IMPROVING THE LOWER BOUND ON THE HIGHER ORDER NONLINEARITY OF BOOLEAN FUNCTIONS WITH PRESCRIBED ALGEBRAIC IMMUNITY

SIHEM MESNAGER

Abstract. The recent algebraic attacks have received a lot of attention in cryptographic literature. The algebraic immunity of a Boolean function quantifies its resistance to the standard algebraic attacks of the pseudo-random generators using it as a nonlinear filtering or combining function. Very few results have been found concerning its relation with the other cryptographic parameters or with the r-th order nonlinearity. As recalled by Carlet at Crypto'06, many papers have illustrated the importance of the rth-order nonlinearity profile (which includes the first-order nonlinearity). The role of this parameter relatively to the currently known attacks has been also shown for block ciphers. Recently, two lower bounds involving the algebraic immunity on the rth-order nonlinearity have been shown by Carlet et *al.* None of them improves upon the other one in all situations. In this paper, we prove a new lower bound on the rth-order nonlinearity profile of Boolean functions, given their algebraic immunity, that improves significantly upon one of these lower bounds for all orders and upon the other one for low orders.

Keywords. stream cipher, block cipher, algebraic attack, Boolean function, algebraic immunity, algebraic degree, higher order nonlinearity, annihilator.

Introduction

Symmetric cryptosystems are commonly used for encrypting and decrypting owing to their efficiency. A classical model of symmetric cryptosystem are stream ciphers. They are composed of one or several Linear Feedback Shift Register (LFSR) combined or filtered by a Boolean function. These cryptosystems have been the objects of a lot of cryptanalyses and several design criteria have been proposed concerning the filtering or combining functions. A survey on this topic can be found in [2]. The most basic requirement concerning Boolean functions used in stream ciphers is to be of algebraic degree as high as possible. We recall that the algebraic degree of a Boolean function f is the degree of its unique representation as a multivariate polynomial over $\mathbb{F}_2$, that we denote by $\deg(f)$.

Date: August 3, 2007.

MAATICAH, Department of Mathematics, University of Paris 8.
email : hachai@math.jussieu.fr.

Recently, new kinds of attacks drawn from an original idea of Shannon [15] has emerged; these attacks are called *algebraic attacks* and *fast algebraic attacks* [6, 12]. They proceed by modelling the problem of recovering the secret key by means of an over-defined system of multivariate nonlinear equations of algebraic degree at most $\deg(f)$. The core of algebraic attacks is to find out low degree Boolean functions $g \neq 0$ and h such that $fg = h$. Meier, Pasalic and Carlet [13] have shown that it is equivalent to the existence of low degree *annihilators* of f, that is, of n-variable Boolean functions g such that $fg = 0$ or $(1 \oplus f)g = 0$. The minimum degree of such g is called the *algebraic immunity* of f, and that we denote by $AI(f)$. It must be as high as possible (the optimum value of $AI(f)$ being equal to $\lceil \frac{n}{2} \rceil$). Fast algebraic attacks proceed in a different way but having a high algebraic immunity is not only a necessary condition for a resistance to standard algebraic attacks but also for a resistance to fast algebraic attacks. Few authors have investigated the relation between the algebraic immunity of Boolean function and other cryptographic parameters. The first result found concerns the Hamming weight $\mathrm{wt}(f)$ of f, that is, the number of 1 in its truth table. Carlet, Dalai and Gupta [3] shown that : $\sum_{i=0}^{AI(f)-1} \binom{n}{i} \leq \mathrm{wt}(f) \leq \sum_{i=0}^{n-AI(f)} \binom{n}{i}$. It implies in particular that a Boolean function with optimum nonlinearity is necessarily balanced in odd dimension, that is outputs 1 with probability $\frac{1}{2}$. Another important cryptographic parameter is the nonlinearity of a Boolean function f, that we denote by $nl(f)$, which equals the number of bits to change in the truth table of f to get an affine Boolean function (that is, a Boolean function of algebraic degree 1). The first lower bound on the nonlinearity of f involving the algebraic immunity was given in [3]. Lobanov [11] improved further upon this lower bound and proved that : $nl(f) \geq 2\sum_{i=0}^{AI(f)-2} \binom{n-1}{i}$ for every n-variable Boolean function f. Moreover, he has exhibited a family of Boolean function achieving the equality $nl(f) = 2\sum_{i=0}^{AI(f)-2} \binom{n-1}{i}$.

Carlet introduced in [1] the term of *nonlinearity profile* of Boolean functions, which is the sequence whose rth-order term equals the rth-order nonlinearity of the function that we denote by $nl_r(f)$, and that is the minimum distance between f and all n-variable Boolean functions of algebraic degrees at most r. This parameter extends the standard (first-order) nonlinearity $nl(f)$ of a Boolean function f. Several papers [5, 8, 9, 10, 14] have shown the role played by this parameter in relation to some cryptanalyses (note that contrary to the (first-order) nonlinearity, it must have low value for allowing the attacks to be realistic). Computing theoretically and algorithmicly the rth-order nonlinearity of an n-variable Boolean function is a hard task for $r > 1$. Therefore the knowledge of upper and lowers bounds for the rth-order nonlinearity on a particular class of Boolean functions is important.

Lobanov's result has been extended to the rth-order nonlinearity $nl_r(f)$ of an n-variable Boolean function f in two different lower bounds [1, 3]. None of the two lower bounds improves upon the other one in all situations.

Basically, these lower bounds say that the rth-order nonlinearity of an n-variable Boolean function f of algebraic immunity k is greater than or equal to the maximum value between $\sum_{i=0}^{k-r-1}\binom{n}{i}$ and $2\sum_{i=0}^{k-r-1}\binom{n-r}{i}$. In this paper, we improve further the lower bound of [3] for all orders and the lower bound of [1] for low orders (which are the most important from a practical point of view) : for every n-variable Boolean function f, we prove that the rth-rder nonlinearity $nl_r(f)$ of a n-variable Boolean function of algebraic immunity k is greater than or equal to $\sum_{i=0}^{k-r-1}\binom{n}{i}+\sum_{i=k-2r}^{k-r-1}\binom{n-r}{i}$.

