Highly Nonlinear Balanced Boolean Functions with a Good Correlation-Immunity

Eric Filiol[1,2] and Caroline Fontaine[1]

[1] INRIA, projet CODES, Domaine de Voluceau
78153 Le Chesnay Cedex, FRANCE
[2] Ecoles militaires de Coëtquidan, DGER/CREC
56381 Guer Cedex, FRANCE
{Eric.Filiol,Caroline.Fontaine}@inria.fr

Abstract. We study a corpus of particular Boolean functions: the idempotents. They enable us to construct functions which achieve the best possible tradeoffs between the cryptographic fundamental properties: balancedness, correlation-immunity, a high degree and a high nonlinearity (that is a high distance from the affine functions). They all represent extremely secure cryptographic primitives to be implemented in stream ciphers.

Keywords: Boolean function, correlation, nonlinearity, balancedness, idempotent, stream cipher.

1 Introduction

One of the most important types of keystream generators is the one generally used for stream cipher. A number n of linear feedback shift registers (LFSRs) are combined by a *Boolean function*, that is to say a function mapping $\mathbf{F}_2^n$ to $\mathbf{F}_2$. Standard cryptographic criteria concerning the LFSRs are well-known (good statistical properties, large period, large period complexity). The Boolean function, whose main goal is to erase, to break the intrinsic linearity of the LFSRs is of great importance and must have some properties to resist certain attacks and particularly the Siegenthaler's correlation attack [23]. The different criteria for a Boolean function (*balancedness*, *correlation-immunity* and high *nonlinearity*) have been extensively separately studied, but it has been shown that it is impossible to combine simultaneously these criteria [22, 14]. Necessary tradeoffs are to be considered and much work leaves to be done in that direction.

Since Boolean functions are important primitives of such keystream generators, achieving the best possible tradeoffs is the main goal, which generally remains a difficult problem [6]. Recent results [7] have shown that if the existence of such functions can be proved, exhibiting some of them is very difficult, as soon as $n > 7$. The *nonlinearity*, *i.e.* the distance from the affine functions, is the criterion which presents the greatest difficulty. Recall that nonlinearity is of great importance in block cipher too, where the substitution-boxes must be as nonlinear as possible.

This paper presents significant results on the search of as good as possible Boolean functions and gives method of construction. 700 functions have been obtained for $n = 9$, which meet theoretical bound defined in [22]. Thus they are of great interest for cryptographic use all the more so since a survey of them show an unexpected additional tradeoff between the correlation-immunity order and the distribution of the (non zero) values of the existing correlations.

In Section 2, basic concepts and notation are given. Section 3 exposes the basic criteria for a Boolean function, related to the main existing attacks. Section 4 presents the corpus of the *idempotent* functions. As an important tool in Coding Theory, the use of the *idempotents* for finding Boolean functions with certain properties has been initiated by C. FONTAINE [7]. It will be presented and extended in this paper to find good cryptographically Boolean functions. In Section 5, results are exposed along with applications. The most significant one is that about 52000 *balanced* Boolean functions achieving the best possible tradeoff between *correlation-immunity* and nonlinearity are obtained, 700 of them meeting the Siegenthaler's bound.

2 Basic Concepts, Definitions and Notation

We will denote by $\mathbf{F}_q$ the finite field with q elements. A *Boolean function f of n variables* is a mapping from $\mathbf{F}_2^n$ into $\mathbf{F}_2$. We denote by $\mathcal{F}_n$ the set of such functions. We will use several representations for Boolean functions:

(1) Let $\mathcal{B} = (b_1, \ldots, b_n)$ be a basis of $\mathbf{F}_2^n$. The *Algebraic Normal Form (ANF) of f relatively to $\mathcal{B}$* is the polynomial $Q_{f,\mathcal{B}}$ of $\mathbf{F}_2[x_1, \ldots, x_n]/(x_1^2 - x_1, \ldots, x_n^2 - x_n)$ given by

$$Q_{f,\mathcal{B}}(x_1, \ldots, x_n) = \sum_{\substack{g = \sum_{i=1}^n g_i b_i \in \mathbf{F}_2^n \\ f(g) = 1}} \prod_{i=1}^{n} (1 + g_i + x_i).$$

This means that there are several ANFs for each function, depending on the basis we consider.

(2) If we identify the vector space $\mathbf{F}_2^n$ with the finite field $\mathbf{F}_{2^n}$, we can also represent f by a formal polynomial of the multiplicative algebra $\mathbf{F}_2[\mathbf{F}_{2^n}, \times]$:

$$f = \sum_{g \in \mathbf{F}_{2^n}^*} f(g)\ (g).$$

We recall that if f_1 and f_2 belong to this algebra, and $f_1 = \sum_{g \in \mathbf{F}_{2^n}^*} f_1(g)\ (g)$, $f_2 = \sum_{g \in \mathbf{F}_{2^n}^*} f_2(g)\ (g)$, then we have $f_1 + f_2 = \sum_{g \in \mathbf{F}_{2^n}^*} (f_1(g) + f_2(g))\ (g)$ and $f_1 f_2 = \sum_{g \in \mathbf{F}_{2^n}^*} (\sum_{hk=g} f_1(h) f_2(k))\ (g)$.

