Boolean functions for symmetric cryptography: immersion and insights

Sihem Mesnager

Department of Mathematics, University of Paris VIII
and University Sorbonne Paris Cité, LAGA, CNRS,
and Telecom Paris, Polytechnic Institute of Paris, France

Young Researchers Algebra Conference 2023
July 26, 2023, L'Aquila, Italy

Outline

- Preliminaries on p-ary functions and some applications
- The (practical) case of Boolean functions in cryptography
- The main mathematical problems in symmetric cryptography

Let q be a power of a prime p and r be a positive integer. The trace function $Tr_{q^r/q} : \mathbb{F}_{q^r} \to \mathbb{F}_q$ is defined as :

$$Tr_{q^r/q}(x) := \sum_{i=0}^{r-1} x^{q^i} = x + x^q + x^{q^2} + \cdots + x^{q^{r-1}}.$$

The trace function from $\mathbb{F}_{q^r} = \mathbb{F}_{p^n}$ to its prime subfield $\mathbb{F}_p$ is called the *absolute trace* function.

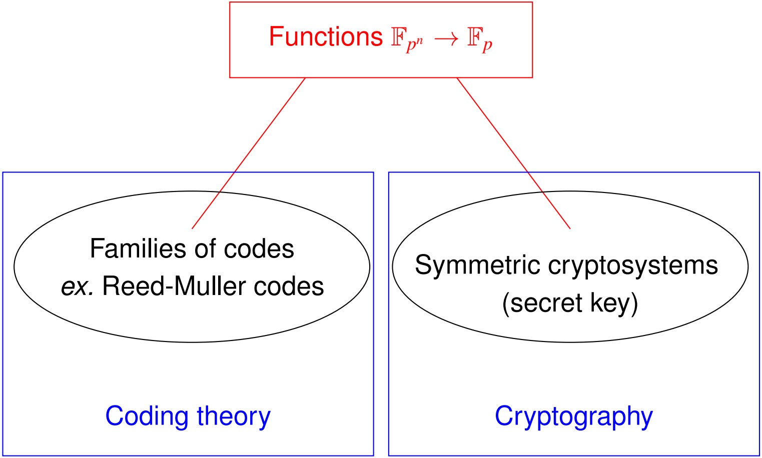
Background on functions over finite fields

Let $q = p^r$ where p is a prime.

- *A vectorial function $\mathbb{F}_q^n \to \mathbb{F}_q^m$ (or $\mathbb{F}_{q^n} \to \mathbb{F}_{q^m}$) is called an (n, m)-q-ary function.*
- *When $q = 2$, an (n, m)-2-ary function will be simply denoted an (n, m)-function. They are called S-boxes (substitution-boxes). when they are used in a block cipher (in symmetric cryptography).*
- *A Boolean function is an $(n, 1)$-function, i.e. a function $\mathbb{F}_2^n \to \mathbb{F}_2$ (or $\mathbb{F}_{2^n} \to \mathbb{F}_2$).*

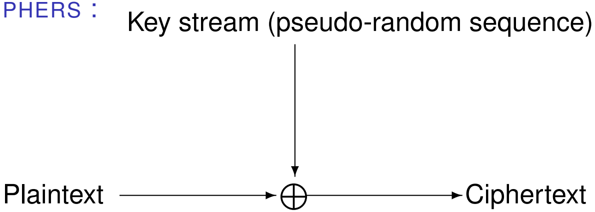
p-ary functions in cryptography and coding theory

- ☞ Functions from the finite field $\mathbb{F}_{p^n}$ to the prime field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ (p-ary functions) play an important role in coding theory and cryptography !