The paper is organized as follows. Firstly, we begin with fixing our main notation in Section 1. Secondly, we study in Section 2 the dimension of the annihilators with prescribed algebraic degrees of Boolean functions with given algebraic degrees. The results of this Section are crucial to obtain in Section 3 a new lower bound on the rth-order nonlinearity of a Boolean function of given algebraic immunity (Theorem 10).

1. Preliminaries

Let n be any positive integer. In this paper, we shall denote by $\mathcal{B}_n$ the set of all n-variable Boolean functions over $\mathbb{F}_2^n$. Any n-variable Boolean function f (that is an application from $\mathbb{F}_2^n$ to $\mathbb{F}_2$) admits a unique *algebraic normal form* (ANF), that is, a representation as a multivariate polynomial over $\mathbb{F}_2$

$$f(x_1,\ldots,x_n)=\bigoplus_{I\subseteq\{1,\ldots,n\}} a_I\prod_{i\in I}x_i,$$

where the a_I's are in $\mathbb{F}_2$. The terms $\prod_{i\in I}x_i$ are called *monomials*. The *algebraic degree* $\deg(f)$ of a Boolean function f equals the maximum degree of those monomials whose coefficients are nonzero in its algebraic normal form. A slightly different form for the algebraic normal form is $f(x)=\bigoplus_{u\in\mathbb{F}_2^n}a_u x^u$, where $a_u\in\mathbb{F}_2$ and where $x^u=\prod_{i=1}^n x_i^{u_i}$. Then $\deg(f)$ equals $\max_{a_u\neq 0}\mathrm{wt}(u)$, where $\mathrm{wt}(u)$ denotes the Hamming weight of u, that is, $\mathrm{wt}(u)=|\{i=1,\ldots,n\mid u_i=1\}|$. Given a positive integer r, we make an abuse of notation and denote by $\mathrm{RM}(r,n)$ the set of all n-variable Boolean functions of algebraic degrees at most r, that is, the so-called r-th order Reed-Muller code of length 2^n. We recall that $\mathrm{RM}(r,n)$ is a vector subspace over $\mathbb{F}_2$ of dimension $\sum_{i=0}^r\binom{n}{i}$.

The Hamming weight $\mathrm{wt}(f)$ of a Boolean function is the size of its support $\{x\in\mathbb{F}_2^n\mid f(x)=1\}$ that we denote by $\mathrm{supp}(f)$. The Hamming distance between two n-variable Boolean functions is the Hamming weight of $f\oplus g$, that is $\mathrm{dist}(f,g)=|\{x\in\mathbb{F}_2^n\mid f(x)\neq g(x)\}|$.

Definition 1 (rth-order nonlinearity)**.** Let f be an n-variable Boolean function. Let r be a positive integer such that $r\leq n$. The *r-th order nonlinearity* of f is the minimum Hamming distance between f and all n-variable Boolean functions from $\mathrm{RM}(r,n)$. We shall denote the r-th order nonlinearity of f by $nl_r(f)$.

The first-order nonlinearity of f is simply called the nonlinearity of f and is denoted by $nl(f)$ (instead of $nl_1(f)$). Clearly we have $nl_r(f) = 0$ if and only if f has degree at most r. So, the knowledge of the nonlinearity profile (i.e. of all the nonlinearities of orders $r \geq 1$) of a Boolean function includes the knowledge of its algebraic degree. It is in fact a much more complete cryptographic parameter than are the single (first-order) nonlinearity and the algebraic degree. Very little is known on $nl_r(f)$. The best known upper bound on $nl_r(f)$ has asymptotic version [4] :

$$nl_r(f) \leq 2^{n-1} - \frac{\sqrt{15}}{2}(1+\sqrt{2})^{r-2}2^{\frac{n}{2}} + O(n^{r-2})$$

for every n-variable Boolean functions f.

The algebraic immunity [13] of a Boolean function f quantifies the resistance to the standard algebraic attack of the pseudo-random generators using it as a nonlinear function. It is defined as follows.

Definition 2 (Algebraic immunity)**.** Let f be an n-variable Boolean function. An n-variable Boolean function g is said to be an *annihilator* of f if the product $f \cdot g$ is null (that is, the support of g is included in the support of $1 \oplus f$). We denote by $An(g)$ the vector space of all annihilators of g. The algebraic immunity of f is the minimum algebraic degree of all the nonzero annihilators of f or of $f \oplus 1$. The *algebraic immunity* of f, is denoted by $AI(f)$.

Clearly, the algebraic immunity of a Boolean function f is less than or equal to its algebraic degree since $1 \oplus f$ is an annihilator of f. As shown in [12], we have $AI(f) \leq \lceil \frac{n}{2} \rceil$. It was shown in [7] that the Hamming weight of a Boolean function f with given algebraic immunity satisfies : $\sum_{i=0}^{AI(f)-1} \binom{n}{i} \leq \mathrm{wt}(f) \leq \sum_{i=0}^{n-AI(f)} \binom{n}{i}$. In particular, if n is odd and f has optimum algebraic immunity then f is balanced.

2. Some results on the dimension of the vector space of prescribed degree annihilators of a Boolean function

An important parameter for evaluating the complexity of algebraic attacks on the systems using a given Boolean function is the number of linearly independent low degree annihilators of this Boolean function g and of the function $g \oplus 1$. We shall see in the next Section that it plays also an important role in relation to the r-th order nonlinearity.

Definition 3. Let g be a Boolean function and let k be a positive integer. We denote by $An_k(g)$ the vector space of those annihilators of degrees at most k of g and by $d_{k,g}$ the dimension of $An_k(g)$.

