On Boolean functions with maximum algebraic immunity

Enes Pasalic
Technical University of Denmark,
Matematiktorvet, Building 303,
DK- 2800 Kgs. Lyngby, DENMARK;
E.Pasalic@mat.dtu.dk

Abstract

In this paper two important issues in theory of algebraic attacks are addressed. We first provide a theoretical framework for better understanding of design rationale in construction of Boolean functions with maximum algebraic immunity. Based on these results, an iterative design of functions with maximum possible algebraic immunity is proposed. In contrast to known constructions, our method generates balanced functions of maximum degree and high nonlinearity, that is functions satisfying all main cryptographic criteria. Additionally, functions in this class have a low implementation cost due to a small number of terms in the ANF. Secondly, for a given Boolean function, a novel algorithm for deciding the existence of annihilators of small degree is presented. The algorithm utilizes the known methods in a slightly different manner which results in a significantly reduced complexity of computation.

Keyword : Algebraic attacks, Algebraic Immunity, Annihilators, Stream ciphers, Nonlinear combiner, Boolean function, Resiliency, Algebraic Degree.

1 Introduction

Boolean functions have important applications in so-called linear transition stream ciphers based on nonlinear filtering of a single or several linear feedback shift registers (LFSR). Two main representatives for this class of ciphers are nonlinear filter generators and nonlinear combiners [12]. Apart from already established cryptographic criteria such as nonlinearity, algebraic degree, and resiliency, it turned out that Boolean functions must also have a certain order of algebraic immunity. This is due to recently introduced algebraic attacks based on the low degree annihilation of Boolean functions [6, 8]. These attacks reflect the property of certain cipher schemes for which the selection of a function f of high algebraic degree to prevent Shannon's attacks [15] and linear complexity attacks [12] is not a sufficient criterion any longer. Instead of setting up a system of equations of degree determined by the degree of function f (this is regarded as Shannon's attack), the attacker can derive a lower degree system provided the existence of a low degree function g, called *annihilator*, such that $fg = 0$ or alternatively $(1 + f)g = 0$ [8, 11]. We will later discuss in more depth how such an attack is practically performed.

In [8], it was proved that for any Boolean function f in n variables there always exists a function g of degree $\lceil \frac{n}{2} \rceil$ such that either $fg = 0$ or $(1+f)g = 0$. Then the minimum degree of nonzero annihilators of either f or $1+f$ is by definition *algebraic immunity* [11]. The existence of functions for which we cannot find a further degree reduction, that is a nonzero function g that annihilates either f or $(1+f)$ is of degree strictly $\geqslant \lceil \frac{n}{2} \rceil$ was first pointed out in [11]. This is not an exceptional case at all, and various computer simulations indicate that the algebraic immunity is mainly concentrated in the range $\{\lceil \frac{n}{2} \rceil - 1, \lceil \frac{n}{2} \rceil\}$. Apparently, there was a need for deterministic techniques to construct functions with maximum algebraic immunity. This was the main cause for several construction methods that generate functions with maximum algebraic immunity to appear recently [5, 9, 2, 10]. However, all these methods fail to optimize other cryptographic criteria at the same time.

This work is mainly motivated by the fact that at the time being all the construction methods fail to provide functions satisfying all important cryptographic criteria. Thus we cannot generate strong cryptographic functions through design methods. Moreover, for a relatively large input variable space (number of variables $n > 20$) the complexity of known algorithms for determining the exact value of algebraic immunity becomes infeasible in most of the cases.

Therefore, this paper deals with these two fundamental tasks in theory of algebraic attacks. In the first place, for the first time we exhibit a construction method that generates functions with maximum algebraic immunity which also succeeds to attain overall good cryptographic properties such as balancedness, high nonlinearity and maximum degree. This is achieved by developing useful theoretical results on functions with maximum algebraic immunity. In connection to the difficulty of determining the exact value of algebraic immunity for large $n > 20$, we propose a deterministic algorithm that repeatedly examine the existence of annihilators for subfunctions of f (these subfunctions are regarded as restrictions of f to a smaller variable space). Based on this, the question about the (non)existence of annihilators of degree $\leqslant d$ for function f can be answered with the computational complexity which is expected to be much smaller compared to known algorithms. The algorithm has therefore an important application when analyzing the algebraic properties of Boolean functions.

The rest of the paper is organized as follows. In Section 2 basic definitions and notations are introduced. Also, a classical application of algebraic attacks is treated in greater details. A theoretical framework regarding the properties of functions with maximum algebraic immunity is developed in Section 3. These results are then used in Section 4 to derive a new construction method for generation of functions with maximum algebraic immunity, with overall good cryptographic properties. A novel algorithm for determining the existence of annihilators of degree $\leqslant d$ is discussed in Section 5. Section 6 concludes the paper.

2 Preliminaries

We denote the Galois field of order 2^n by $\mathbb{F}_{2^n}$ and the corresponding vector space by $\mathbb{F}_2^n$. A Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is usually represented via so called *algebraic normal form* (ANF),

$$f(x_1, \ldots, x_n) = \sum_{u \in \mathbb{F}_2^n} \lambda_u \left(\prod_{i=1}^{n} x_i^{u_i} \right) \ , \quad \lambda_u \in \mathbb{F}_2 \ , u = (u_1, \ldots, u_n). \tag{1}$$

For the rest of this paper, if otherwise not stated, x will denote a vector containing n input binary variables, that is $x = (x_1, \ldots, x_n) \in \mathbb{F}_2^n$. The *algebraic degree* of f, denoted by $deg(f)$ or sometimes simply d, is the maximal value of the Hamming weight of u such that $\lambda_u \neq 0$. The set of all Boolean functions in n variables is denoted by $\mathcal{B}_n$, and functions of degree at most one are called *affine* functions. An affine function with constant term equal to zero is called a *linear* function. The support set of function $f \in \mathcal{B}_n$, denoted by $supp(f)$, is the set of input values where f has a nonzero evaluation, that is,

