A family of boolean functions with good cryptographic properties

Guillermo Sosa Gómez
guillermo.sosa@cimat.mx
Departamento de Matemáticas, CUCEI
Universidad de Guadalajara
Guadalajara, México

Octavio Páez Osuna
octavio.paezosuna@ronininstitute.org
Ronin Institute for Independent Scholarship
Montclair, NJ 070043
USA

February 19, 2019

Abstract

In 2005, [2] Philippe Guillot presented a new construction of Boolean functions using linear codes as an extension of Maiorana-McFarland's construction of bent functions. In this paper, we study a new family of Boolean functions with cryptographically strong properties such as nonlinearity, propagation criterion, resiliency and balance. The construction of cryptographically strong boolean functions is a daunting task and there is currently a wide range of algebraic techniques and heuristics for constructing such functions , however these methods can be complex, computationally difficult to implement and not always produce a sufficient variety of functions. We present in this paper a construction of Boolean functions using algebraic codes following Guillot's work.

1 Introduction

Here we follow [1]. Let $\mathbb{F}_2^n$ be the binary vector space of dimension n over the Galois Field of two elements $\mathbb{F}_2$. Given two vectors a,$b \in \mathbb{F}_2^n$, we define the scalar product

$$a \cdot b = (a_1 b_1 \oplus \ldots \oplus a_n b_n)$$

and the sum as

$$a \oplus b = (a_1 \oplus b_1, \ldots, a_n \oplus b_n),$$

where the product and sum $\oplus$ (also called XOR) are over $\mathbb{F}_2$.

A n-variable **boolean function** f is a mapping

$$f : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2.$$

We will denote by $\mathcal{B}_n$ the set of all Boolean functions of n variables. The set $\mathcal{B}_n$ is a vector space over $\mathbb{F}_2$ with the addition $\oplus$ defined by

$$(f \oplus g)(x) = f(x) \oplus g(x),$$

for any $f, g \in \mathcal{B}_n$ and any $x \in \mathbb{F}_2^n$. The **polar form** $\hat{f} : \mathbb{F}_2^n \longrightarrow \mathbb{R}$, or sign function, of a boolean function $f \in \mathcal{B}_n$, is defined by

$$\hat{f}(x) = (-1)^{f(x)}.$$

The **support** f, denoted by $Supp(f)$, is the set of vectors in $\mathbb{F}_2^n$ whose image under f is 1. That is

$$Supp(f) = \{x \in \mathbb{F}_2^n \mid f(x) = 1\}.$$

The **weight** of a boolean function $f \in \mathcal{B}_n$, denoted by $w(f)$, is the cardinality of its support, that is $w(f) = |Supp(f)|$. We will say that a function $f \in \mathcal{B}_n$ is **balanced** if $w(f) = 2^{n-1}$, that is, the truth table of f contains the same number of 0 and 1. This property is desirable in a Boolean function to resist differential attacks such as those introduced by A. Shamir against the DES algorithm.

A boolean function $f \in \mathcal{B}_n$ is called **affine** if we can write it as

$$f(x) = \langle a, x \rangle \oplus b$$

for some $a \in \mathbb{F}_2^n$ and $b \in \mathbb{F}_2$. If $b = 0$, we say that f is **linear function**. The set of affine functions will be denoted by $\mathcal{A}_n$. Let $f, g \in \mathcal{B}_n$. The **distance**, $d(f, g)$, between f and g, is the weight of the function $f \oplus g$, *i.e.*,

$$d(f, g) = w(f \oplus g).$$

The **nonlinearity** of a boolean function $f \in \mathcal{B}_n$, denoted by $\mathcal{N}_f$, is the minimum distance between f and the set of affine functions $\mathcal{A}_n$, *i.e.*,

$$\mathcal{N}_f = min\{d(f, \varphi) \mid \varphi \in \mathcal{A}_n\}.$$

A high nonlinearity is desired to reduce the effect of linear cryptanalysis attacks.