The dimension $d_{k,g}$ is an affine invariant, that is, we have $d_{k,g} = d_{k,g \circ A}$ for every affine automorphism A of $\mathbb{F}_2^n$ (this comes from the affine invariance of the algebraic degree and the fact that p is an annihilator of g if and only

if $p \circ A$ is an annihilator of $g \circ A$). Little is known on the behavior of $d_{k,g}$. Carlet [1] proved the following upper bound on $d_{k,g}$.

Proposition 1. *For every n-variable Boolean function g of algebraic degree at most r, we have $d_{k,g} \leq \sum_{i=0}^{k} \binom{n}{i} - \sum_{i=0}^{k} \binom{n-r}{i}$. This upper bound is achieved by the indicators of an $(n-r)$-dimensional affine subspace of $\mathbb{F}_2^n$ for which the dimension $d_{k,g}$ is exactly equal to $\sum_{i=0}^{k} \binom{n}{i} - \sum_{i=0}^{k} \binom{n-r}{i}$*

We can derive from this upper bound a lower bound on $d_{k,g}$. Let us introduce some notation before. For every n-variable Boolean function g and every positive integer k, we denote by $Mul_k(g)$ the vector space of all n-variable Boolean functions p that can be written as $p = gh$ where h is of algebraic degree at most k. There exists a simple relation between $d_{k,g}$ and $\dim Mul_k(g)$.

Lemma 2. *Let g be an n-variable Boolean function of algebraic degree r. Let k be any positive integer less than n. Then* $\dim Mul_k(g) = \sum_{i=0}^{k} \binom{n}{i} - d_{k,g}$.

Proof. Let ϕ_g be the linear map from $\mathrm{RM}(k,n)$ to $Mul_k(g)$ which maps h to gh. This linear map is onto and its kernel equals $An_k(g)$. Thus, by applying the rank theorem to ϕ_g, one gets that $\dim \mathrm{RM}(k,n) = \sum_{i=0}^{k} \binom{n}{i} = \dim \mathrm{Im}(\phi_g) + \dim \ker(\phi_g) = \dim Mul_k(g) + d_{k,g}$. □

The upper bound of [1] (that we have recalled in Proposition 1) and Lemma 2 lead us to a lower bound on $d_{k,g}$ achieved by the complements of the indicators of affine subspaces of $\mathbb{F}_2^n$. More precisely,

Proposition 3. *Let g be an n-variable Boolean function of algebraic degree at most r. Then, for every positive integer k, one has $d_{k,g} \geq \sum_{i=0}^{k-r} \binom{n-r}{i}$. If g is the complement of the indicator of an $(n-r)$-dimensional affine subspace of $\mathbb{F}_2^n$ then $d_{k,g} = \sum_{i=0}^{k-r} \binom{n-r}{i}$.*

Proof. Let g be an n-variable Boolean function of algebraic degree at most r. We can assume that $k \geq r$ (otherwise the lower bound is trivial). Take $h \in An_r(g)$. We have $d_{k,h} \leq \sum_{i=0}^{k-r} \binom{n}{i} - \sum_{i=0}^{k-r} \binom{n-r}{i}$ by Proposition 1. Now, according to Lemma 2, $\dim Mul_{k-r}(h) = \sum_{i=0}^{k-r} \binom{n}{i} - d_{k-r,h}$. Thus $\dim Mul_{k-r}(h) \geq \sum_{i=0}^{k-r} \binom{n-r}{i}$. Moreover, we have the inclusion $Mul_{k-r}(h) \subseteq An_k(g)$. Therefore, it holds that $d_{k,g} \geq \dim Mul_{k-r}(h) \geq \sum_{i=0}^{k-r} \binom{n-r}{i}$. This latter inequality becomes an equality whenever g is the complement of an $(n-r)$-dimensional affine subspaces of $\mathbb{F}_2^n$ because it has been shown in [1] that $d_{k,g}$ is equal to $\sum_{i=0}^{k-r} \binom{n-r}{i}$ for such Boolean functions. □

We prove a result that we shall use to improve the lower bound of [1, 3]. To this aim, we need to introduce some additional notation. We shall use the partial ordering $\preceq$ on $\mathbb{F}_2^n$ defined as follows :

$$u, v \in \mathbb{F}_2^n, \quad (u \preceq v) \iff (\mathrm{supp}(u) \subset \mathrm{supp}(v))$$

Given an element u of $\mathbb{F}_2^n$, we call the subset $\{i \in \{1, \ldots, n\} \mid u_i = 1\}$ the support of u, and we denote it by $\mathrm{supp}(u)$. The Hamming weight of

u, denoted by $\mathrm{wt}(u)$, is the cardinality of $\mathrm{supp}(u)$. Moreover, for every pair (u,v) of elements of $\mathbb{F}_2^n$, we denote by $u \vee v$ the element of $\mathbb{F}_2^n$ defined as: $\forall i = 1,\ldots,n$, $(u \vee v)_i = \max(u_i, v_i)$, that is, the element of $\mathbb{F}_2^n$ whose support is the union of the two supports $\mathrm{supp}(u)$ and $\mathrm{supp}(v)$. We say that an element u of a subset Π of $\mathbb{F}_2^n$ is a maximal element of Π with respect to the word partial ordering $\preceq$ if : $v \in \Pi$, $u \preceq v \Rightarrow v = u$. For every element u of $\mathbb{F}_2^n$, we denote by $\bar{u}$ the bitwise complement of u, that is, the element of $\mathbb{F}_2^n$ defined by : $\forall i \in \{1,\ldots,n\}, \bar{u}_i = 1 \oplus u_i$. We begin with proving the following key Lemma.