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

A function f is said to be balanced if it outputs equal number of zeros and ones, that is $\#\{x \in \mathbb{F}_2^n : f(x) = 1\} = \#\{x \in \mathbb{F}_2^n : f(x) = 0\}$. The *nonlinearity* of an n-variable function f is defined as the minimum distance from the set of all n-variable affine functions,

$$\mathcal{N}_f = min_{g \in \mathcal{A}_n}(d_H(f, g)), \tag{2}$$

where d_H denotes the Hamming distance, i.e. $d_H(f, g) = \#\{x \in \mathbb{F}_2^n \mid f(x) \neq g(x)\}$.

The notion of algebraic immunity (AI) of order d was introduced in [11] meaning the ability of function f or $(1 + f)$ not to admit aninhilation by any function of degree $< d$. Using elementary arguments it was shown that $\mathrm{AI}(f) \leqslant \lceil n/2 \rceil$ for any $f \in \mathcal{B}_n$, see [8]. The set of annihilators of f is denoted $An(f)$ and the minimum nonzero degree of functions in this set is $d^0(An(f))$. A function $f \in \mathcal{B}_n$ whose algebraic immunity attains the upper bound $\mathrm{AI}(f) = \lceil n/2 \rceil$ is called a function with maximum AI.

The main idea behind algebraic attacks on additive stream ciphers is that annihilation of f or $1 + f$ by a low degree function g results in a system of nonlinear equations with a relatively small number of terms. These attacks, though generic in its nature, are applicable in a straightforward manner to so-called nonlinear combiners and filtering generators [12]. Assuming a known-plaintext attack this class of ciphers is characterized by the fact that each output bit induces a multivariate equation of certain degree in the secret key bits. Then in the case of the existence of annihilators of f or $1 + f$ of lower degree than f, the low degree system of multivariate equations (given by degree of annihilators) may be set up which substantially reduces the complexity of solving such a system. The new system of equations is usually solved using so-called linearization technique, that is each nonlinear equation is turned into linear one [8] by replacing nonlinear terms with new variables.

When the system is overdefined, which means that the number of equations is larger than the number of unknowns, it may be solved by Gaussian elimination. Notice that the total complexity only depends on the state size of the cipher S and the degree d of annihilator g. An approximate estimate for this complexity is given as $\binom{S}{d}^{\omega}$ in [8], where ω is the complexity of Gaussian elimination (usually one takes $\omega = 3$). For further understanding how these attacks work the reader is referred to [11, 8] where a rather detailed description is given.

3 Optimizing algebraic immunity

Several construction methods [9, 5, 2, 10] providing functions with maximum algebraic immunity have recently appeared. Nevertheless, all proposed techniques

achieve the optimization of algebraic immunity at the price of rather severe degradation of other cryptographic criteria such as degree and nonlinearity. It is hard to believe that these classes of functions are suitable for cryptographic applications.

The purpose of this section is to identify some basic conditions that any function with maximum AI must satisfy. It is an open problem to find relationship between annihilators of f and $(1+f)$. However, when $f \in \mathcal{B}_n$ is balanced (that is $supp(f) = 2^{n-1}$) and n is odd, the following result could be derived.

Proposition 1 *[4] When n is odd, a balanced function $f \in \mathcal{B}_n$ has a maximum algebraic immunity, that is $AI(f) = (n+1)/2$, if and only if $d^0(An(f)) = (n+1)/2$.*

This means that for odd n and a balanced function $f \in \mathcal{B}_n$, it is sufficient to show that the minimum degree of nonzero annihilators of f is $(n+1)/2$ and the same minimum degree for annihilators of $1+f$ is automatically obtained.

Let us consider a concatenation of two arbitrary functions $f_1, f_2 \in \mathcal{B}_n$, commonly denoted by $f = f_1||f_2$, and specified by

$$\begin{aligned} f(x_1, \ldots, x_n, x_{n+1}) &= (1 + x_{n+1})f_1(x) + x_{n+1}f_2(x) = \\ &= x_{n+1}\{f_1(x) + f_2(x)\} + f_1(x). \end{aligned}$$

For shortness of notation we sometimes use $f = x_{n+1}\{f_1 + f_2\} + f_1$. Obviously any annihilator $g \in \mathcal{B}_{n+1}$ of f can be written as

$$g = x_{n+1}\{g_1 + g_2\} + g_1, \tag{3}$$

where $g_1, g_2 \in \mathcal{B}_n$ annihilate f_1 respectively f_2. Note that it as allowable to take either g_1 or g_2 to be a zero function. The following relation is then deduced in [4],

$$AI(f) = \begin{cases} \min(AI(f_1), AI(f_2)) + 1, & AI(f_1) \neq AI(f_2); \\ AI(f_1) \text{ or } AI(f_1) + 1, & AI(f_1) = AI(f_2); \end{cases}$$

Then selecting $f_1, f_2 \in B_n$ with maximum $\mathrm{AI}(f_i) = (n+1)/2$ (n being odd) the algebraic immunity of function $f = f_1||f_2$ satisfies $\mathrm{AI}(f) \in \{(n+1)/2, (n+1)/2+1\}$. On the other hand, $\mathrm{AI}(f) \leqslant (n+1)/2$ and therefore any $f_1, f_2 \in \mathcal{B}_n$ of maximum AI for odd n, gives a maximum AI function $f = f_1||f_2 \in \mathcal{B}_{n+1}$. Moreover, the equation (3) implies the following.