The **Truth Table** of a Boolean function f is the vector, indexed by the elements of $\mathbb{F}_2^n$ (in lexicographical order),

$$(f(\bar{0}), f(\bar{1}), \ldots, f(\overline{2^n - 1}))$$

where $\bar{0} = (0, \ldots, 0, 0)$, $\bar{1} = (0, \ldots, 0, 1)$, $\ldots, \overline{2^n - 1} = (1, \ldots, 1, 1)$. The **polar truth table** of f is the $(1, -1)$ sequence defined by

$$\left((-1)^{f(\bar{0})}, \ldots, (-1)^{f(\overline{2^n-1})}\right).$$

A Boolean function in $\mathbb{F}_2^n$ can be expressed uniquely as a polynomial in

$$\mathbb{F}_2[x_1, \ldots, x_n] / \left(x_1^2 \oplus x_1, \ldots, x_n^2 \oplus x_n\right)$$

through its **Algebraic Normal Form (ANF)**

$$f(x) = \sum_{a \in \mathbb{F}_2^n} c_a x_1^{a_1} \cdots x_n^{a_n}, \tag{1}$$

where $c_a \in \mathbb{F}_2$ and $a = (a_1, \ldots, a_n)$, with $c_a = \sum_{x \leq a} f(x)$, where $x \leq a$ means that $x_i \leq a_i$, for all $1 \leq i \leq n$. That is, $c_a = g(a_1, \ldots, a_n)$, and g is a function in $\mathcal{B}_n$ called the **Möbius Transform** of f, denoted by $g = \mu(f)$. The **Algebraic Degree** of a boolean function f is the degree of its ANF. It follows that the algebraic degree of $f \in \mathcal{B}_n$ does not exceed $n - 1$.

The **Walsh-Hadamard Transform** of a function f in $\mathbb{F}_2^n$ is the mapping $H(f) : \mathbb{F}_2^n \to \mathbb{R}$, defined by:

$$H(f)(h) = \sum_{x \in \mathbb{F}_2^n} f(x)(-1)^{h \cdot x}, \tag{2}$$

Let $f \in \mathcal{B}_n$ be a boolean function, let S be an arbitrary subspace of $\mathbb{F}_2^n$ and $S^\perp$ the dual(annihilator) of S, *i.e.*,

$$S^\perp = \{x \in \mathbb{F}_2^n : x \cdot s = 0, \, \forall s \in S\}$$

then,

$$\sum_{u \in S} H(f)(u) = 2^{dim S} \sum_{u \in S^\perp} f(u). \tag{3}$$

From the definition of the Walsh-Hadamard Transform, it follows that $H(\hat{f})(u)$ equals the number of zeros minus the number of ones in the binary vector $f \oplus l_u (l_u \in \mathcal{A}_n$, or, $l_u(v) = \sum_{i=1}^n u_i v_i)$ and such that

$$H(\hat{f})(u) = 2^n - 2d(f, \sum_{i=1}^n u_i v_i) \tag{4}$$

$$d(f, \sum_{i=1}^n u_i v_i) = \frac{1}{2}(2^n - H(\hat{f})(u)) \tag{5}$$

$$d(f, 1 \oplus \sum_{i=1}^n u_i v_i) = \frac{1}{2}(2^n + H(\hat{f})(u)) \tag{6}$$

We summarize these earlier results in the following theorem

Theorem 1.1. *The nonlinearity f is determined by the Walsh-Hadamard Transform of f,* i.e.

$$\mathcal{N}_f = 2^{n-1} - \frac{1}{2} \max_{u \in \mathbb{F}_2^n} |H(\hat{f})(u)|. \tag{7}$$

In what follows we summarize some factors which are important in the design of Boolean functions with good cryptographic properties [3]:

A n-variable boolean function is said to have **Correlation immunity** of order m if and only if $H(\hat{f})(u) = 0$, with $1 \leq w(u) \leq m$. A Boolean function with Correlation Immunity of order m and balanced is called m-**resilient**. The fundamental relationship between the number of variables

n, algebraic degree d and order of correlation immunity m of a boolean function is

$$m + d \leq n.$$

The **autocorrelation function** $r_{\hat{f}}(s)$ for a Boolean function f is defined from its polar representation as

$$r_{\hat{f}}(s) = \sum_{x \in \mathbb{F}_2^n} \hat{f}(x)\hat{f}(x \oplus s).$$

This value is proportional to the imbalance of all the first-order derivatives of the Boolean function. Small autocorrelation values are desirable while boolean functions having larger values are considered weak.