(3) We can also represent f by a polynomial of $\mathbf{F}_{2^n}[Z]/(Z^{2^n-1}-1)$, its *Mattson-Solomon (MS) polynomial*:

$$MS_f(Z)=\sum_{j=0}^{2^n-2} A_j Z^{2^n-1-j}, \quad A_j=\sum_{i=0}^{2^n-1} f(\alpha^i)\alpha^{ij}$$

where α is a primitive element of $\mathbf{F}_{2^n}$.

Example 1. For $n=3$, let α be a root of the primitive polynomial X^3+X+1. We identify $\mathbf{F}_2^n$ with the finite field $\mathbf{F}_{2^n}$.

(1) We consider two bases $\mathcal{B}_1=\{\alpha^0,\alpha^1,\alpha^2\}$ (the canonical basis) and $\mathcal{B}_2=\{\alpha^3,\alpha^6,\alpha^5\}$ of this field. Now let f be defined by $f(\alpha^0)=f(\alpha^1)=f(\alpha^2)=f(\alpha^4)=1$ and $f(0)=f(\alpha^3)=f(\alpha^5)=f(\alpha^6)=0$. We have $Q_{f,\mathcal{B}_1}(x_1,x_2,x_3)=x_1+x_2+x_3+x_2x_3$ and $Q_{f,\mathcal{B}_2}(x_1,x_2,x_3)=x_1x_2+x_2x_3+x_1x_3$.
(2) $f=(\alpha^0)+(\alpha^1)+(\alpha^2)+(\alpha^4)$.
(3) $MS_f(Z)=Z^3+Z^5+Z^6$.

Now let us introduce some definitions concerning Boolean functions:

- The *degree of f* is the global degree of the ANFs of f (they have all the same global degree). It is denoted by $deg(f)$.
- The *support of f*, denoted by $supp(f)$, is the set of the elements x such that $f(x)\neq 0$.
- the *distance between two Boolean functions f and g* is given by $d(f,g)=|supp(f+g)|$, where $+$ denotes the bitwise exclusive-or.
- f is said to be *balanced* if $|supp(f)|=2^{n-1}$.
- The *Walsh-Hadamard transform of f* is, for x in $\mathbf{F}_2^n$, the mapping $\hat{f}(x)=\sum_{y\in\mathbf{F}_2^n}(-1)^{x.y}f(y)$, where $x.y=\sum_{i=1}^{n}x_iy_i$ and the sum is evaluated over the real numbers. In fact, for a given Boolean function f, we will use only the Walsh-Hadamard transform of $(-1)^f=1+2f\ (mod\ 2)$ — the representation of f with ±1 — that is finally:

$$\widehat{\chi_f}(x)=\sum_{y\in\mathbf{F}_2^n}(-1)^{x.y+f(y)}.$$

Let $\mathcal{A}_n$ be the set of the *affine Boolean functions*, that is with degree at most 1. Since the set of the Boolean functions of $\mathcal{F}_n$ with degree at most r is the *r-th order binary Reed-Muller code of length 2^n and dimension $n+1$* [11, p. 373], $\mathcal{A}_n$ is in fact the first-order Reed-Muller code. The distance between a Boolean function f and $\mathcal{A}_n$ is given by $d(f,\mathcal{A}_n)=\min_{g\in\mathcal{A}_n}d(f,g)$ and it is called the *nonlinearity of f*.

We will denote by $NL(\mathcal{F}_n)$ the maximal nonlinearity for functions in $\mathcal{F}_n$. We recall in the following table the current knowledge about $NL(\mathcal{F}_n)$:

- even n : we know that $NL(\mathcal{F}_n)=2^{n-1}-2^{\frac{n}{2}-1}$ and the functions whose nonlinearity is equal to $NL(\mathcal{F}_n)$ are called *bent functions* [18].

– odd n : we only know that

n	knowledge about $NL(\mathcal{F}_n)$
$3, 5, 7$	$NL(\mathcal{F}_n) = 2^{n-1} - 2^{\frac{n-1}{2}}$ [2,15]
$9, 11, 13$	$2^{n-1} - 2^{\frac{n-1}{2}} \leq NL(\mathcal{F}_n) \leq 2^{n-1} - 2^{\lfloor \frac{n}{2} \rfloor - 1}$
odd ≥ 15	$2^{n-1} - 2^{\frac{n-1}{2}} < NL(\mathcal{F}_n) \leq 2^{n-1} - 2^{\lfloor \frac{n}{2} \rfloor - 1}$ [16,17,3]