Lemma 1 *Let n be odd, and let f be a function whose subfunctions $f_1, f_2 \in \mathcal{B}_n$ are arbitrary functions having $AI(f_i) = (n+1)/2$. Then there must exist some g_1 and g_2 of degree $(n+1)/2$ which annihilates f_1 respectively f_2 (alternatively annihilating $1+f_1$ and $1+f_2$) such that $deg(g_1 + g_2) < (n+1)/2$.*

Note that taking f_1 and f_2 with maximum AI is sufficient but not necessary condition. One might for instance consider $f_1, f_2 \in \mathcal{B}_n$ (for odd n) such that $\mathrm{AI}(f_i) = (n+1)/2-1$ and still obtain $f = f_1||f_2$ which has optimized $AI = (n+1)/2$. The following theorem is the key result for the iterative construction of functions with maximum AI given in the next section.

Theorem 1 *Let $f_1 \in \mathcal{B}_n$ be a balanced maximum* AI *function, that is* $\mathrm{AI}(f_1) = \lceil \frac{n}{2} \rceil$. *Then the* AI *of $f_2(x) = f_1(x) + x_1 \cdots x_n$ satisfies,*

$$\mathrm{AI}(f_2) \in \{\left\lceil \frac{n}{2} \right\rceil - 1, \left\lceil \frac{n}{2} \right\rceil\}.$$

Proof. Assume the existence of nonzero annihilator g_2 of f_2 of degree $< \lceil \frac{n}{2} \rceil - 1$. Then,

$$f_2(x)g_2(x) = 0 \implies f_1(x)g_2(x) = x_1 \cdots x_n g_2(x),$$

where the product $x_1 \cdots x_n g_2(x)$ is either 0 or $x_1 \cdots x_n$ depending on the parity of the number of terms in the ANF of g_2. If this parity is even then $x_1 \cdots x_n g_2(x) = 0$, hence g_2 of degree $< \lceil \frac{n}{2} \rceil - 1$ is annihilator of f_1, a contradiction. Then assuming the ANF of g_2 is of odd parity,

$$f_1(x)g_2(x) = x_1 \cdots x_n.$$

Multiplying the above equation with $(1 + x_i)$ for any $i \in [1, n]$ gives,

$$(1 + f_1(x))g_2(x)(1 + x_i) = 0,$$

that is the function $g_2(x)(1 + x_i)$, which is of degree $< \lceil \frac{n}{2} \rceil$, annihilates $1 + f_1$, a contradiction again. This means that the minimum degree of nonzero annihilators of f_2 is either $\lceil \frac{n}{2} \rceil - 1$ or $\lceil \frac{n}{2} \rceil$.

To prove the assertion it remains to show that $1 + f_2$ does not admit annihilators of degree less than $\lceil \frac{n}{2} \rceil - 1$. On contrary, assume g_2 is an annihilator of $1 + f_2$ of degree $< \lceil \frac{n}{2} \rceil - 1$. Then $(1 + f_2(x))g_2(x) = 0$ gives,

$$(1 + f_1(x))g_2(x) = x_1 \cdots x_n g_2(x).$$

Similarly to above, the even parity of g_2's ANF implies a contradiction (g_2 cannot annihilate $(1 + f_1)$). For the ANF of g_2 of odd parity we get,

$$(1 + f_1(x))g_2(x) = x_1 \cdots x_n,$$

which again, after multiplying the above equation with $(1+x_i)$, gives a contradiction. Thus, $\mathrm{AI}(f_2) \in \{\lceil \frac{n}{2} \rceil - 1, \lceil \frac{n}{2} \rceil\}$ as stated. □

The concatenation of functions with maximum AI may be extended to consider $f = f_1||f_2||f_3||f_4 \in \mathcal{B}_{n+2}$, where each $f_i \in \mathcal{B}_n$ has maximum AI. Using the shortened notation (f_i denoting $f_i(x)$), the ANF of function f is given by:

$$f = x_{n+1}x_{n+2}(f_1 + f_2 + f_3 + f_4) + x_{n+1}(f_1 + f_2) + x_{n+2}(f_1 + f_3) + f_1.$$

A similar expression is then valid for any annihilator g of f,

$$g = x_{n+1}x_{n+2}(g_1 + g_2 + g_3 + g_4) + x_{n+1}(g_1 + g_2) + x_{n+2}(g_1 + g_3) + g_1, \qquad (4)$$

where g_i is arbitrary annihilator of f_i (including the trivial annihilation $g_i = 0$). Let g_i denote any minimum degree nonzero annihilator of $f_i \in \mathcal{B}_n$. If $deg(g_i) = d$ then we also use,

$$g_i(x) = g_i^d(x) + g_i^{d-1}(x) + \cdots + g_i^0(x),$$

where each g_i^r, for $0 \leqslant r \leqslant d$, contains only degree r monomial terms. Then in connection to the representation of annihilator g of f given in (4), the following simple property is obtained.

Lemma 2 *Let $f = f_1||f_2||f_3||f_4$, where $f_i \in \mathcal{B}_n$ are functions with maximum* AI. *Then if there exists g such that $deg(g) < \lceil \frac{n}{2} \rceil + 1$ then $deg(g_i) = \lceil \frac{n}{2} \rceil$ for all $i \in [1, 4]$ and furthermore,*

$$g_1^d = g_2^d = g_3^d = g_4^d; \qquad \sum_{i=1}^{4} g_i^{d-1} = 0. \qquad (5)$$

Proof. Assuming $deg(g) < \lceil \frac{n}{2} \rceil + 1$ gives that,

$$\sum_{i=1}^{4} g_i^d = 0 \qquad \sum_{i=1}^{4} g_i^{d-1} = 0 \qquad g_1^d + g_2^d = 0 \qquad g_1^d + g_3^d = 0,$$

and the result easily follows. ∎

The result of Lemma 2 is a useful tool for establishing the algebraic properties of given function. Showing that subfunctions $f_1, \ldots, f_4$ of maximum AI are chosen so that conditions in Lemma 2 cannot be satisfied for neither f nor $1+f$ is equivalent to proving that $f = f_1||f_2||f_3||f_4$ has a maximum AI.