Lemma 4. *Let g be an n-variable Boolean function whose algebraic normal form is : $\forall x \in \mathbb{F}_2^n$, $g(x) = \bigoplus_{u \in \mathbb{F}_2^n} a_u x^u$. Set $\Pi = \{u \in \mathbb{F}_2^n \mid a_u = 1\}$. Let $\mathbf{u}$ be a maximal element of Π with respect to the partial ordering $\preceq$. Set $\Theta = \{v \in \mathbb{F}_2^n \mid v \preceq \bar{\mathbf{u}}\}$. Then $\{x^v \cdot g,\, v \in \Theta\}$ is a linearly independent family of $\mathcal{B}_n$.*

Proof. Let $(c_v)_{v\in\Theta}$ be a collection of elements of $\mathbb{F}_2$ such that : $\forall x \in \mathbb{F}_2^n$, $\bigoplus_{v\in\Theta} c_v x^v g(x) = 0$. Replacing g by its algebraic normal form yields to : $\forall x \in \mathbb{F}_2^n$, $\bigoplus_{(u,v)\in\Pi\times\Theta} c_v x^{u\vee v} = 0$. We now prove that, for every $v \in \Theta$, the monomial $x^{\mathbf{u}\vee v}$ appears only once time in the sum $\bigoplus_{(u,v)\in\Pi\times\Theta} c_v x^{u\vee v}$. To this end, let us fix $v \in \Theta$ and let us look forward $v' \in \Theta$ and $u \in \Pi$ such that $u \vee v' = \mathbf{u} \vee v$. This requires that $\mathbf{u} \preceq u \vee v'$. The support of $\mathbf{u}$ being disjoint from the support of v', we must have $\mathbf{u} \preceq u$ which is possible only if $u = \mathbf{u}$ because $\mathbf{u}$ is a maximal element of Π with respect to the word ordering $\preceq$. The equality $u \vee v' = \mathbf{u} \vee v$ becomes $\mathbf{u} \vee v' = \mathbf{u} \vee v$ from which we deduce that $v = v'$ (since they are both disjoint from $\mathbf{u}$). We hence prove that, for every $v \in \Theta$, the monomial $x^{\mathbf{u}\vee v}$ appears only once time in the sum $\bigoplus_{(u,v)\in\Pi\times\Theta} c_v x^{u\vee v}$ which vanishes for every word x in $\mathbb{F}_2^n$. That requires that $x \mapsto c_v x^{\mathbf{u}\vee v}$ is null on $\mathbb{F}_2^n$ yielding to $c_v = 0$. The element v being arbitrary, that proves that the collection $\{x^v \cdot g, v \in \Theta\}$ is a linearly independent family of $\mathcal{B}_n$. □

We then use Lemma 4 to show the following result.

Proposition 5. *Let g be an n-variable Boolean function of algebraic degree at most r and $g(x) = \bigoplus_{u\in\mathbb{F}_2^n} a_u x^u$ be its ANF. Let k be a positive integer less than n. Set $\Pi = \{u \in \mathbb{F}_2^n \mid a_u = 1\}$. Let $\mathbf{u}$ be a maximal element of Π with respect to the partial ordering $\preceq$. Then*

1. *The vector space $An_k(g \oplus 1)$ is contained in $Mul_k(g)$.*
2. $\dim Mul_k(g) \geq \sum_{i=k-\mathrm{wt}(\mathbf{u})+1}^{k} \binom{n-\mathrm{wt}(\mathbf{u})}{i} + d_{k,1\oplus g}$.

Proof.

1. Every annihilator h of $1 \oplus g$ satisfies $gh = h$ and thus is an element of $Mul_k(g)$.
2. The algebraic normal form of g can be rewritten as $g(x) = \bigoplus_{u\in\Pi} x^u$. Define $\Theta = \{v \in \Pi \mid v \preceq \bar{\mathbf{u}}\}$. Let Σ be the subset of Θ defined by $\Sigma = \{v \in \Theta \mid k - \mathrm{wt}(\mathbf{u}) + 1 \leq \mathrm{wt}(v) \leq k\}$ (this subset is non

empty because $\max_{v\in\Theta} \mathrm{wt}(v) = n - \mathrm{wt}(\mathbf{u}) \geq n - r \geq k - r + 1$). Now, $\{x^v \cdot g, v \in \Sigma\}$ is a subfamily of $\{x^v \cdot g, v \in \Theta\}$ which is a linearly independent family of $\mathcal{B}_n$ according to Lemma 4. Thus, $\{x^v \cdot g, v \in \Sigma\}$ is also a linearly independent family of $\mathcal{B}_n$. Moreover, every element of this family belongs to $Mul_k(g)$ since, for every $v \in \Sigma$, we have that $\mathrm{wt}(v) \leq k$.

Now, let V be the vector subspace spanned by all the Boolean functions $x^v g$ where v ranges over Σ. The vector subspace V is by construction a vector subspace of $Mul_k(g)$ and its dimension over $\mathbb{F}_2$ equals the cardinality of the family $\{x^v \cdot g, v \in \Sigma\}$, that is, its dimension equals $\sum_{i=k-\mathrm{wt}(\mathbf{u})+1}^{k} \binom{n-\mathrm{wt}(\mathbf{u})}{i}$.

We are now going to prove that the vector sum $V + An_k(1 \oplus g)$ is a direct sum of $Mul_k(g)$. The ANF of an element of V is of the form $\bigoplus_{(u,v)\in\Pi\times\Sigma} c_v x^{u\vee v}$. The algebraic degree of such a Boolean function is at least $k+1$. Indeed, for every $v \in \Sigma$, the monomial $x^{\mathbf{u}\vee v}$ appears at most once time in the sum $\bigoplus_{(u,v)\in\Pi\times\Sigma} c_v x^{u\vee v}$ (see proof of Lemma 4) and is of algebraic degree $\mathrm{wt}(\mathbf{u}) + \mathrm{wt}(v) \geq k+1$. Hence, the intersection $V \cap An_k(1 \oplus g)$ is reduced to $\{0\}$ because every non null element of V is of algebraic degree at least $k+1$ while every non null element of $An_k(1 \oplus g)$ is of algebraic degree at most k. This proves that the vector sum $V + An_k(1 \oplus g)$ is a direct sum. This implies that $\dim Mul_k(g) \geq \dim V + \dim An_k(1 \oplus g) = \sum_{i=k-\mathrm{wt}(\mathbf{u})+1}^{k} \binom{n-\mathrm{wt}(\mathbf{u})}{i} + d_{k,1\oplus g}$.