Cryptographic framework for Boolean functions

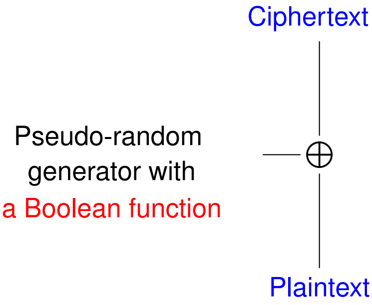
STREAM CIPHERS :



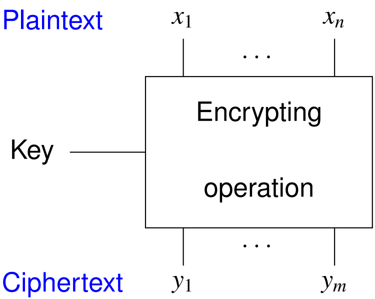
- To generate the key stream, we use models (eg. the combiner model, the filtered model) involving Boolean functions.
- The key stream has to follow properties related to the two fundamental principles introduced by Shannon : confusion and diffusion.
- The level of security of the cryptosystem against the known attacks can be quantified through some fundamental characteristics of the Boolean functions.

Boolean functions : cryptographic framework

Stream ciphers



Block ciphers (AES, DES, etc)



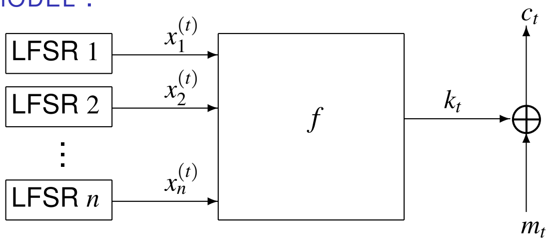
outputs of Boolean functions (depending on the key) over $x_1, \cdots, x_n$

Cryptographic framework for Boolean functions

The two models of pseudo-random generators with a Boolean function :

COMBINER MODEL :

m_t : plain text
c_t : cipher text
k_t : key stream

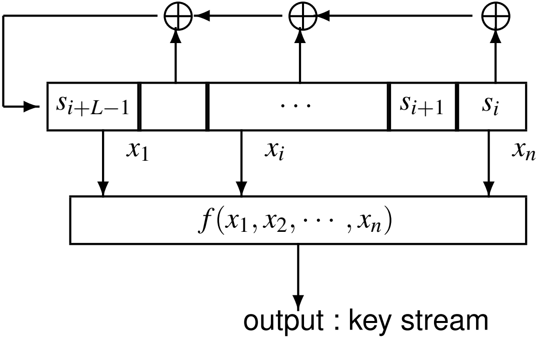


LFSR : Linear Feedback Shift Register

- A Boolean function combines the outputs of several LFSR to produce the keystream : **a combining (Boolean) function** f.
- The initial state of the LFSR's depends on a secret key.

Cryptographic framework for Boolean functions

FILTER MODEL :



- A Boolean function takes as inputs several bits of a single LFSR to produce the keystream : **a filtering (Boolean) function** f

  ☞ To make the cryptanalysis very difficult to implement, we have to pay attention when choosing the Boolean function : several recommendations (cryptographic criteria) !

Main cryptographic criteria for Boolean functions

• CRITERION 1 : To protect the system against distinguishing attacks, the cryptographic function f must be balanced, that is, its Hamming weight $wt(f) := \#\{x \in \mathbb{F}_{2^n}, f(x) \neq 0\}$ equals 2^{n-1}.

• CRITERION 2 : The cryptographic function must have a high algebraic degree to protect against the Berlekamp-Massey attack.

The Hamming distance $d_H(f,g) := \#\{x \in \mathbb{F}_{2^n} \mid f(x) \neq g(x)\}$.

• CRITERION 3 : To protect the system against linear attacks and correlation attacks, the Hamming distance from the cryptographic function to all affine functions must be large. i.e. high nonlinearity $\mathrm{nl}(f) := \min_{l \in A_n} d_H(f,l)$; l : affines functions.

• CRITERION 4 : To be resistant against correlation attacks on combining registers, a combining function f must be m-resilient where m is as large as possible. i.e. f must stay balanced if we fix at most m coordinates.

• CRITERION 5 : To be resistant against algebraic attacks, f must be of high algebraic immunity that is, close to the maximum $\lceil \frac{n}{2} \rceil$. Algebraic immunity of f : $\mathrm{AI}(f)$ is the lowest degree of any nonzero function g such that $f \cdot g = 0$ or $(1+f) \cdot g = 0$.
But this condition is insufficient because of Fast Algebraic Attacks !