Definition 1. *For a function f of $\mathcal{F}_n$, the* coset C_f of $\mathcal{A}_n$ generated by f *is the set $\{f + g, g \in \mathcal{A}_n\}$. The* weight distribution of C_f *is the polynomial of the form $WD(f) = \sum_{i=0}^{2^n} W_i X^i$, where W_i is the number of functions g in $\mathcal{A}_n$ such that $d(f, g) = i$. Remark that the nonlinearity of f corresponds to the smallest i such that $W_i \neq 0$. It is obvious that all the functions lying in C_f have the same nonlinearity, and the same degree.*

The Walsh-Hadamard transform of a function f and the weight distribution of C_f are related by the following theorem:

Theorem 1. *[11, p. 415] Let us consider the coset C_f of $\mathcal{A}_n$, and $WD(f)$ its weight distribution. Then we have $W_i \neq 0$ for all $i \in \left\{\frac{1}{2}\left[2^n \pm \widehat{\chi_f}(u)\right],\ u \in \mathbf{F}_2^n\right\}$.*

3 Cryptographic Criteria of Boolean Functions

Boolean functions are of great importance in the design of running-key generators for stream ciphers. These latter are a very important class of encryption algorithms. Encrypting binary digits of a plaintext one at a time, stream ciphers are widely used, being very fast and particularly well adapted to telecommunications applications (allowing *stream decryption*).

To be considered cryptographically secure, the sequence produced by the running-key generator must fulfill the following properties: the period must be large, as well the period complexity and good statistical properties must be achieved.

A well-known method for designing such a pseudo-random generator consists in using n linear feedback shift registers (LFSRs) with primitive feedback polynomials whose lengths are relatively prime. Additionally they are supposed not to be sparse (see [5,13]). Their output sequences $x_1, x_2, \ldots, x_n$ are taken as arguments of a Boolean function f of n variables whose output $f(x_1, \ldots, x_n)$ forms the running-key s (Fig. 1). The secret key of the system then consists of the initialization of all the LFSRs.

There exists a huge theoretical knowledge of such combining generators [19–21].

Let us review the different necessary criteria a Boolean function must fulfill to yield a cryptographically secure scheme, at least to resist known attacks.

3.1 Balancedness Criterion

In this case exactly half of the values of f are 0. The output of f must obviously be uniformly distributed to behave as an unpredictable variable.

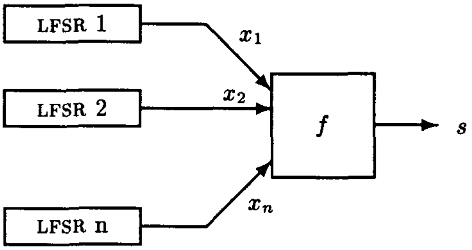

Fig. 1. Nonlinear combination generator

3.2 Nonlinearity Criterion

In order to break the linear properties of the LFSRs, to increase the period complexity of the output sequence s, and to avoid some well-known attacks [24, 25], the combining function f must be highly nonlinear. But as J.L. MASSEY pointed it out [12] the main difficulty is to quantify what cryptographers need and call "nonlinearity". The first (historically) approach is to consider it as the degree of the function.

But let us consider for example $f(x_1, x_2, x_3) = x_1 + x_2 + x_3 + x_1x_2x_3$. f is of degree 3, but it is easy to see that in fact f is very close to a linear function (whose order is 1). Then this approach of nonlinearity is not sufficient. The second approach then consists in defining the nonlinearity of a Boolean function as its distance from affine functions, using the theoretical tools of Coding Theory. From this point of view, a combining function must be as far as possible from any linear (or affine) function.

3.3 Correlation-Immunity Criterion

This criterion was defined in response of the correlation-attack of such schemes, developed by T. SIEGENTHALER [23] and which constitutes their main cryptanalytical approach. This cryptanalysis method aims to recover the different initializations separately - or at least an enough part of the secret key to greatly reduce the cost of the remaining exhaustive search.

The complexity of a brute-force attack (*i.e.* the number of trials) is $\prod_{i=1}^{n}(2^{L_i} - 1)$ where L_i is the length of the i-th LFSR. But if the combining function is such that there exists a correlation between the keystream s and the output sequence of the i-th LFSR *i.e.* if

$$P[x_i = f(x)] = \frac{1}{2}\left(1 \pm \frac{\widehat{\chi_f}(2^i)}{2^n}\right) \neq \frac{1}{2}$$

it is possible to try in a first step, all the $2^{L_i} - 1$ possible initializations of the i-th register only. The correct one will then be detected with a high probability. Thus the complexity of the brute-force attack is considerably reduced to $\prod_{j=1, j\neq i}^{n}(2^{L_j} - 1) + 2^{L_i} - 1$.

More generally, the *correlation-immunity order* is defined as follows:

Definition 2. *A Boolean function f of n variables is t-*th order correlation-immune *if, for any subset $T \subset \{1, 2, \ldots, n\}$ of size t, the probability distribution of its output is unaltered when the x_i are fixed, for $i \in T$. Moreover a balanced t-th order correlation-immune function is said to be t-*resilient.