We say that a Boolean function has **Propagation Criteria** of order l, denoted by $PC(l)$ if $f(x) \oplus f(x \oplus u)$ is balanced for all u with $1 \leq w(u) \leq l$.

The **Strict Avalanche Criterion (SAC)**, refers to the effect of changing all input bits. A boolean function f is said to satisfy SAC if $f(x) \oplus f(x \oplus u)$ is balanced for all u with $w(u) = 1$.

Let $q = 2^m$, and let $\mathbb{F}_q$ be the finite field with q elements. An $\mathbb{F}_q-$linear **error correcting code** C of length n is an $\mathbb{F}_q-$linear subspace of $\mathbb{F}_q^n$. The elements of C are called words. The weight $wt(x)$ of a word x in C is the number of its non-zero coordinates. The minimum weight d of the code C is defined as the minimum of the weights among all non-zero words occurring in C. For $x, y \in C$, we define the Hamming distance $d(x, y)$ between x and y as $wt(x - y)$. The **minimum distance** of a code C is defined as

$$d = min\{d(x, y)|x, y \in C, x \neq y\}.$$

If k is the dimension of C as a vector space over $\mathbb{F}_q$, then we say that C is a

$$[n, k, d]_q$$

error correcting code. The *Singleton bound* states that the parameters of a code C must satisfy

$$n + 1 \geq k + d.$$

A code satisfying the previous inequality with equality is called a maximum distance separable code, or simply a MDS-Code.

For $q \geq 2, h \geq 1$. Let $Q = q^h$. Consider two codes which we call outer code and inner code. Let C be outer code with parameters $[N, K, D]_Q$ and let I be inner code with parameters $[n, h, d]_q$. The concatenation method constructs a code F over $\mathbb{F}_q$ out of a code over $\mathbb{F}_Q$. The first step is to fix any isomorphism $\varphi : \mathbb{F}_Q \longrightarrow I \subseteq \mathbb{F}_q^n$. Then

$$F := \{(\varphi(c_1)), \ldots, \varphi(x_N))|(x_1, \ldots, x_N) \in C\}.$$

The code F has parameters

$$[N \cdot n, K \cdot h, D \cdot d]_q.$$

2 Maiorana-McFarland-Guillot's construction

The Maiorana-McFarland construction was originally designed to obtain bent functions. It has been extended to construct resilient functions [2].

For $n \geq 2$ an integer and $\mathbb{F}_2^n = E \oplus F$ a decomposition into two complementary subspaces: E of dimension p y F of dimension $q = n - p$.

For any application $\pi : E \longrightarrow \mathbb{F}_2^n$ and any application $h : E \longrightarrow \mathbb{F}_2$ the Maiorana-McFarland(MM) construction defines a Boolean function f as follows:

$$\begin{array}{rccc} f: & E \oplus F & \longrightarrow & \mathbb{F}_2 \\ & x + y & \mapsto & \pi(x) \cdot y + h(x), \end{array}$$

The application π is defined on $\mathbb{F}_2^n$, but since $\pi(x)$ is wrapped by an internal product with an element of F, the value of f it is invariant when $\pi(x)$ is moved by a vector of $F^\perp$. So, π can be considered to be defined over the space $\mathbb{F}_2^n / F^\perp \cong E^\perp$, so $\pi : E \longrightarrow E^\perp$.

One of the properties we are interested in from a Boolean function is the Propagation Criteria, in [2] it is shown that for a Boolean function to have Propagation Criteria of order k it is enough that the coset $x_0 + F$, with $x_0 \in E$, has $w(x_0 + F) > k$. Therefore, to find a Boolean function with $PC(k-1)$ it is enough to select an appropriate x_0 in the complement of F, such that the lateral class $x_0 + F$ has weight $\geq k$.