Representations of p-ary functions

There is a unique representation of F from $\mathbb{F}_{p^n}$ to $\mathbb{F}_{p^n}$: $F(x) = \sum_{i=0}^{2^n-1} a_i x^i$ with $a_i \in \mathbb{F}_{p^n}$. A unique univariate form of a p-ary function, called *trace representation*, is given by

$$f(x) = \sum_{j \in \Gamma_n} Tr_{p^{o(j)}/p}(A_j x^j) + A_{p^n-1} x^{p^n-1}$$

- Γ_n is the set of integers obtained by choosing the smallest element, called the *coset leader*, in each p- cyclotomic coset modulo $p^n - 1$;
- $o(j)$ is the size of the cyclotomic coset containing j (that is the smallest positive integer such that $jp^{o(j)} \equiv j \pmod{p^n - 1}$; $o(j)$ divides n ;
- $A_j \in \mathbb{F}_{p^{o(j)}}$ and $A_{p^n-1} \in \mathbb{F}_p$.

The *algebraic degree* of f : $deg(f) := max\{w_p(j) \mid A_j \neq 0\}$ where $w_p(j)$ is the number of nonzero entries in the p-ary expansion of j. For example, affine functions f : $d_{alg}(f) = 1$.

Background on Boolean functions : existence of the polynomial form

Any function $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ admits a unique representation :
$f(x) = \sum_{j=0}^{2^n-1} A_j x^j$; $A_j \in \mathbb{F}_{2^n}$;

- f is Boolean iff

$A_0, A_{2^n-1} \in \mathbb{F}_2$ and $A_{2j \ mod \ 2^n-1} = (A_{j \ mod \ 2^n-1})^2; 0 < j < 2^n - 1$

- $[1, 2^n - 2] = \cup_{r=1}^{c} \Gamma_r$;

$\Gamma_r = \{j_r \ mod \ 2^n - 1, 2j_r \ mod \ 2^n - 1, \cdots, 2^{o(j_r)-1} j_r \ mod \ 2^n - 1\}$

$$
\begin{aligned}
f(x) &= A_0 + A_{2^n-1} x^{2^n-1} + \sum_{r=1}^{c} \sum_{s=0}^{o(j_r)-1} A_{2^s j_r \ mod \ 2^n-1} x^{2^s j_r} \\
&= A_0 + A_{2^n-1} x^{2^n-1} + \sum_{r=1}^{c} \sum_{s=0}^{o(j_r)-1} (A_{j_r \ mod \ 2^n-1} x^{j_r})^{2^s} \\
&= A_0 + A_{2^n-1} x^{2^n-1} + \sum_{r=1}^{c} Tr_{2^{o(j_r)}/2}(A_{j_r \ mod \ 2^n-1} x^{j_r})
\end{aligned}
$$

where $A_0, A_{2^n-1} \in \mathbb{F}_2$, $A_{j_r \ mod \ 2^n-1} \in \mathbb{F}_{2^{o(j_r)}}$.

Background on Boolean functions : representation

Example : Let $n = 4$. $f : \mathbb{F}_{2^4} \to \mathbb{F}_2$,

$f(x) = \sum_{j \in \Gamma_4} Tr_{2^{o(j)}/2}(A_j x^j) + A_{2^4-1} x^{2^4-1}$;

Γ_4 is the set obtained by choosing one element in each cyclotomic class of 2 modulo $2^n - 1 = 2^4 - 1 = 15$. $C(j)$ the cyclotomic coset of 2 modulo 15 containing j.

$C(j) = \{j, j2, j2^2, j2^3, \cdots, j2^{o(j)-1}\}$ where $o(j)$ is the smallest positive integer such that $j2^{o(j)} \equiv j \pmod{2^n - 1}$.