In term of cryptanalysis, if the combining function is t-th order correlation-immune, any correlation attack must consider at least $(t + 1)$ different LFSRs simultaneously. Suppose for example that f is 2-resilient (let us say that there is a correlation between with LFSRs 1, 2 and 3). We then have:

$$P[x_1 + x_2 + x_3 = f(x)] = \frac{1}{2}\left(1 \pm \frac{\widehat{\chi_f}(7)}{2^n}\right)$$

Being 2-resilient, all entries of $\widehat{\chi_f}$ of weight less than 3 are zero. But according to the Parseval equation:

$$\sum_{u \in \mathbf{F}_2^n} \widehat{\chi_f}^2(u) = 2^{2n}$$

the value $\widehat{\chi_f}(7)$ corresponding to LFSRs 1, 2 and 3 will be higher, that is to say the correlation will be stronger.

The problem is how to exploit such an important bias without performing a too complex exhaustive search. Despite refinements of Siegenthaler's attack and attempts to give an answer to this problem, made by J. Golic and M. Mihaljevic [8–10], it is still an open problem.

A combining function should then fulfill all these criteria (high degree, balancedness, correlation-immunity and high nonlinearity). But some of these conditions are incompatible:

- T. Siegenthaler [22] showed that there is a necessary tradeoff between achieving high degree and high-correlation immunity.
- Since the nonlinearity is a global property of the functions, contrary to the correlation-immunity which is a local one, only a tradeoff can be achieved for these two criteria.
- When n (number of entries) is even, the functions of highest nonlinearity are the bent functions and it is a well-known fact that they cannot be balanced[11, p. 426]. Then balanced functions having the highest possible nonlinearity must be considered. Until now, finding such functions is a very difficult problem[6]. When n is odd, exhibiting functions of the highest nonlinearity is a hard problem in itself. Among the available candidates, balanced ones exist.

Finding some Boolean functions which achieve a good tradeoff between all these criteria is then of great importance in cryptography. Until now only theoretical, nonconstructive characterizations are known. We do not even know the value of the highest possible nonlinearity of a balanced Boolean function. It is also a very difficult problem to exhibit highly nonlinear functions meeting the Siegenthaler's bound for the t-resilient functions:

$$\deg(f) + t \leq n - 1 \qquad \text{unless } t = n - 1 \tag{1}$$

In the following part, a constructive method is given to obtain balanced Boolean functions with a high nonlinearity and a good correlation-immunity order, some of them meeting the Siegenthaler's bound.

4 The Corpus of the Idempotents

A first approach to the problem of finding balanced Boolean functions with a high nonlinearity could be to pick functions at random, in hope that they have the required properties. It is obviously not a suitable algorithm since the proportion of such functions seems to be small.

It is then necessary to restrict our investigation to a corpus of particular Boolean functions. Here we are interested in those which lie in cosets of $\mathcal{A}_n$ generated by *idempotents*. We choose *idempotents* for their role in the statement of important results in Coding Theory [1]; moreover, N. J. PATTERSON and D. H. WIEDEMANN have shown that $NL(\mathcal{F}_{15}) > 2^{15-1} - 2^{\frac{15-1}{2}}$ with in fact the help of *idempotents* [16,17].

We will first recall some definitions and properties of *idempotents*, and then give our numerical results.

4.1 Definitions

Definition 3. *Using the representation (2) introduced page 2, f is an* idempotent *if and only if $f^2 = f$, that is:*

$$\sum_{g\in \mathbf{F}_{2^n}^*} f(g)\,(g^2) = \sum_{g\in \mathbf{F}_{2^n}^*} f(g)\,(g).$$

This means that:

- $\forall g \in \mathbf{F}_{2^n}$, $f(g) = f(g^2)$. Then, if α is a primitive element of $\mathbf{F}_{2^n}$, $supp(f)$ is a union of conjugacy classes of some α^i, that is $supp(f) = \bigcup_{i\in I}\{\alpha^i, \alpha^{2i}, \dots, \alpha^{2^{n-1}i}\}$, where I is an arbitrary set of representative elements of the 2-cyclotomic cosets modulo $2^n - 1$ (we recall that the 2-cyclotomic coset generated by i is the set $\{i, 2i, \dots, 2^{n-1}i\}$, with the elements taken modulo $2^n - 1$).
- The coefficients of the MS polynomial of f belong to $\mathbf{F}_2$. Moreover, $A_j = A_k$ for all k in the 2-cyclotomic coset modulo $2^n - 1$ generated by j. Then we use a short MS polynomial: $\sum_{j\in Rep} A_j Z^{2^n-1-j}$, where $A_j \in \mathbf{F}_2$ and Rep denotes the set of all the representative elements of the 2-cyclotomic cosets modulo $2^n - 1$.
- The ANF of f expressed relatively to a normal basis — that is if $(x_1, \dots, x_n)$ corresponds to a basis of the form $(\gamma, \gamma^2, \dots, \gamma^{2^{n-1}})$ where γ is a primitive element of $\mathbf{F}_{2^n}$ — remains invariant if all subscripts are permuted with a circular shift. So we can keep only one term per class of shifts to express it as a short ANF.

