$
2. f_2 is balanced while f_1 is not,
3. f_2 has maximum degree $n-1$, while f_1 has high degree $n-2$
4. better lower bound for nonlinearity, the experiment results for $k \leq 20$ show that the actual nonlinearities of f_1 and f_2 are very close to $2^{2k-1} - 2^{k-1}$.
5. it can checked that, for small variables, both f_1 and f_2 have good behavior against fast algebraic attacks.

A construction with a general similar conjecture (Jin,2011)

A general conjecture

For any integer $0 < t < 2^k - 1$, and any $u, v \in \mathbb{Z}^*_{2^k-1}$, the set

$$S_{t,u,v} = \{(a,b) : 0 \leq a, b < 2^k,\ ua+vb \equiv t \pmod{2^k-1},\ wt(\bar{a})+wt(\bar{b}) < k\}$$

at most 2^{k-1} elements.

This general conjecture is Conjecture 1 (2) when $(u,v) = (1,1)$ $((1,-1))$.

Let u be an positive integer such that $\gcd(2^k-1, u) = 1$. Let g a function on $\mathbb{F}_{2^k}$ with $supp(g) = \{\alpha^s, \alpha^{s+1}, \cdots, \alpha^{s+2^{k-1}-1}\}$. Then

$$f(x,y) = g(xy^{2^k-1-u})$$

has $AI = k$. In particular, if $u = 2^l$, $0 \leq l < k$, the function $f(x,y)$ is Bent function.

A construction with a general similar conjecture

If the general conjecture is true, the function $f(x, y)$ has maximum AI. Furthermore,

$$f_1(x,y) = \begin{cases} g(xy^{2^k-1-u}) & \text{if } x \neq 0, \\ g(y) & \text{if } x = 0, \end{cases}$$

where the function g has $supp(g) = \{\alpha^s, \alpha^{s+1}, \cdots, \alpha^{s+2^{k-1}-1}\}$, is a balanced function with $AI = k$.

Properties:

1. the function f_1 has maximum degree $2k - 1$.
2. if $u = 1$, f_1 is the balanced function given by Tu-Deng; and if $u = 2^k - 2$, f_1 is the balanced function given by Tang-Carlet
3. better lower bound for nonlinearity, and actual value of nonlinearity is very good
4. it has not been checked (but is believed) that f_1 have good behavior against fast algebraic attacks.

Conclusion

- The research of constructing BF over vector space with optimum AI and other good cryptographical properties is not satisfactory enough.
- In general, the functions constructed over finite field have surprisingly better properties.
- **Further work**
  - The conjectures in combinatorial field are to be proven, and the actual nonlinearities of the constructed functions are to be theoretically determined
  - The basic idea of the known constructions of BF with optimum AI originated from BCH codes, may be there are some other codes can be used to construct BF with good properties.

Thanks for attention!