The cyclotomic cosets modulo 15 are :

$C(0) = \{0\}$

$C(1) = \{1, 2, 4, 8\}$

$C(3) = \{3, 6, 12, 9\}$

$C(5) = \{5, 10\}$

$C(7) = \{7, 14, 11, 13\}$

We find $\Gamma_4 = \{0, 1, 3, 5, 7\}$

$f(x) =$

$Tr_{o(1)/2}(A_1 x^1) + Tr_{o(3)/2}(a_3 x^3) + Tr_{o(5)/2}(A_5 x^5) + Tr_{o(7)/2}(A_7 x^7) + A_0 + A_{15} x^{15}$

$= Tr_{4/2}(A_1 x) + Tr_{4/2}(A_3 x^3) + Tr_{2/2}(A_5 x^5) + Tr_{4/2}(A_7 x^7) + A_0 + A_{15} x^{15}$

where $A_1, A_3, A_7 \in \mathbb{F}_{2^4}$, $A_5 \in \mathbb{F}_{2^2}$ and $A_0, A_{15} \in \mathbb{F}_2$;

$Tr_{4/2} : \mathbb{F}_{2^4} \to \mathbb{F}_2$; $x \mapsto x + x^2 + x^{2^2} + x^{2^3}$;

$Tr_{2/2} : \mathbb{F}_{2^2} \to \mathbb{F}_2$; $x \mapsto x + x^2$.

Representations of p-ary functions

Viewed over $\mathbb{F}_p^n$, a p-ary function f has a representation as a unique multinomial in $x_1, \cdots, x_n$, where the variables x_i occur with exponent at most $p-1$. This is called the *multivariate representation* or algebraic normal form (ANF) of f :

$$f(x_1, \ldots, x_n) = \sum_{(j_1,\ldots,j_n)\in\mathbb{F}_p^n} a_{(j_1,\ldots,j_n)} \prod_{i=1}^{n} x_i^{j_i},$$

with coefficients $a_{(j_1,\ldots,j_n)} \in \mathbb{F}_p$. The degree of a monomial $\prod_{i=1}^n x_i^{j_i}$. is $j_1 + \cdots + j_n$.

The *algebraic degree* of f (denoted by $d_{alg}(f)$) is the global (total) degree of its multivariate representation, that is, the largest degree of all monomials in its ANF with a nonzero coefficient $a_{(j_1,\ldots,j_n)}$.

Example : A Boolean function $f(x) = f(x_1, \ldots, x_n)$ can be (uniquely) written as $f(x) = \bigoplus_{I\subseteq\{1,\ldots,n\}} a_I \left(\prod_{i\in I} x_i\right)$;

where " $\oplus$ " is the addition is made modulo 2 and a_I belongs to $\mathbb{F}_2$.

Background on Boolean functions

Example : Let $n = 3$.

x_1	x_2	x_3	$f(x_1,x_2,x_3)$	$g(x_1,x_2,x_3)$	$h(x_1,x_2,x_3)$
0	0	0	0	0	1
0	0	1	0	1	0
0	1	0	1	1	0
0	1	1	1	1	1
1	0	0	1	1	0
1	0	1	0	0	1
1	1	0	1	1	1
1	1	1	0	0	0

$f(x_1,x_2,x_3) = x_1 + x_2 + x_1x_2 + x_1x_3$; $d_{alg}(f) = 2$;
$g(x_1,x_2,x_3) = x_1 + x_2 + x_3 + x_1x_2 + x_2x_3 + x_1x_2x_3$; $d_{alg}(g) = 3$;
$h(x_1,x_2,x_3) = 1 + x_1 + x_2 + x_3$; $d_{alg}(h) = 1$.

Symmetric encryption

$F : \mathbb{F}_{2^n} \to \mathbb{F}_{2^m}$

- **How does symmetric cryptography work ?**
  Attack on a cryptographic system $\Rightarrow$ cryptographic parameter of F $\Rightarrow$ cryptographic criterion

  Main Goal : Design "optimal" cryptographic functions to resist attacks !

- two big problems :

1. it is also necessary to study the properties of the functions which satisfy several of them cryptographic criteria (and not just one) compromise $\hookrightarrow$ to be found mathematically !
2. space too large (doubling exponential 2^{2^n} if $m = 1$) $\hookrightarrow$ need math !

☞ mathematical work to do for each cryptographic property :

a studies its algebraic properties of the function F satisfying cryptographic properties ;

b provides an efficient mathematical characterisation of each cryptographic criterion ;

c design functions F (given by their algebraic representations), which are optimal with respect to a cryptographic property.