Example 2. We take $n = 3$, and we consider the idempotent f whose support is $\{\alpha^0, \alpha, \alpha^2, \alpha^4\}$, where α is a root of the primitive irreducible polynomial $X^3 + X + 1$:

- its MS polynomial is $Z^3 + Z^5 + Z^6$. The nonzero coefficients are A_1, A_2, A_4. But $1, 2, 4$ are all in the cyclotomic coset containing 1. Then the short MS polynomial of f has only one nonzero coefficient — A_1 — and is equal to Z^6.
- the ANF expressed in the normal basis $\{\alpha^3, \alpha^6, \alpha^5\}$ is $x_1x_2 + x_2x_3 + x_1x_3$. These terms are all in the same shift class, and then the short ANF is x_1x_2.

Then the corpus of idempotent functions has the useful property to be short to represent. It is a gain of space in computation, and then a gain of memory. Another important point is that the short MS polynomial has its coefficients into $\mathbf{F}_2$: it is then a gain of time for computations since idempotents can be stored as a computer word (shorter than 2^n which is the number of the values taken by f).

4.2 Results

As explained at the beginning of this section, our aim is to study cosets of $\mathcal{A}_n$ generated by idempotents. This means that for a given idempotent function f in $\mathcal{F}_n$ we look at the coset $C_f = \{f + g, g \in \mathcal{A}_n\}$ and the weight distribution $WD(f) = \sum_{i=0}^{2^n} W_i X^i$. The nonlinearity of any function in C_f is the smallest i such that $W_i \neq 0$, and the number of balanced functions in C_f is given by $W_{2^{n-1}}$. It is obvious that all the Boolean functions of C_f have the same degree, and since they all differ from each other only on their affine terms, they are said to be *equivalent* [11, p. 416]. We will say that the weight distribution $WD(f)$ is *maximal* if the nonlinearity of f is equal to $NL(\mathcal{F}_n)$.

We will first resume our results for $n = 5, 6, 7, 8$ and then give in a table the number of highest nonlinear Boolean functions we obtained. Then we will expose our results for $n = 9$, the first case when $NL(\mathcal{F}_n)$ is unknown.

For our examples, we will use the short ANF of the idempotent functions, relatively to a normal basis $\gamma, \gamma^2, \ldots, \gamma^{2^{n-1}}$. α will be a root of the primitive polynomial $P_n(X)$ of degree n and γ is the $pow_n{}^{th}$ power of α. Here, we take $P_5(X) = X^5 + X^2 + 1, pow_5 = 3$, $P_6(X) = X^6 + X + 1, pow_6 = 5$, $P_7(X) = X^7 + X + 1, pow_7 = 13$, $P_8(X) = X^8 + X^4 + X^3 + X^2 + 1, pow_8 = 11$, $P_9(X) = X^9 + X^5 + 1, pow_9 = 13$.

When $NL(\mathcal{F}_n)$ is known

$n = 5, 7$ We have computed all the idempotent functions with the highest nonlinearity ($NL(\mathcal{F}_5) = 12, NL(\mathcal{F}_7) = 56$). They are of degree $2, 3$ for $n = 5$ and $2, 3, 4, 5, 6$ for $n = 7$. Moreover, the cosets generated by them contain a lot of balanced functions. Another important point is that we obtained four distinct maximal weight distributions for $n = 7$.

We give some examples of the highest nonlinear idempotent balanced functions we have found (using their short ANF):

Example 3. $n = 5$: x_1x_3 and $x_1x_3 + x_1x_2x_3 + x_1x_2x_4$
$n = 7$:

- $x_1x_2 + x_1x_3 + x_1x_4$
- $x_1x_3 + x_1x_2x_3 + x_1x_2x_4 + x_1x_2x_5 + x_1x_3x_5$
- $x_1x_3 + x_1x_2x_5 + x_1x_3x_5 + x_1x_2x_3x_4 + x_1x_2x_3x_6 + x_1x_2x_4x_5 + x_1x_2x_4x_6$
- $x_1 + x_1x_2 + x_1x_2x_3 + x_1x_2x_5 + x_1x_2x_6 + x_1x_2x_3x_5 + x_1x_2x_3x_6 + x_1x_2x_4x_6 + x_1x_2x_3x_4x_6$
- $x_1x_2 + x_1x_4 + x_1x_2x_3x_4 + x_1x_2x_3x_4x_5 + x_1x_2x_3x_4x_5x_6$

$n = 6, 8$ We obtained all the idempotent functions with the highest nonlinearity, that is the bent functions ($NL(\mathcal{F}_6) = 28, NL(\mathcal{F}_8) = 120$). We have obtained all the possible degrees for these functions: $2, 3$ for $n = 6$ and $2, 3, 4$ for $n = 8$. But since they are bent, they can not be balanced.