It is well-known that the algebraic immunity is invariant under composition with linear permutation, which is in particular true for a permutation of subfunctions $f_1, \ldots, f_4$. Therefore the AI of function $f_1||f_2||f_3||f_4$ is the same as for $f_1||f_3||f_2||f_4$ for instance. This fact will be frequently used later to simplify some proofs and therefore it is given in the form of statement.

Lemma 3 *The algebraic immunity of f is invariant under composition with linear permutation, and in particular it is invariant under permutation of its subfunctions.*

4 Construction of functions with maximum AI

The results introduced in the previous section will now contribute in proposing some recursive constructions of functions with maximum algebraic immunity. The existence of such functions is strongly supported by computer simulations (at least for a relatively small variable space $n \leqslant 15$). The goal is to investigate the possibility of concatenating four suitable functions, say $f_1, \ldots, f_4 \in \mathcal{B}_n$ in order to generate a function with maximum AI on $\mathbb{F}_2^{n+2}$.

In what follows we use a single function $f_1 \in \mathcal{B}_n$ from which we derive f_2, f_3, f_4 by suitable modifications.

Construction 1 *Let $f_1 \in \mathcal{B}_n$ be a balanced function with maximum* AI*, n odd. Let the ANF of f_1 contain even number of terms. Then the function $f = f_1||f_2||f_3||f_4 \in \mathcal{B}_{n+2}$ has maximum* AI *for the following choice of f_2, f_3 and f_4:*

$$f_2 = f_1 + x_1 x_2 \cdots x_n; \qquad f_3 = f_1; \quad f_4 = 1 + f_1 + x_1 x_2 \cdots x_n.$$

Moreover, the ANF of f is given by,

$$f(x_1, \ldots, x_{n+2}) = x_{n+1}x_{n+2} + x_1 \cdots x_{n+1} + f_1(x_1, \ldots, x_n).$$

Proof. We show that $deg(g) \geqslant \lceil \frac{n}{2} \rceil + 1$ for any nonzero annihilator g of either f or $1+f$. The equations $f_1 g_1 = 0$ and $f_2 g_2 = 0$ give,

$$f_1(x)(g_1(x) + g_2(x)) = x_1 \cdots x_n g_2(x) = 0 \quad \text{or} \quad x_1 \cdots x_n, \tag{6}$$

depending on the parity of the ANF of g_2. As f_1 has maximum AI, then either $g_1 = 0$ or $deg(g_1) \geqslant \lceil \frac{n}{2} \rceil$. We have to consider several cases depending on the choice of g_i.

i) Setting $g_1 = 0$ implies $g_3 = 0$, as due to the term $x_{n+2}(g_1 + g_3)$ we would have $deg(g) \geqslant \lceil \frac{n}{2} \rceil + 1$. Then assuming $g_1 = g_3 = 0$, a nonzero choice of g_2 implies the following. If $deg(g_2) \geqslant \lceil \frac{n}{2} \rceil$ then $deg(g) \geqslant \lceil \frac{n}{2} \rceil + 1$ due to $x_{n+1}(g_1 + g_2)$ in (4).

So we assume $deg(g_2) < \lceil \frac{n}{2} \rceil$. By Theorem 1, the minimum degree of nonzero g_2 is $\lceil \frac{n}{2} \rceil - 1$. Then for even parity of the ANF of g_2, the equation (6) becomes,

$$f_1(x)(0 + g_2(x)) = 0,$$

contradicting that f_1 is of maximum AI. So we must assume that the ANF of g_2 is of odd weight, that is $f_1(x)g_2(x) = x_1x_2 \cdots x_n$. This is impossible as f_1 has an even number of terms in its ANF. To summarize, taking $g_1 = 0$ leads now to the only remaining case $g_1 = g_2 = g_3 = 0$ and $g_4 \neq 0$. By Theorem 1, $\mathrm{AI}(f_4) \geqslant \lceil \frac{n}{2} \rceil - 1$, and therefore $deg(g) \geqslant \lceil \frac{n}{2} \rceil + 1$ due to term $x_{n+2}x_{n+1}g_4(x)$.

ii) Clearly, if $g_1 \neq 0$ then also $g_2 \neq 0$, as taking $g_2 = 0$ gives $deg(g) \geqslant \lceil \frac{n}{2} \rceil + 1$ due to term $x_{n+1}(g_1(x) + 0)$ in (4). Also from equation (4), nonzero $g_1, g_2 \neq 0$ must also satisfy $deg(g_1 + g_2) < \lceil \frac{n}{2} \rceil$, as otherwise $deg(g) \geqslant \lceil \frac{n}{2} \rceil + 1$. The case $f_1(x)(g_1(x) + g_2(x)) = 0$ in equation (6), gives annihilator of f_1 of degree $< \lceil \frac{n}{2} \rceil$, a contradiction. Thus, we must have

$$f_1(x)(g_1(x) + g_2(x)) = x_1x_2 \cdots x_n,$$

contradicting the assumption that the ANF of f_1 is of even parity. Therefore $d^0(An(f)) = \lceil \frac{n}{2} \rceil + 1$, that is maximum degree is achieved.