And that is not enough : they must also be classified (think of the notions of equivalence) !

☞ **Key tool to study f :** need discrete Fourier theory

$f : \mathbb{F}_{p^n} \to \mathbb{F}_p$ (resp. $F : \mathbb{F}_{p^n} \to \mathbb{F}_{p^m}$), p prime

The discrete Hadamard Walsh (Fourier) transform W_f de f (resp. de F) :

Symmetric Cryptography

☞ **Key tool to study f : need discrete Fourier theory**

$f : \mathbb{F}_{p^n} \to \mathbb{F}_p$ (resp. $F : \mathbb{F}_{p^n} \to \mathbb{F}_{p^m}$), p prime

The discrete Hadamard Walsh (Fourier) transform W_f of f (resp. of F) :

$$W_f(a) := \sum_{x \in \mathbb{F}_{p^n}} \zeta_p^{f(x) - Tr_{p^n/p}(ax)}$$

$$W_F(a,b) := \sum_{x \in \mathbb{F}_{p^n}} \zeta_p^{Tr_{p^m/p}(bF(x)) - Tr_{p^n/p}(ax)}$$

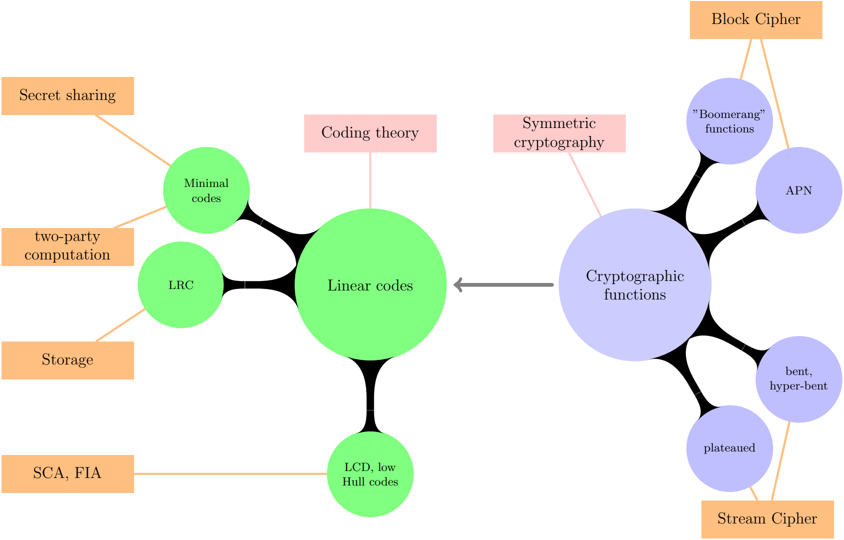
where $(a,b) \in \mathbb{F}_{p^n} \times \mathbb{F}_{p^m} \setminus \{0\}$

- The p-ary functions $F_b : x \mapsto b \cdot F(x) := Tr_{p^m/p}(bF(x))$, $b \in \mathbb{F}_{p^m}$ where $b \neq 0$ (F_0 is the null function) are called the *components* of F
- W_f (resp. W_F) is with values in the cyclotomic field $\mathbb{Q}(\zeta_p)$ where $\zeta_p = exp\left(\frac{2\pi i}{p}\right)$ is a p-th primitive root of the unit.

- There is an algorithm for calculating W_f (resp. W_F) but that is not enough ! (complexity too high $\hookrightarrow$ evaluate W_f mathematically !)
- Parseval identity : $\sum_{b\in\mathbb{F}_{p^n}} |W_f(b)|^2 = p^{2n}$
- Note that the notion of a Walsh transform refers to a scalar product, it is convenient to choose the isomorphism such that the canonical scalar product "$\cdot$" in $\mathbb{F}_{p^n}$ coincides with the canonical scalar product in $\mathbb{F}_{p^n}$, which is the trace of the product $b \cdot x := Tr_{p^n/p}(bx)$.
- Walsh transform of a very simple (but important function : Kloosterman sums on $\mathbb{F}_{2^m}$: $K_m(a) := \sum_{x\in\mathbb{F}_{2^m}} (-1)^{Tr_{2^m/2}(ax+\frac{1}{x})}$.