□

We can deduce from the Proposition 5 the following lower bound on the difference $\dim Mul_k(g) - d_{k,1\oplus g}$ valid for every Boolean function of degree at most r.

Corollary 6. *Let k be a positive integer. Then, for every n-variable Boolean function g of algebraic degree at most r, we have*

$$\dim Mul_k(g) - d_{k,1\oplus g} \geq \sum_{i=k-r+1}^{k} \binom{n-r}{i}$$

Proof. Assume that the algebraic normal form of g is : $\forall x \in \mathbb{F}_2^n$, $g(x) = \bigoplus_{u\in\mathbb{F}_2^n} a_u x^u$. Set $\Pi = \{u \in \mathbb{F}_2^n \mid a_u = 1\}$. The algebraic degree of g equals r then there exists at least one maximal element $\mathbf{u}$ of Π with respect to the word partial ordering $\preceq$ whose hamming weight equals r. We then deduce the result from Proposition 5. □

Remark 1. Proposition 5 says that, for every $w \leq r$,

$$\dim Mul_k(g) - d_{k,1\oplus g} \geq \sum_{i=k-w+1}^{k} \binom{n-w}{i}$$

if the algebraic normal form of g contains a monomial x^ω, with $\mathrm{wt}(\omega) = w$, which is not contained in any another monomial of g. Now, we have

$$\sum_{i=k-w+1}^{k} \binom{n-w}{i} \geq \sum_{i=k-r+1}^{k} \binom{n-r}{i}.$$

This follows from the identity $\binom{n-w}{i} = \sum_{p=i-r+w}^{i} \binom{n-r}{p}\binom{r-w}{i-p}$ and the sequence of equalities $\sum_{i=k-w+1}^{k} \binom{n-w}{i} = \sum_{i=k-w+1}^{k} \sum_{p=i-r+w}^{i} \binom{n-r}{p}\binom{r-w}{i-p} = \sum_{p=k-r+1}^{k} \binom{n-r}{p} \sum_{i=\max(p,k-w+1)}^{\min(p-w+r,k)} \binom{r-w}{i-p} \geq \sum_{p=k-r+1}^{k} \binom{n-r}{p}$.

Therefore, the preceding lower bound on $\dim Mul_k(g) - d_{k,1\oplus g}$ is better than that of Corollary 6 if we take $w < r$. However, it requires more information on the n-variable Boolean function g than that of Corollary 6 that simply depends on the algebraic degree of g. Now, we shall need a lower bound that does not depend on the n-variable Boolean function g to get our result. This is the reason why we shall restrict ourselves to use Corollary 6 rather than Proposition 5 in the sequel.

Remark 2. The lower bound of Corollary 6 is achieved by the complements of the indicators of $(n-r)$-dimensional affine subspaces of $\mathbb{F}_2^n$, that is, whenever g is the complement of an $(n-r)$-dimensional affine subspace of $\mathbb{F}_2^n$, it holds $\dim Mul_k(g) - d_{k,1\oplus g} = \sum_{i=k-r+1}^{k} \binom{n-r}{i}$. Indeed, we have that $d_{k,1\oplus g} = \sum_{i=0}^{k} \binom{n}{i} - \sum_{i=0}^{k} \binom{n-r}{i}$ (Proposition 1) and $d_{k,g} = \sum_{i=0}^{k-r} \binom{n-r}{i}$ (Proposition 3). Therefore, according to Lemma 2, $\dim Mul_k(g) = \sum_{i=0}^{k} \binom{n}{i} - d_{k,g} = \sum_{i=0}^{k} \binom{n}{i} - \sum_{i=0}^{k-r} \binom{n-r}{i} = \sum_{i=0}^{k} \binom{n}{i} - \sum_{i=0}^{k} \binom{n-r}{i} + \sum_{i=k-r+1}^{k} \binom{n-r}{i} = d_{k,1\oplus g} + \sum_{i=k-r+1}^{k} \binom{n-r}{i}$.

However, we do not know whether there exists or not another Boolean functions that achieve the equality $\dim Mul_k(g) - d_{k,1\oplus g} = \sum_{i=k-r+1}^{k} \binom{n-r}{i}$. The only fact that we are able to say is deduced from the arguments exposed in Remark 1, that is, if an n-variable Boolean function g achieves the equality, then all the maximal elements x^w in the ANF of g are all of algebraic degree r.

Lemma 7. *Let g be an n-variable Boolean functions of algebraic immunity k and of algebraic degree r. Suppose that $k > r$. Then the subspace $Mul_{k-r}(1\oplus g)$ is contained in $An_k(g)$.*

Proof. Let p be an element of $Mul_{k-r}(1 \oplus g)$. Assume that $p = (1 \oplus g)q$ where $q \in \mathrm{RM}(k-r, n)$. Now, $\deg(p) \leq \deg(1 \oplus g) + \deg(q) \leq r + k - r = k$. Moreover, one has $p(x) = 0$ for every $x \in \mathrm{supp}(g)$, that is, p is an annihilator of g. Thus, $Mul_{k-r}(1 \oplus g) \subset An_k(g)$. □

Remark 3. In the particular case where the n-variable Boolean function g is the complement of the indicator of an $(n-r)$-dimensional affine subspace of $\mathbb{F}_2^n$, the subspaces $Mul_{k-r}(1\oplus g)$ and $An_k(g)$ coincide because their dimensions are equal.