3 Reed-Solomon Codes

The class of Reed-Solomon Codes is considered of great importance in coding theory. They are members of the family of algebraic codes. Recall one of the standard descriptions of an extended Reed-Solomon code over $\mathbb{F}_q$([4]). Let $\mathbb{F}_q = \{0, 1, \alpha, \alpha^2, \cdots, \alpha^{q-2}\}$. Consider the set

$$L = \{f(x) \in \mathbb{F}_q[q] \mid degree(f(x)) < r\}.$$

The code Reed-Solomon code $RS(r, q)$ of length $n = q$ is defined by

$$RS(r, q) := \{c = (f(0), f(1), f(\alpha), f(\alpha^2), \cdots, f(\alpha^{q-2})) \mid f(x) \in L\}$$

Because a polynomial of degree l has at most l zeros in $\mathbb{F}_q$, we see that $RS(r, q)$ has minimum distance $d = q - r + 1$, which is the best possible, *i.e.*, $RS(r, q)$ is a maximum distance separable(MDS) code [4]. The code $RS(r, q)$ has parameters

$$[q, r, q - r + 1]_q .$$

In this paper we will assume that $q = 2^m$, then $RS(r, q)$ has parameters

$$[2^m, r, 2^m - r + 1]_{2^m}$$

4 Boolean functions from $RS(r, 2^m)$

For our construction of boolean functions we will use a concatenated Reed-Solomon code. Let $C = RS(r, 2^m)$, this is our outer code. Let I be the all even weight codewords, then with parameters $[m+1, m, 2]_2$. After concatenation we obtain a code F with parameters

$$[(m+1)2^m, m \cdot r, 2(2^m - r + 1)]_2 .$$

We will use our code F as the main ingredient to the Maiorana-McFarland construction. Obtaining a new family of Boolean functions, in $n = (m+1)2^m$ variables. The dimension of the complementary vector space E is therefore $b = (m+1)2^m - m \cdot r$. And $\mathbb{F}_2^{(m+1)2^m} = E \oplus F$.

We focus now in the lateral class $x_0 + F$. As F is constructed by evaluating all polynomials of degree less than r over $\mathbb{F}_{2^m}[x]$, we can assume that x_0 is also constructed by evaluating a polynomial $L(x)$ over $\mathbb{F}_{2^m}[x]$. A polynomial $L(x)$ can be obtained using Lagrange interpolation whose evaluation produces a suitable concatenated x_0.

Let $a_1, ..., a_r$ be a set of information coordinates for the code $RS(r, 2^m)$, by Lagrange interpolation, we can obtain a polynomial $L(x)$ of degree r such that $L(a_i) = 0$ for $i = 1, ..., r$ and $L(b) \neq 0$ for all $b \in \mathbb{F} - \{a_1, \ldots, a_r\}$. The vector $ev(L)$ is a vector in the complement of $RS(r, 2^m)$ as a vector space over $\mathbb{F}_{2^m}$, and the lateral class $ev(L) + RS(r, 2^m)$ has minimum weight $\geq 2^m - r$. Let x_0 be the image of $ev(L)$ under concatenation, it follows that x_0 is a vector in the complement of F as a binary vector space and, by construction, the minimum weight of the lateral class $x_0 + F$ is $\geq 2(2^m - r)$. Thus, by using our proposed F and x_0 in Guillot's construction, we obtain a boolean function satisfying $PC(2^{m+1} - 2r - 1)$.

5 Example

Suppose we want to build a 12 variable boolean function. As the main ingredient we use the Reed-Solomon code $C = RS(3, 4)$ over $\mathbb{F}_4$ with parameters $[4, 3, 2]$
A generator matrix for C is

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & \alpha & \alpha+1 & 1 \\ 0 & \alpha+1 & \alpha & 1 \end{pmatrix}.$$