In [6], H. DOBBERTIN give for even n a result on the highest nonlinearity for balanced Boolean functions in term of the values of the Walsh-Hadamard transform. In term of nonlinearity (using theorem 1) and in our case, this result is the following: for $n = 6$, the highest nonlinearity for balanced functions is 26, and for $n = 8$, it is 116 or 118.

For $n = 6$ we have computed all the idempotent functions, and we have found balanced functions with nonlinearity 26. All the highest nonlinear balanced functions we have found are of degree 5 .

For $n = 8$, we did not compute all the idempotent functions. But we applied the following algorithm: we fix the number of nonzero coefficients nb in the short MS polynomial, and then look at all the cosets generated by the idempotents corresponding to short MS polynomial with nb nonzero coefficients. Fixing nb to $1, \ldots, 8$ we obtained a lot of balanced functions with nonlinearity 116, but no balanced one with nonlinearity 118. We then focused on the nonlinearity 118, looking at all the possible short MS polynomials, but we did not find any balanced function with this nonlinearity.

We give some examples of the idempotent bent functions we have found (using their short ANF):

Example 4. $n = 6$: $x_1 + x_1x_4$ and $x_1 + x_1x_4 + x_1x_2x_3 + x_1x_3x_5$
$n = 8$:

- x_1x_5
- $x_1x_2 + x_1x_5 + x_2x_5 + x_1x_2x_3 + x_1x_2x_4 + x_1x_2x_7 + x_1x_3x_6$
- $x_1x_2 + x_1x_3 + x_1x_4 + x_1x_2x_3 + x_1x_2x_5 + x_1x_2x_7 + x_1x_3x_5 + x_1x_2x_3x_4 + x_1x_2x_4x_5 + x_1x_2x_4x_6$

The balanced functions we obtained with the highest nonlinearity are not idempotents. Their ANFs are too long to be written here.

We now give in the following table the number of highest nonlinear Boolean functions we have found for $n = 5, 6, 7, 8$:

	highest nonlinear functions			highest nonlinear balanced functions		
n	$NL(\mathcal{F}_n)$	nb. of f.	up to equiv.	max. nonlin. for bal. f.	nb. of f.	up to equiv.
5	12	576	9	12	288	9
6	28	1536	12	26	832	40
7	56	754432	2947	56	259420	9
8	120	1933312	3776	116 (it is perhaps 118)	328480	4737

The First Case when $NL(\mathcal{F}_n)$ is Not Known: $n = 9$ In this case, we do not know the real value of $NL(\mathcal{F}_9)$, we only know that $NL(\mathcal{F}_9) \geq 240$. Our aim was to obtain Boolean functions with nonlinearity at least 240, and perhaps higher. We decided to restrict the set of the functions to look at, fixing the number of nonzero coefficients in the short MS polynomial, as previously explained in the case $n = 8$ for the search of balanced functions. We looked at the short MS polynomials with at most 11 nonzero coefficients, and we have found 83 new weight distributions corresponding to the nonlinearity 240. We give in the following table a sample of them, showing that the corresponding cosets contain a lot of balanced functions.

	240	242	244	246	248	250	252	254	256
i	272	270	268	266	264	262	260	258	
	117		210		72		46		134
	126		190		76		66		108
	138		172		64		84		108
	144		163		58		93		108
	145		156		72		84		110
W_i	162		136		40		120		108
	162		156				100		188
	180		120				136		152
	211				180				242
	217				156				278
	220				144				296
	226				120				332

In total, we have found 1169812480 functions with nonlinearity 240, that is 1142395 functions up to equivalence; 549339200 of these functions are balanced, that is 1142390 up to equivalence (this means that there are balanced functions in almost any coset with nonlinearity 240 we have found). They are of degree $2, 3, 4, 5, 6, 7$.

5 Applications

In general, the functions we obtained with the maximal nonlinearity are not correlation immune. It is actually very difficult to find a balanced function which has a high correlation-immunity order and a high nonlinearity.

In 1991, P. Camion, C. Carlet, P. Charpin and N. Sendrier have presented in [4] a construction of t-resilient functions: let f be a t-resilient function of $\mathcal{F}_n$, and g be the function of $\mathcal{F}_{n+1}$ defined by (the ANFs are expressed in the canonical basis):

$$g(x_1, \dots, x_n, x_{n+1}) = f(x_1, \dots, x_n) + x_{n+1}.$$

Then g is $(t+1)$-resilient.

We will apply it to our balanced highly nonlinear functions in order to construct balanced highly nonlinear functions with a good Correlation Immunity order.

Theorem 2. *The degree and nonlinearity of g can be deduced from the ones of f by*

$$\begin{aligned} deg(g) &= deg(f) \\ d(g, \mathcal{A}_{n+1}) &= 2d(f, \mathcal{A}_n) \end{aligned}$$

Proof. It is obvious that f and g have the same degree. We will now prove the result on the nonlinearity.

We have $d(g, \mathcal{A}_{n+1}) = \min\{d(g,l), l \in \mathcal{A}_{n+1}\}$. Let l be an affine function of $\mathcal{A}_{n+1}$: it can be written as $l(x_1, \dots, x_{n+1}) = l'(x_1, \dots, x_n) + \varepsilon x_{n+1}$ with l' in $\mathcal{A}_n$ and ε in $\mathbf{F}_2$.