Indeed, note first that $\dim Mul_{k-r}(1 \oplus g) = \sum_{i=0}^{k-r} \binom{n}{i} - d_{k-r,1\oplus g} = \sum_{i=0}^{k-r} \binom{n-r}{i}$ (since $d_{k-r,1\oplus g} = \sum_{i=0}^{k-r} \binom{n}{i} - \sum_{i=0}^{k-r} \binom{n-r}{i}$ by virtue of Proposition 1). On the other hand, Proposition 3 says that $d_{k-r,g} = \sum_{i=0}^{k-r} \binom{n-r}{i}$. Thus, $\dim Mul_{k-r}(1 \oplus g) = \sum_{i=0}^{k-r} \binom{n-r}{i} = d_{k-r,g}$.

3. A new lower bound on the r-th-order nonlinearity of n-variable Boolean function with respect to their algebraic immunity

In this section, we shall see that the dimension of the vector subspace of all annihilators with prescribed algebraic degree of a Boolean function plays also an important role in relation to the r-th order nonlinearity of this Boolean function.

Given an n-variable Boolean function f and a positive integer r, we denote by $\Re_f(r,n)$ the restriction of the generator matrix of the rth-order Reed-Muller code to the support of f, that is, the columns of this matrix correspond to the evaluation of the monomials of algebraic degree at most k on the support of f. This matrix has $\mathrm{wt}(f)$ rows and $\sum_{i=0}^{k} \binom{n}{i}$ columns. But above, we have

Proposition 8. *An n-variable Boolean function f has no annihilator of algebraic degree at most k if and only if all the matrices $\Re_f(r,n)$, $r \leq k-1$, are of full rank. Moreover, one has, for every positive integer $k \leq n$,*

$$d_{k,f} + \mathrm{rank}\,(\Re_f(k,n)) = \sum_{i=0}^{k} \binom{n}{i}. \tag{1}$$

Proof. We begin with proving the first assertion. We shall prove it by contraposition, that is, we prove that an n-variable Boolean function f admits an annihilator of algebraic degree at most k if and only if the matrix $\Re_f(k,n)$ is singular.

Suppose first that f admits an annihilator of algebraic degree at most k, that is, there exists an n-variable Boolean function $p \in \mathrm{RM}(k,n)$ such that $f(x)p(x) = 0$ for every $x \in \mathbb{F}_2^n$. This is equivalent to say that $p(x) = 0$ for every $x \in \mathrm{supp}(f)$ or, in matrix form, that $\Re_f(k,n)A_p = 0$ (where A_p is the column vector whose entries are the coefficients a_v of the ANF of p, that we assume to be $p(x) = \bigoplus_{\mathrm{wt}(v)\leq k} a_v x^v$). Now, the latter equality is equivalent to say that the matrix $\Re_f(k,n)$ is singular.

Conversely, suppose that the matrix $\Re_f(k,n)$ is singular. The columns vectors $(C_v)_{\mathrm{wt}(v)\leq k}$ of $\Re_f(k,n)$ are then linearly dependent, that is, there exists a family $\{a_v,\ \mathrm{wt}(v) \leq k\}$ of elements of $\mathbb{F}_2$ such that $\bigoplus_{\mathrm{wt}(v)\leq k} a_v C_v = 0$. Now, a column C_v is the truth table of the restriction of the monomial x^v to $\mathrm{supp}(f)$. Thus, we have $\bigoplus_{\mathrm{wt}(v)\leq k} a_v x^v = 0$ for every $x \in \mathrm{supp}(f)$. Let then $p \in \mathrm{RM}(k,n)$ be the n-variable Boolean function whose ANF is $p(x) = \bigoplus_{\mathrm{wt}(v)\leq k} a_v x^v$. The latter equality is hence equivalent to say that the n-variable Boolean function p is an annihilator of f.

Identity (1) is obtained by noting that the dimension of the subspace $Mul_k(f)$ and the rank of $\Re_f(k,n)$ are equal. The result follows then from the fact that $\dim Mul_k(f) = \sum_{i=0}^{k} \binom{n}{i} - d_{k,f}$ (Lemma 2). □

The rth-order nonlinearity of a Boolean function g is the minimum Hamming distance from f to an n-variable Boolean function g of algebraic degree at most r. Our approach is to establish a lower bound on $\mathrm{dist}(f,g)$ holding for every Boolean function g of algebraic degree r. To this end, we first establish a lower bound on $\mathrm{dist}(f,g)$ involving the sum of the two dimensions $d_{k-1,g}$ and $d_{k-1,1\oplus g}$. This is the key result that will enable to improve further the lower bound of [3, 1].

Lemma 9. *Let f be an n-variable Boolean function. Suppose that $AI(f) = k$. Let r be a positive integer less than k. Then, for every n-variable Boolean function g of algebraic degree at most r, we have*

$$\mathrm{dist}(f,g) \geq d_{k-1,g} + d_{k-1,1\oplus g}.$$

Proof. Denote by d the number of bits to be modified in the truth table of f to obtain g. Denote by d_i, $i \in \{0,1\}$, the number of words of $\mathrm{supp}(i \oplus f)$ for which we modify the output value of $i \oplus f$. Clearly, we have $\mathrm{dist}(f,g) = d = d_0 + d_1$.

Now, for every positive integer ℓ, The matrix $\Re_g(\ell,n)$ is deduced from the matrix $\Re_f(\ell,n)$ by deleting d_0 rows and adding d_1 rows. The matrix $\Re_f(k-1,n)$ being of full rank according to proposition 8, we hence have that $\mathrm{rank}(\Re_g(k-1,n)) \geq \sum_{i=0}^{k-1} \binom{n}{i} - d_0$ and thus that $d_0 \geq \sum_{i=0}^{k-1} \binom{n}{i} - \mathrm{rank}(\Re_g(k-1,n)) = d_{k-1,g}$.