Extension of the theory of cryptographic Boolean functions to :

1. Vectorial Boolean functions
2. Functions in odd characteristic
3. Generalized functions



Block Cipher
Secret sharing
Coding theory
Symmetric cryptography
"Boomerang" functions
Minimal codes
APN
two-party computation
LRC
Linear codes
Cryptographic functions
Storage
bent, hyper-bent
plateaued
SCA, FIA
LCD, low Hull codes
Stream Cipher

Approaches and tools used to solve problems in this topic

Approaches : an algebraic approach, combinatoric approach, asymptotic approach, and geometric approach.

Mathematical tools :

- discrete Fourier/Walsh transforms
- polynomials over finite fields (polynomials, Linearized polynomials, permutation polynomials, involutions, Dickson polynomials, polynomials e- to-1, etc.)
- functions over finite fields (symmetric functions, quadratic forms, etc.)
- tools from algebraic geometry (algebraic, elliptic curves, hyper-elliptic curves, etc.)
- finite geometry (oval polynomials, hyperovals, etc.)
- linear algebra and group theory
- tools from combinatorics
- tools from arithmetic number theory

The nonlinearity of p-ary functions (where $p = 2$)

The *nonlinearity* of f defined over $\mathbb{F}_2^n$ is the minimum Hamming distance to the set A_n of all affine functions :

$$nl(f) = \min_{g \in A_n} d(f, g),$$

where $d(f, g)$ is the Hamming distance between f and g, that is $d(f, g) := \#\{x \in \mathbb{F}_2^n \mid f(x) \neq g(x)\}$. The relationship between nonlinearity and Walsh spectrum of f is

$$nl(f) = 2^{n-1} - \frac{1}{2} \max_{\omega \in \mathbb{F}_2^n} |W_f(\omega)|.$$

By Parseval's identity $\sum_{\omega \in \mathbb{F}_2^n} W_f(\omega)^2 = 2^{2n}$, it can be shown that $\max\{|W_f(\omega)| : \omega \in \mathbb{F}_2^n\} \geq 2^{\frac{n}{2}}$ which implies that $nl(f) \leq 2^{n-1} - 2^{\frac{n}{2}-1}$.

Let n be an even integer. An n-variable Boolean function is said to be bent if the upper bound $2^{n-1} - 2^{n/2-1}$ on its nonlinearity $nl(f)$ is achieved with equality.

- ☞ Bent Boolean functions function f defined over $\mathbb{F}_{2^n}$ exist only when n is even !
- ☞ The notion of bent function was introduced by [Rothaus 1976] and attracted a lot of research of more than four decades. Such functions are extremal combinatorial objects with several application areas, such as coding theory, maximum length sequences, and cryptography !
- ☞ f is bent over $\mathbb{F}_{2^n}$ if and only if, $W_f(\omega) \in \{2^{\frac{n}{2}}, -2^{\frac{n}{2}}\}$, for all $\omega \in \mathbb{F}_{2^n}$ ([Dillon 1974]).

☞ Linear Cryptanalysis [Matsui (1993)] $\Rightarrow$ nonlinearity $\mathrm{Nl}(F)$ de $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$,

$$\mathrm{Nl}(F) = \min_{b \in \mathbb{F}_{2^m}, b \neq 0} \{\mathrm{nl}(Tr_{2^m/2}(bF))\} = 2^{n-1} - \frac{1}{2} \max_{a \in \mathbb{F}_2^n, b \in \mathbb{F}_2^m, b \neq 0} |W_F(a,b)|.$$

- ✭ The higher the value of $\mathrm{Nl}(F)$, the better resistance to linear cryptanalysis.
- ✭ When $n = m$ odd, $\mathrm{Nl}(F)$ is bounded by $2^{n-1} - 2^{\frac{n-1}{2}}$. The functions reaching this upper bound are the AB functions.
- ✭ When $m = 1$, $\mathrm{Nl}(F)$ is bounded by $2^{n-1} - 2^{\frac{n}{2}-1}$. The functions reaching this upper bound are the bent functions (n even).

$\hookrightarrow$ A very difficult parameter to study mathematically ! !