Let V_f denote the vector corresponding to the values of f. We also introduce V_g, V_l and $V_{l'}$. We have $V_g = (V_f|\overline{V_f})$, where $\overline{V_f}$ denotes the complement of V_f, and $V_l = (V_{l'}|\overline{V_{l'}})$. And since the distance between two Boolean functions g and l is the sum of the elements of the vector $V_g + V_l$, we have:

If $\varepsilon = 1$, then $d(g,l) = d(f,l') + d(\overline{f}, \overline{l'}) = 2\ d(f,l')$.

If $\varepsilon = 0$, then $d(g,l) = d(f,l') + d(\overline{f}, l') = d(f,l') + d(f, \overline{l'}) = 2^n$.

And then $d(g, \mathcal{A}_{n+1}) = \min_{l' \in \mathcal{A}_n}\{2\ d(f,l'),\ 2^n\} = \min_{l' \in \mathcal{A}_n}\{2\ d(f,l')\} = 2\ d(f, \mathcal{A}_n)$.

Then if we use a 0-correlation-immune balanced function of $\mathcal{F}_n$ with nonlinearity NL_f, we obtain by iterating t times this construction a t-resilient function of $\mathcal{F}_{n+t}$ with nonlinearity $2^t NL_f$.

Example 5. We take the idempotent function of $\mathcal{F}_7$ whose short ANF relatively to the normal basis $(\alpha^{13}, \dots, \alpha^{70})$ (where α is a root of the primitive polynomial $X^7 + X + 1$) is

$f = x_2 + x_1x_7 + x_1x_2x_3 + x_1x_5x_7 + x_2x_5x_6 + x_2x_5x_7 + x_2x_6x_7 + x_1x_2x_3x_4 + x_1x_2x_3x_6 + x_1x_3x_4x_7 + x_2x_5x_6x_7 + x_1x_3x_5x_6x_7 + x_1x_2x_3x_4x_5x_6$.

This function is balanced of degree 6 and has the highest nonlinearity, 56. We apply the construction twice, and then obtain a 2-resilient function g of $\mathcal{F}_9$ with degree 6 and nonlinearity $2^2 * 56 = 224$. The inequality given in Equation 1 is here $6 + 2 \leq 9 - 1$ and is then an equality. We have optimized the degree regarding to the resilience order.

This function g is representative of all the functions we obtained, they are particularly well-suited for implementation in stream ciphers as the most secure cryptographic primitive of this kind, since they optimally combine all three necessary criteria, precedently presented. One more important point in this kind of application is a very good behavior (better than expected) toward the correlation-immunity. Being 2-resilient, the different nonzero values of $\hat{f}(w)$, yet more important than for 1-resilient functions (due to the Parseval's equation), are however very well distributed. It can be considered as an additional tradeoff between the correlation-immunity order and the most unfavorable existing correlation values. Cryptographically speaking, not only we are forced to consider 3 LFSRs simultaneously in a correlation attack (which is generally intractable if the sum of their length exceeds about 80) but also the correlation values will be weak enough to offer far more resistance since it will oblige the cryptanalyst to consider a generally too long keystream s to recover the secret elements with acceptable probability of success. From this point of view, these obtained functions are cryptographically secure regarding to known attacks.

And since we have 700 functions with the same characteristics as f (154 functions up to equivalence), we can construct 700 functions with the same characteristics as g.

Moreover we have 51744 functions of 7 variables which are balanced, of degree 5 and with nonlinearity 56 (they are 1176 up to equivalence). They enable us to construct 51744 functions of 9 variables which are 2-resilient, of degree 5 and nonlinearity 224.

6 Conclusion

The need for the most possible secure cryptographic primitives in cipher systems is of great importance. In case of stream ciphers most of the reliability and security lies in the Boolean functions which must combine different criteria (balancedness, correlation-immunity, nonlinearity) to ensure resistance to known attacks.

Precedent studies showed that only tradeoffs could be envisaged, these criteria being impossible to be obtained simultaneously. Until now, only existence results were known and no effective construction method were given, which could have been used to exhibit best as possible Boolean functions. In this paper, the use of a particular of Boolean functions, called idempotents, has been widely used to give such a constructive method. 700 Boolean functions for $n = 9$ have been obtained. They not only present the best possible tradeoff between the desired criteria but also meet the theoretical Siegenthaler's bound. An additional tradeoff has been observed, between the correlation-immunity order and the distribution of the nonzero values of the existing correlations. All that make these functions particularly well-suited for implementation in stream ciphers thus resisting all the known attacks.

Acknowledgements

We are very grateful to P. Charpin for her enriching discussions and suggestions. We wish to thank A. Canteaut for her motivating discussions on Boolean functions and N. Sendrier for valuable improvements concerning the implementation.