Similarly, the matrix $\Re_{1\oplus g}(\ell,n)$ is deduced from the matrix $\Re_{1\oplus f}(\ell,n)$ by deleting d_1 rows and adding d_0 rows. The matrix $\Re_f(k-1,n)$ being also of full rank, we hence deduce by similar arguments as those exposed previously that $d_1 \geq d_{k-1,1\oplus g}$. □

Remark 4. Collecting together Lemma 3 applied to affine Boolean functions and Lemma 9 leads to $\mathrm{dist}(f,l) \geq d_{k-1,l} + d_{k-1,1\oplus l} = 2\sum_{i=0}^{k-2} \binom{n-1}{i}$ for every n-variable affine Boolean functions, that is, we recover the lower bound of [11].

Similarly, Applying Lemma 9 to n-variable Boolean functions of algebraic degree at most r leads to $\mathrm{dist}(f,g) \geq 2\sum_{i=0}^{k-r-1} \binom{n-r}{i}$, that is, we recover the first lower bound of [1, Theorem 1].

We then deduce from Lemma 5 and Lemma 9 our lower bound on the rth-order linearity of an n-variable Boolean function with prescribed algebraic immunity. Our idea is to get a lower bound on this sum rather than considering separately the two dimensions $d_{k-1,g}$ and $d_{k-1,1\oplus g}$.

Theorem 10. *Let f be an n-variable Boolean function of algebraic immunity k and let r be a positive integer less than k. Then*

$$nl_r(f) \geq \sum_{i=0}^{k-r-1} \binom{n}{i} + \sum_{i=k-2r}^{k-r-1} \binom{n-r}{i}$$

Proof. Let g be an arbitrary n-variable Boolean function of algebraic degree at most r. According to Lemma 9, we have

$$\mathrm{dist}(f,g) \geq d_{k-1,g} + d_{k-1,1\oplus g}.$$

Now, according to Lemma 7, one has $An_{k-1}(g \oplus 1) \supset Mul_{k-r-1}(g)$ and $An_{k-1}(g) \supset Mul_{k-r-1}(1 \oplus g)$. Hence

$$\mathrm{dist}(f,g) \geq d_{k-1,g} + d_{k-1,1\oplus g} \geq \dim Mul_{k-r-1}(g) + \dim Mul_{k-r-1}(1 \oplus g)$$

Next, thanks to Lemma 2, we get

$$\begin{aligned} \mathrm{dist}(f,g) &\geq \dim Mul_{k-r-1}(g) + \dim Mul_{k-r-1}(1 \oplus g) \\ &= \sum_{i=0}^{k-r-1} \binom{n}{i} + \dim Mul_{k-r-1}(g) - d_{k-r-1,1\oplus g}. \end{aligned}$$

We finally conclude thanks to Corollary 6 that says that

$$\dim Mul_{k-r-1}(g) - d_{k-r-1,1\oplus g} \geq \sum_{i=k-2r}^{k-r-1} \binom{n-r}{i}.$$

□

Remark 5. In the particular case where $r = 1$, Theorem 10 says that

$$nl(f) \geq \sum_{i=0}^{k-2} \binom{n}{i} + \binom{n-1}{k-2}. \tag{2}$$

Now, use the identity $\binom{n}{i} = \binom{n-1}{i} + \binom{n-1}{i-1}$ in the first summation of the right-hand side of (2) :

$$\sum_{i=0}^{k-2} \binom{n}{i} = 1 + \sum_{i=1}^{k-2} \binom{n-1}{i} + \sum_{i=1}^{k-2} \binom{n-1}{i-1} = 2\sum_{i=0}^{k-3} \binom{n-1}{i} + \binom{n-1}{k-2}$$

Thus, we get

$$nl(f) \geq 2\sum_{i=0}^{k-2} \binom{n-1}{i}$$

which is exactly the lower bound of [11].

Remark 6. Theorem 10 improves further the result of [3] for all orders. We present in Table 3 the comparison between our lower bound and the lower bound of [3]. On the other hand, it only improves partially the result of [1]. We present in table 2 the comparison between the lower bound of Theorem 10 and the lower bound of [1]. Moreover, we give in table 1 the best lower

bound between ours (that we write in bold text) and those of [1]. We have checked by computer experiments that, for every $n \leq 60$, our lower bound improves the lower bound of [1] for $2 \leq k \leq \lceil \frac{n}{2} \rceil$ and $2 \leq k \leq \lceil \frac{n}{2} \rceil$ while it does not improve the lower bound of [1] for $2 \leq k \leq \lceil \frac{n}{2} \rceil$ and $\lfloor \frac{k-1}{2} \rfloor + 3 \leq r \leq k$. However, we do not know whether it holds for every positive integer n or not. Concerning the cases where $r \in \left\{ \lfloor \frac{k-1}{2} \rfloor + 1, \lfloor \frac{k-1}{2} \rfloor + 2 \right\}$, we have found by computer experiments that our lower bound is better than the lower bound of [1] for some values of (k, n) with $n \leq 60$ and $2 \leq k \leq \lceil \frac{n}{2} \rceil$.