It remains to prove that the same is true for annihilators of $1 + f$. But the conditions of Proposition 1 are satisfied so we have that $\mathrm{AI}(f) = \lceil \frac{n}{2} \rceil + 1$.
The choice of subfunctions of f and representation of f as,

$$f = x_{n+1}x_{n+2}(f_1 + f_2 + f_3 + f_4) + x_{n+1}(f_1 + f_2) + x_{n+2}(f_1 + f_3) + f_1,$$

give the ANF of f as stated. □

4.1 Cryptographic properties of the new maximum AI class

The main cryptographic properties of the class of functions proposed in Construction 1 are summarized below.

Theorem 2 *The function $f \in \mathcal{B}_{n+2}$ in Construction 1 satisfies the following:*

i) *f is a balanced function on $\mathbb{F}_2^{n+2}$ of maximum degree, that is $deg(f) = n + 1$.*

ii) *If the nonlinearity of f_1 is $\mathcal{N}_{f_1}$, then $\mathcal{N}_f \in \{2^n + 2\mathcal{N}_{f_1}, 2^n + 2\mathcal{N}_{f_1} \pm 2\}$. Furthermore, if nonlinearity of f_1 reaches the bent concatenation bound, i.e. $\mathcal{N}_{f_1} = 2^{n-1} - 2^{\frac{n-1}{2}}$, then*

$$\mathcal{N}_f \in \{2^{n+1} - 2^{\frac{n+1}{2}}, 2^{n+1} - 2^{\frac{n+1}{2}} \pm 2\}.$$

Proof. i) f is clearly balanced due to the choice of its subfunctions. From the ANF of f we have $deg(f) = n + 1$.
ii) It is well-known that the nonlinearity of $f' = f_1||f_1||f_1||1+f_1$ is $\mathcal{N}_{f'} = 2^n + 2\mathcal{N}_{f_1}$, see for instance [3]. Since our method only differs in complementation of two bits in the truth table of f', then the statement is obvious.
In particular, when $\mathcal{N}_{f_1} = 2^{n-1} - 2^{\frac{n-1}{2}}$ then $\mathcal{N}_{f'} = 2^n + 2\mathcal{N}_{f_1} = 2^{n+1} - 2^{\frac{n+1}{2}}$, so that the nonlinearity of f is as claimed. □

Remark 1 *The only cryptographic criterion that is not covered by this construction is resiliency.*[1] *Though the function f cannot be resilient function (this would violate Siegenthaler's upper bound on the degree [16] given by $deg(f) \leqslant n - t - 1$ for any t-resilient f) one may show that assuming f_1 is t-resilient function then $f(x) + l(x)$ is only slightly unbalanced for any linear function $l(x)$ having at most t linear terms. More precisely, the correlation coefficient ε (that measures susceptibility to correlation attacks) is equal to $\varepsilon = 0.5 + 2/2^n$ for linear functions of at most t terms. This is a small deviation from the ideal value $\varepsilon = 0.5$ in the case of resilient functions.*

This is the first time that a construction comprising most of the cryptographic criteria has been proposed. More importantly, the method is recursive so we can generate infinite sequences of functions with maximum AI and overall good cryptographic properties. To use the construction in a recursive manner the resulting function $f \in \mathcal{B}_{n+2}$ should have an even number of terms in its ANF. This is satisfied by the construction as,

$$f(x_1, \ldots, x_{n+2}) = x_{n+1}x_{n+2} + x_1 \cdots x_{n+1} + f_1(x_1, \ldots, x_n),$$

and therefore if the ANF of f_1 has an even number of terms so does f. To construct a function $f' \in \mathcal{B}_{n+4}$ with maximum AI from maximum AI function $f \in \mathcal{B}_{n+2}$ the following subfunctions of f' are used,

$$f_1 = f; \;\; f_2 = f + x_1 \cdots x_{n+2}; \;\; f_3 = f; \;\; f_4 = 1 + f + x_1 \cdots x_{n+2}.$$

Note that the nonlinearity of the functions in this class is well approximated by the bent concatenation bound if the recursion is initiated by $f_1 \in \mathcal{B}_n$ such that $\mathcal{N}_{f_1} = 2^{n-1} - 2^{\frac{n-1}{2}}$. In the worst case after the first iteration $\mathcal{N}_f = 2^{n+1} - 2^{\frac{n+1}{2}} - 2$ for the function $f \in \mathcal{B}_{n+2}$. Then after some i iterative steps, the nonlinearity of $f^{(i)} \in \mathcal{B}_{n+2i}$ in the worst case is given by $\mathcal{N}_{f^{(i)}} \geqslant 2^{n'-1} - 2^{\frac{n'-1}{2}} - 2^i$ for $n' = n + 2i$.

Example 1 *Let us take a balanced $f_1 \in \mathcal{B}_{11}$ with maximum $\mathrm{AI}(f_1) = 6$, and $\mathcal{N}_{f_1} = 992$. Then after two iterations we get $f \in \mathcal{B}_{15}$ satisfying $\mathrm{AI}(f) = 8$, $deg(f) = 14$, $\mathcal{N}_f \geqslant 16252$. Note that the bent concatenation bound for $n = 15$ is 16256, while the best known nonlinearity for $f \in \mathcal{B}_{15}$ is $\mathcal{N}_f = 16276$ but such f is not balanced, see [13, 14].*

It is worth noticing that the ANF of functions in this class in general contains a relatively small number of terms. This is especially true if we start with f_1 on a small input space. Then in each step of iteration the number of terms is increased by two. Using the same notation as above the ANF of function $f^{(i)} \in \mathcal{B}_{n+2i}$ will contain $\#\mathrm{ANF}(f_1) + 2i$ terms, where $\#\mathrm{ANF}(f_1)$ denotes the number of terms for the initial function f_1.