Where $\alpha^2 + \alpha + 1 = 0$. We now obtain a binary code from C by concatenation with the even weight code $I = \{000, 101, 011, 110\}$ with parameters $[3, 2, 2]$. Any other 2-dimensional binary code will serve as an inner code. The next step is to choose any homomorphism ν between F_4 and I as vector spaces over $\mathbb{F}_2$. For our example we choose $0 \mapsto 000, 1 \mapsto 101, \alpha \mapsto 011, \alpha + 1 \mapsto 110$. After concatenation we obtain a binary code F with parameters $[12, 6, 4]$. A systematic generator matrix for F is given by

$$G_F = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

The row span of G_F is the binary vector space F in the MM construction. As G_F is systematic, that is, the first 6 columns are the information coordinates of code F, we may easily describe the complementary space E with generator matrix

$$G_E = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

In this example we have $n = 12, p = 6, q = 6$, so we will build a two to one function π. The next step is to build $x_0 \in E$ by concatenation of the evaluation vector of $L(x) = x^2 + x$. We obtain $x_0 = \{0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0\} \in E$. For each lateral class $u + F^{\perp}$ with $u \in E^{\perp}$ we construct the sets

$$E_0 = \{v \in u + F^{\perp} : v \cdot x_0 = 0\}$$

and

$$E_1 = \{v \in u + F^{\perp} : v \cdot x_0 = 1\}.$$

Let $d_0 = d(E_0), d_1(E_1)$ be the minimum distances of E_0 and E_1 respectively, and let $d_j = max\{d_0, d_1\}$. Next we store in an array the pairs

(u, j). In this example the array is given by

$$(u, 0/1) = \begin{pmatrix} u & h_u \\ \{0,0,1,0,1,1\} & 0 \\ \{0,0,1,1,0,1\} & 0 \\ \{0,0,1,1,1,0\} & 0 \\ \{0,1,0,0,1,1\} & 0 \\ \{0,1,0,1,1,0\} & 0 \\ \{0,1,0,1,1,1\} & 1 \\ \{0,1,1,0,0,1\} & 0 \\ \{0,1,1,0,1,0\} & 1 \\ \{0,1,1,0,1,1\} & 0 \\ \{0,1,1,1,0,0\} & 0 \\ \{0,1,1,1,0,1\} & 1 \\ \{0,1,1,1,1,0\} & 0 \\ \{0,1,1,1,1,1\} & 1 \\ \{1,0,0,0,1,1\} & 0 \\ \{1,0,0,1,0,1\} & 0 \\ \{1,0,0,1,1,0\} & 1 \\ \{1,0,0,1,1,1\} & 0 \\ \{1,0,1,0,0,1\} & 1 \\ \{1,0,1,0,1,1\} & 0 \\ \{1,0,1,1,0,0\} & 0 \\ \{1,0,1,1,0,1\} & 0 \\ \{1,0,1,1,1,0\} & 0 \\ \{1,0,1,1,1,1\} & 1 \\ \{1,1,0,0,0,1\} & 0 \\ \{1,1,0,0,1,0\} & 0 \\ \{1,1,0,1,0,0\} & 0 \\ \{1,1,0,1,0,1\} & 1 \\ \{1,1,0,1,1,0\} & 0 \\ \{1,1,0,1,1,1\} & 1 \\ \{1,1,1,0,0,0\} & 0 \\ \{1,1,1,0,0,1\} & 0 \\ \{1,1,1,0,1,0\} & 0 \end{pmatrix}$$

As you may have noticed all u in the previous array have weight ≥ 3, as expected from Guillot's results, so the boolean function we will construct will have resilience order 2. For $x \in E$ we define $\pi(x) = \pi(x + x_0) \in \mathbb{F}_2$ at random, and define $h(x) = h_u$ and $h(x + X_0) = h_u + h_t$ where h_t is a random value in $\mathbb{F}_2$.

By using π and h defined above in the Maoirana-McFarland construction the following cryptographic paramenters for the boolean function f were checked using sage:

- Balanced
- Non linearity: 1984
- Algebraic Immunity of order 5
- Propagation criteria of order 3
- Resilience of order 2

References

[1] Thomas W Cusick and Pantelimon Stanica. *Cryptographic Boolean functions and applications*. Academic Press, 2009.

[2] Philippe Guillot. Cryptographical boolean functions construction from linear codes. *Boolean Functions: Cryptography and Applications*, 387:141, 2005.

[3] Francisco Rodríguez Henríquez. De la búsqueda de funciones booleanas con buenas propiedades criptográficas, 2007.

[4] Jacobus Hendricus Van Lint. *Introduction to coding theory*, volume 86. Springer Science & Business Media, 2012.