References

1. D. Augot and N. Sendrier. Idempotents and the BCH bound. *IEEE Transactions on Information Theory*, 40(1):204–207, 1994.
2. E. R. Berlekamp and L. R. Welch. Weight distributions of the cosets of the (32,6) Reed-Muller code. *IEEE Transactions on Information Theory*, 18(1):203–207, January 1972.
3. R. A. Brualdi, N. Cai, and V. S. Pless. Orphan structure of the first-order Reed-Muller codes. *Discrete Mathematics*, (102):239–247, 1992.
4. P. Camion, C. Carlet, P. Charpin, and N. Sendrier. On correlation-immune functions. In J. Feigenbaum, editor, *Advances in Cryptology - CRYPTO'91*, number 576 in Lecture Notes in Computer Science, pages 86–100. Springer-Verlag, 1992.
5. V. Chepyzhov and B. Smeets. On a fast correlation attack on certain stream ciphers. In D.W. Davis, editor, *Advances in Cryptology - EUROCRYPT'91*, number 547 in Lecture Notes in Computer Science, pages 176–185. Springer-Verlag, 1991.
6. H. Dobbertin. Construction of bent functions and balanced Boolean functions with high nonlinearity. In B. Preneel, editor, *Fast Software Encryption*, number 1008 in Lecture Notes in Computer Science, pages 61–74. Springer-Verlag, 1994.
7. C. Fontaine. The nonlinearity of a class of Boolean functions with short representation. In J. Přibyl, editor, *PRAGOCRYPT'96*, pages 129–144. CTU Publishing House, 1996.
8. J. Golic and M. Mihaljevic. A fast iterative algorithm for a shift register initial state reconstruction given the noisy output sequence. In *Advances in Cryptology - AUSCRYPT'90*, number 453 in Lecture Notes in Computer Science, pages 165–175. Springer-Verlag, 1990.
9. J. Golic and M. Mihaljevic. A generalized correlation attack on a class of stream ciphers based on the Levenshtein distance. *Journal of Cryptology*, (3):201–212, 1991.
10. J. Golic and M. Mihaljevic. Convergence of a Bayesian iterative error-correction procedure on a noisy shift register sequence. In *Advances in Cryptology - EUROCRYPT'92*, number 658 in Lecture Notes in Computer Science, pages 124–137. Springer-Verlag, 1993.
11. F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error Correcting Codes*. North-Holland, 1992.
12. J. L. Massey. Some applications of Coding Theory in Cryptography. In P.G. Farrel, editor, *Codes and Cyphers: Cryptography and Coding IV*, pages 33–47, 1995.
13. W. Meier and O. Staffelbach. Fast correlation attacks on stream ciphers. In C.G. Günther, editor, *Advances in Cryptology - EUROCRYPT'88*, number 330 in Lecture Notes in Computer Science, pages 301–314. Springer-Verlag, 1988.
14. W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. In J.Vandewalle J.-J.Quisquater, editor, *Advances in Cryptology - EUROCRYPT'89*, number 434 in Lecture Notes in Computer Science, pages 549–562. Springer-Verlag, 1990.
15. J. Mykkeltveit. The covering radius of the [128,8] Reed-Muller code is 56. *IEEE Transactions on Information Theory*, 26(3):358–362, May 1980.
16. N.J. Patterson and D. H. Wiedemann. The covering radius of the $[2^{15}, 16]$ Reed-Muller code is at least 16276. *IEEE Transactions on Information Theory*, 29(3):354–356, May 1983.
17. N.J. Patterson and D. H. Wiedemann. Correction to [16]. *IEEE Transactions on Information Theory*, 36(2):443, March 1990.

18. O.S. Rothaus. On bent functions. *Journal of Combinatorial Theory*, (20):300–305, 1976.
19. R. A. Rueppel. *Analysis and Design of Stream Ciphers.* Springer-Verlag, 1986.
20. R. A. Rueppel. *Contemporary Cryptology: the science of Information Integrity*, chapter Stream Ciphers, pages 65–134. IEEE Press, 1992.
21. B. Schneier. *Applied Cryptography.* Wiley, second edition, 1996.
22. T. Siegenthaler. Correlation-immunity of nonlinear combining functions for cryptographic applications. *IEEE Transactions on Information Theory*, 30(5):776–780, September 1984.
23. T. Siegenthaler. Decrypting a class of stream ciphers using ciphertext only. *IEEE Transactions on Computers*, 34(1):81–84, January 1985.
24. K. Zeng and M. Huang. On the linear syndrome method in cryptanalysis. In S. Goldwasser, editor, *Advances in Cryptology - CRYPTO'88*, number 403 in Lecture Notes in Computer Science, pages 469–478. Springer-Verlag, 1990.
25. K. Zeng, C. H. Yang, and T. R. Rao. An improved linear syndrome algorithm in cryptanalysis with applications. In A.J. Menezes S.A. Vanstone, editor, *Advances in Cryptology - CRYPTO'90*, number 537 in Lecture Notes in Computer Science, pages 34–47, 1991.