For instance, starting with function $f_1 \in \mathcal{B}_7$ that has 30 terms we can generate a maximum AI function in 15 variables containing only 38 terms in its ANF. This is advantageous feature of the construction since Boolean functions are commonly employed as filtering functions in the context of low complexity circuit environment, thus function should have a sparse ANF from the implementation point of view.

Open Problem 1 *Is there any particular class of attacks that might exploit the sparseness of functions in this class ?*

[1] A function $f \in \mathcal{B}_n$ is said to be resilient of order t if and only if $f(x) + l(x)$ is a balanced function for any linear function $l(x)$ having at most t linear terms.

4.2 Other construction possibilities

An alternative construction method is based on a similar modification of the following design $f^{(2)} = f_1||1 + f_1||1 + f_1||f_1$. The difference to the method given in Construction 1 is that the function $f^{(2)}$ is $(t+2)$-resilient if f_1 is t-resilient and nonlinearity of $f^{(2)}$ is given by $\mathcal{N}_{f^{(2)}} = 4\mathcal{N}_{f_1}$ which is strictly less than $2^n + 2\mathcal{N}_{f_1}$. Thus the higher resiliency order is traded-off against lower nonlinearity.[2]

Construction 2 *Let $f_1 \in \mathcal{B}_n$ be a function with maximum AI , and let the ANF of f_1 contain even number of terms. Then the function $f = f_1||f_2||f_3||f_4$ is also optimized AI function for the following choice of f_2, f_3 and f_4:*

$$f_2 = 1 + f_1 + x_1x_2\cdots x_n; \qquad f_3 = 1 + f_1; \quad f_4 = f_1 + x_1x_2\cdots x_n.$$

Moreover, the ANF of f is given by,

$$f(x_1, \ldots, x_{n+2}) = x_{n+2} + x_1\cdots x_{n+1} + f_1(x_1, \ldots, x_n).$$

Proof. We give a somewhat shortened proof that $deg(g) \geqslant \lceil \frac{n}{2} \rceil + 1$ for any nonzero annihilator g of either f or $1 + f$. Since by Lemma 3 the AI is invariant under permutation of subfunctions we consider f' defined by,

$$f'_1 = f_1; \quad f'_2 = f_1 + x_1x_2\cdots x_n; \qquad f'_3 = 1 + f_1; \quad f'_4 = 1 + f_1 + x_1x_2\cdots x_n.$$

From $f'_1g_1 = 0$ and $f'_2g_2 = 0$ we have,

$$f_1(x)(g_1(x) + g_2(x)) = x_1x_2\cdots x_ng_2(x) = 0 \quad \text{or} \quad x_1\cdots x_n, \tag{7}$$

depending on the parity of the ANF of g_2.

i) Taking $g_1 = 0$ implies $g_3 = 0$. Thus if $g_1 = g_3 = 0$ then $deg(g_2) < \lceil \frac{n}{2} \rceil$. Then the both case in (7) gives a contradiction, similarly as in the proof of Construction 1. Thus $g_1 = 0$ leads to $g_1 = g_2 = g_3 = 0$ and $g_4 \neq 0$. By Theorem 1, $\mathrm{AI}(f_4) \geqslant \lceil \frac{n}{2} \rceil - 1$, and therefore $deg(g) \geqslant \lceil \frac{n}{2} \rceil + 1$ due to term $x_{n+2}x_{n+1}g_4(x)$.

ii) Clearly, $g_1 \neq 0$ implies $g_2 \neq 0$ if $deg(g) < \lceil \frac{n}{2} \rceil + 1$. Also we must have $deg(g_1+g_2) < \lceil \frac{n}{2} \rceil$, as otherwise $deg(g) \geqslant \lceil \frac{n}{2} \rceil + 1$. The case $f_1(x)(g_1(x)+g_2(x)) = 0$ in equation (7), gives a contradiction. Thus, we must have

$$f_1(x)(g_1(x) + g_2(x)) = x_1x_2\cdots x_n,$$

contradicting the assumption on the parity of ANF of f_1.
It remains to prove that the same is true for annihilators of $1 + f'$. Since $1 + f'$ is only a permutation of the subfunctions of f', by Lemma 3 the minimum degree of annihilators of $1 + f'$ is the same as for f'. The same is true for original function f. The choice of subfunctions of f gives the ANF of f as stated. □

The function $f \in \mathcal{B}_{n+2}$ defined in Construction 2 is a balanced function of maximum degree $deg(f) = n + 1$. Also, assuming that the nonlinearity of f_1 is $\mathcal{N}_{f_1}$, then $\mathcal{N}_f \in \{4\mathcal{N}_{f_1}, 4\mathcal{N}_{f_1} \pm 2\}$. The necessary conditions for the recursive use of the

[2] We cannot speak about such a trade-off for function f in Construction 2 as f is not a resilient function in a strict sense. Nevertheless, $\varepsilon = 0.5 + \frac{1}{2^{n-1}}$ for any $f(x) + l(x)$, $l(x)$ being linear function of at most $t + 2$ terms.

construction are satisfied by noting that the ANF of f_1 of even parity implies even parity of,

$$f(x_1, \ldots, x_{n+2}) = x_1 \cdots x_{n+1} + x_{n+2} + f_1(x_1, \ldots, x_n).$$

Note that there is no restriction on the evenness of n in Construction 2. We believe that the design methods of functions with maximum AI described above do not exhaust the possibilities of finding more good classes based on a similar approach.