References

1. C. Carlet, *On the higher order nonlinearities of algebraic immune Boolean functions*, CRYPTO 2006, Lecture notes in Computer Science, vol. 4117, 2006, pp. 584–601.
2. ———, *Boolean functions for cryptography and error correcting codes*, Cambridge University Press, 2007.
3. C. Carlet, D. K. Dalai, K. C. Gupta, and S. Maitra., *Algebraic immunity for cryptographically significant boolean functions: Analysis and construction*, IEEE Transactions on Information Theory **52** (2006), no. 7, 3105–3121.
4. C. Carlet and S. Mesnager, *Improving the upper bounds on thecovering radii of binary Reed-Muller codes*, IEEE Transactions on Information Theory **53** (2007), no. 1, 162–173.
5. N. Courtois, *Higher Order Correlation Attacks, XL algorithm, and Cryptanalysis of Toyocrypt*, ICISC 2002, Lecture notes in Computer Science, vol. 2587, Spinger-Verlag, 2002, The extended version of the paper can be found on `http://eprint.iacr.org/2002/087/`, pp. 182–199.
6. ———, *Fast algebraic attacks on stream ciphers with linear feedback*, Advances in Cryptology - CRYPTO 2003, LNCS, vol. 2729, Spinger-Verlag, 2003, pp. 176–194.
7. D. K. Dalai, K. C. Gupta, and S. Maitra, *Notion of algebraic immunity and its evaluation related to fast algebraic attacks*, International Workshop on Boolean Functions : Cryptography and Applications, 2006, 13–15 March, Rouen, France, 2006.
8. J. Golic, *Fast low order approximation of cryptographic functions*, EUROCRYPT'96, Lecture notes in Computer Science, vol. 1070, Spinger-Verlag, 1996, pp. 268–282.
9. T. Iwata and K. Kurosawa, *Probabilistic higher order differential attack and higher order bent functions*, ASIACRYPT'99, Lecture notes in Computer Science, vol. 1716, Spinger-Verlag, 1999, pp. 62–74.
10. L. R. Knudsen and M. J. B. Robshaw, *Nonlinear approximations in linear cryptanalysis*, EUROCRYPT'96, Lecture notes in Computer Science, vol. 1070, Spinger-Verlag, 1996, pp. 224–236.
11. M. Lobanov, *Tight bound between nonlinearity and algebraic immunity*, Cryptology ePrint Archive, Report 2005/441, 2005, `http://eprint.iacr.org/`.
12. W. Meier and N. Courtois, *Algebraic Attacks on Stream Ciphers with Liner Feedback*, Eurocrypt 2003, Lecture notes in Computer Science, vol. 2656, Spinger-Verlag, 2003, This attack has been now implemented by Nicolas Courtois, see the extended version of the paper at `http://www.minrank.org/toyolili.pdf`. Also reproduced (illegally) at `http://www.esat.kuleuven.ac.be/~jlano/stream/papers/coumei03.pdf`, pp. 345–359.
13. W. Meier, E. Pasalic, and C. Carlet, *Algebraic Attacks and Decomposition of Boolean Functions*, Eurocrypt 2004, Lecture notes in Computer Science, vol. 3027, Spinger-Verlag, 2004, Slides at `http://www.zurich.ibm.com/eurocrypt2004/slides/session14talk1.pdf`, pp. 474–491.

n \ r	2	3	4	5	6	7
18	**43556**	**17439**	**5518**	1976	344	38
19	**126008**	**57992**	**21592**	**6507**	2320	382
20	**188368**	**81404**	**28568**	8826	2702	422
21	**527900**	**257396**	**103784**	**34780**	15094	3124
22	**803860**	**369748**	**141064**	**44844**	18218	3588
23	**2195580**	**1123220**	**483680**	**176660**	**53954**	21806
24	**3396320**	**1645660**	**672784**	**233827**	**68071**	25902
25	**9080772**	**4838490**	**2202164**	**863975**	**289301**	136812
26	**14239032**	**7211198**	**3125248**	**1169920**	**374371**	167364
27	**37392864**	**20633040**	**9846132**	**4104275**	**1484042**	**458054**
28	**59333408**	**31214643**	**14221898**	**5670245**	**1963795**	**581338**
29	**153434536**	**87279291**	**43393566**	**19055725**	**7355234**	**2462995**
30	**246025562**	**133797407**	**63665462**	**26799567**	**9928262**	**3194667**

TABLE 1. Best lower bounds on $nl_r(f)$ for $18 \leq n \leq 30$, $AI(f) = \lceil \frac{n}{2} \rceil$, $r \leq 7$

n \ r	2	3	4	5	6	7
18	1.46	1.76	1.88	0.69	0.73	0.82
19	1.53	1.95	2.18	1.01	0.66	0.71
20	1.49	1.87	2.07	0.92	0.67	0.72
21	1.55	2.04	2.38	2.30	0.63	0.65
22	1.52	1.96	2.26	2.38	0.64	0.66
23	1.58	2.13	2.57	2.83	1.33	0.62
24	1.55	2.05	2.44	2.67	1.21	0.62
25	1.60	2.20	2.74	3.13	2.11	0.59
26	1.57	2.12	2.60	2.95	2.24	0.60
27	1.61	2.27	2.90	3.42	3.74	1.77
28	1.59	2.19	2.76	3.22	3.50	1.60
29	1.63	2.33	3.05	3.69	4.17	2.12
30	1.60	2.26	2.90	3.48	3.90	2.08

TABLE 2. The new lower bound over the Lower bound of [1] for $18 \leq n \leq 30$, $AI(f) = \lceil \frac{n}{2} \rceil$, $r \leq 7$

14. W. Millan, *Low order approximations of cipher functions*, Cryptographic Policy and Algortihms, Lecture notes in Computer Science, vol. 1029, Springer-Verlag, 1996, pp. 144–155.
15. C.E. Shannon, *Communication theory of secrecy systems*, vol. 28, pp. 656–715, Bell system technical journal, 1949.

APPENDIX A. TABLES

n \ r	2	3	4	5	6	7
18	1.40	1.38	1.36	1.38	1.46	1.63
19	1.34	1.32	1.30	1.29	1.33	1.41
20	1.37	1.35	1.32	1.31	1.35	1.44
21	1.31	1.30	1.26	1.25	1.26	1.30
22	1.34	1.32	1.28	1.27	1.28	1.32
23	1.29	1.27	1.24	1.21	1.21	1.23
24	1.32	1.29	1.25	1.23	1.23	1.25
25	1.28	1.26	1.22	1.19	1.18	1.18
26	1.30	1.27	1.23	1.20	1.19	1.20
27	1.26	1.24	1.20	1.17	1.15	1.15
28	1.28	1.26	1.22	1.18	1.17	1.16
29	1.25	1.23	1.19	1.16	1.14	1.13
30	1.27	1.24	1.20	1.17	1.15	1.14

TABLE 3. The new lower bound over the Lower bound of [3] for $18 \leq n \leq 30$, $AI(f) = \left\lceil \frac{n}{2} \right\rceil$, $r \leq 7$