5 A fast algorithm for finding annihilators of small degree

Security estimates for stream cipher schemes that employ nonlinear filtering of a single or several LFSR's strongly depend on the choice of nonlinear function $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Due to time-memory trade-off attacks for these schemes the state size of the cipher is commonly taken to be twice as large as key length. For a standard key length $k = 128$ and the state size $S = 256$, an application of algebraic attack (approximating the complexity of the attack as $\binom{S}{AI(f)}^3$) is less than exhaustive key search for $\mathrm{AI}(f) \leqslant 7$. Noticing that $\mathrm{AI}(f) \leqslant \lceil \frac{n}{2} \rceil$, this implies that the input space of Boolean function must be at least $n = 15$.

The basic approach to check the existence of annihilators of degree d is to form the matrix of size $supp(f) \times \sum_{i=0}^{d} \binom{n}{i}$ (see [8]), where the columns of this matrix correspond to evaluation of all monomials of degree up to d restricted to $supp(f)$,

$$1, x_1, \ldots, x_n, x_1x_2, \ldots, x_1 \cdots x_d, \ldots, x_{n-d+1} \cdots x_d, \quad x \in supp(f).$$

The Gaussian elimination on the above defined matrix for a balanced function f induces the complexity of

$$2^{n-1}\Big(\sum_{i=0}^{d} \binom{n}{i}\Big)^2,$$

which for the critical value $d = 7$ gives "infeasible" computational complexity $> 2^{50}$ for $n > 18$.

In [11], two algorithms (called Algorithm 1 and Algorithm 2) were proposed for finding low degree annihilators. The both algorithms are faster than the straightforward approach based on Gaussian elimination. The estimated computational complexity for Algorithm 2 is of order $\frac{1}{8}\binom{n}{d}^3$ to decide the existence of annihilators of degree at most d (except for the cases $d \leqslant 5$ with somewhat improved performance). Here again, the computational complexity for $d = 7$ becomes larger than 2^{50} for $n > 22$.

Any function g, being an annihilator of $f : \mathbb{F}_2^n \to \mathbb{F}_2$, may be viewed as concatenation of subfunctions $g_{[\tau]}$,

$$g(y, x) = \sum_{\tau \in \mathbb{F}_2^{n-k}} \Big(\prod_{i=1}^{n-k} (y_i + \tau_i + 1)\Big) g_{[\tau]}(x), \tag{8}$$

where $g_{[\tau]} : \mathbb{F}_2^k \to \mathbb{F}_2$ is any annihilator of $f_{[\tau]}$ for f represented as,

$$f(y, x) = \sum_{\tau \in \mathbb{F}_2^{n-k}} \Big(\prod_{i=1}^{n-k} (y_i + \tau_i + 1)\Big) f_{[\tau]}(x). \tag{9}$$

The next proposition makes the basis for a novel algorithm for determining the existence of annihilators to be described later.

Proposition 2 *Any Boolean function $f \in \mathcal{B}_n$ is a degree non-decreasing function with respect to the degrees of its subfunctions. That is, for any $1 \leqslant k \leqslant n-1$ and,*

$$f(y,x) = \sum_{\tau \in \mathbb{F}_2^{n-k}} \big(\prod_{i=1}^{n-k} (y_i + \tau_i + 1) \big) f_{[\tau]}(x); \;\; (y,x) \in \mathbb{F}_2^{n-k} \times \mathbb{F}_2^k,$$

we have the following relation,

$$deg(f) \geqslant \max_{\tau \in \mathbb{F}_2^{n-k}} deg(f_{[\tau]}).$$

Proof. Let $k = n-1$. This means that $f = f_1 || f_2$, that is,

$$f(x_1, \ldots, x_n) = x_n(f_1(x) + f_2(x)) + f_1(x),$$

where x stands for $(x_1, \ldots, x_{n-1})$. It is easily verified that $deg(f) \geqslant \max_{i=1,2} deg(f_i)$. But the same is true for f_1 and its subfunctions, holding also for f_2 as well. Thus, however we decompose f into subfunctions from a smaller space the assertion above always holds. □

Corollary 1 *Let $f \in \mathcal{B}_n$ be any Boolean function. If f admits annihilators of degree d, then all subfunctions of f defined on some smaller variable space must admit annihilators of degree at most d.*

The interpretation of this result is that nonexistence of annihilators of degree at most d for any subfunction of f (these subfunctions defined on a smaller variable space) implies the nonexistence of annihilators of degree d for function f as well. Computer simulations suggest that only a negligible small fraction of functions have algebraic immunity less than $\lceil \frac{n}{2} \rceil - 1$. Thus, a primary goal of the algorithm below is to confirm the nonexistence of annihilators of degree $\leqslant d$ when d is significantly less than $\lceil \frac{n}{2} \rceil$. For practical applications the case $d = 7$ is of special importance as in this case the cipher is protected against standard algebraic attacks[3].

Let $p_i < 1$ denote the probability that a k-variable subfunction f_i of $f \in \mathcal{B}_n$ has an annihilator of degree $\leqslant d$, where $d < \lceil \frac{k}{2} \rceil$. Then the total probability that all 2^{n-k} subfunctions admit annihilators of degree $\leqslant d$ becomes $p_i^{2^{n-k}}$ which for reasonably small p_i tends to zero. This immediately gives a faster method than known algorithms for d relatively small compared to n, and in particular for $d = 7$. For instance, "infeasible" computational complexity of 2^{50} for $n = 22$ and $d = 7$ applying the Algorithm 2 in [11] becomes,

$$Compl. = 2^7 \times \frac{1}{8} \binom{15}{7}^3 = 2^{42},$$

when the same algorithm is repeatedly applied to 2^7 subfunctions of f, each subfunction being a 15-variable function.

[3] It might be a good idea to consider slightly larger d, e.g. $d = 8$ to introduce protection against fast algebraic attacks [7, 1].

Note that the estimate above corresponds to the worst case scenario, as in the best case our algorithm may terminate after only checking the first subfunction resulting in the best case complexity,

$$Compl. = \frac{1}{8}\binom{15}{7}^3 = 2^{35}.$$

However if algorithm does not terminate, in a sense that all subfunctions admit annihilators of degree $\leqslant d$, it simply increases $k \leftarrow k+1$ and the same procedure is repeated. For the same example above it is quite unlikely that any 17-variable function admits annihilators of degree $d = 7$. Then the complexity of the algorithm is most likely to be,

$$Compl. = \frac{1}{8}\binom{17}{7}^3 = 2^{39}.$$

Figure 1 summarizes the formal steps of the algorithm. Note that the algorithm is most likely to terminate without increasing k in step 6.

Input: A function $f \in \mathcal{B}_n$, and annihilator degree $d < \lceil \frac{n}{2} \rceil$.

1. Set $k = 2d + 1$, and $i = 0$.
2. While $i < 2^{n-k} - 1$, let $(x_{k+1}^{(i)}, \ldots, x_n^{(i)}) \in \mathbb{F}_2^{n-k}$ be a binary representation of i.
3. Check whether $f(x_1, \ldots, x_k, x_{k+1}^{(i)}, \ldots, x_n^{(i)})$ admits nontrivial annihilation of degree $\leqslant d$.
4. If NO, OUTPUT: No annihilators of degree $\leqslant d$, TERMINATE.
5. If YES, go to 2 and increase $i \leftarrow i + 1$.
6. If $i = 2^{n-k} - 1$ then go to 1 and $k \leftarrow k + 1$.
7. OUTPUT: There are annihilators of degree $\leqslant d$.

Figure 1: Fast annihilator algorithm.

6 Conclusions

In this paper we have addressed two important issues in theory of algebraic attacks. We have set out a method for designing Boolean functions, which for the first time unifies all important cryptographic criteria. The only criterion that is not covered by the construction is resiliency. Nevertheless, apart from the fact that resiliency is not decisive criterion for certain stream cipher schemes, correlation coefficient is shown to be extremely small $\varepsilon = 0.5 + 1/2^{n-1}$ which for reasonably large n (say $n > 16$) makes the correlation attacks quite likely impractical. We believe that the technique presented here may be further developed to yield more classes of functions with maximum AI.

References

[1] F. Armknecht. Improving fast algebraic attacks. In *Fast Software Encryption 2004*, volume LNCS 3017, pages 65–82. Springer-Verlag, 2004.

[2] A. Braeken and B. Preneel. On the algebraic immunity of symmetric Boolean functions. Accepted at Indocrypt 2005.

[3] P. Camion, C. Carlet, P. Charpin, and N. Sendrier. On correlation-immune functions. In *Advances in Cryptology—EUROCRYPT'91*, volume LNCS 547, pages 86–100. Springer-Verlag, 1991.

[4] A. Canteaut. Invited talk: Open problems related to algebraic attacks stream ciphers. In *Proceedings of the 2005 International Workshop on Coding and Cryptography (WCC 2005), Bergen, Norway*, To appear in LNCS, 2005.

[5] C. Carlet. Improving the algebraic immunity of resilient and nonlinear functions and constructing bent functions. Cryptology ePrint Archive, Report 2004/276, 2002. http://eprint.iacr.org/.

[6] N. Courtois. Higher order correlation attacks, XL algorithm and cryptoanalysis of Toyocrypt. In *ICISC 2002*, volume LNCS 2587, pages 182–199. Springer-Verlag, 2002.

[7] N. Courtois. Fast algebraic attacks on stream ciphers with linear feedback. In *Advances in Cryptology—CRYPTO 2003*, volume LNCS 2729, pages 176–194. Springer-Verlag, 2003.

[8] N. Courtois and W. Meier. Algebraic attacks on stream ciphers with linear feedback. In *Advances in Cryptology—EUROCRYPT 2003*, volume LNCS 2656, pages 346–359. Springer-Verlag, 2003.

[9] D. K. Dalai, K. C. Gupta, and S. Maitra. Significant Boolean functions: construction and analysis in terms of algebraic immunity. In *Fast Software Encryption 2005*, volume LNCS 3557, pages 98–111. Springer-Verlag, 2005.

[10] K. D. Dalai, S. Maitra, and S. Sarkar. Basic theory in construction of Boolean functions with maximum annihilator immunity. Cryptology ePrint Archive, Report 2005/229.

[11] W. Meier, E. Pasalic, and C. Carlet. Algebraic attacks and decomposition of Boolean functions. In *Advances in Cryptology—EUROCRYPT 2004*, volume LNCS 3027, pages 474–491. Springer-Verlag, 2004.

[12] A. Menezes, P. van Oorschot, and S. Vanstone. *Handbook of Applied Cryptography*. CRC Press, Boca Raton, 1997.

[13] N. J. Patterson and D. H. Wiedemann. The covering radius of the $(2^{15}, 16)$ Reed-Muller code is at least 16276. *IEEE Trans. on Inform. Theory*, IT-29(3):354–356, 1983.

[14] N. J. Patterson and D. H. Wiedemann. Correction to – the covering radius of the $(2^{15}, 16)$ Reed-Muller code is at least 16276. *IEEE Trans. on Inform. Theory*, IT-36(2):443, 1990.

[15] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, Vol. 27:379–423 (Part I) and 623–656 (Part II), 1948.

[16] T. Siegenthaler. Correlation-immunity of nonlinear combining functions for cryptographic applications. *IEEE Trans. on Inform. Theory*, IT-30:pages 776–780, 1984.