Cryptographically Significant Boolean functions: Construction and Analysis in terms of Algebraic Immunity

Deepak Kumar Dalai, Kishan Chand Gupta and Subhamoy Maitra

Applied Statistics Unit, Indian Statistical Institute,
203, B T Road, Calcutta 700 108, INDIA
{deepak_r, kishan_t, subho}@isical.ac.in

Abstract. Algebraic attack has recently become an important tool in cryptanalysing different stream and block cipher systems. A Boolean function, when used in some cryptosystem, should be designed properly to resist this kind of attack. The cryptographic property of a Boolean function, that resists algebraic attack, is known as Algebraic Immunity ($\mathcal{AI}$). So far, the attempt in designing Boolean functions with required algebraic immunity was only ad-hoc, i.e., the functions were designed keeping in mind the other cryptographic criteria, and then it has been checked whether it can provide good algebraic immunity too. For the first time, in this paper, we present a construction method to generate Boolean functions on n variables with highest possible algebraic immunity $\lceil \frac{n}{2} \rceil$. Such a function can be used in conjunction with (using direct sum) functions having other cryptographic properties.
In a different direction we identify that functions, having low degree subfunctions, are weak in terms of algebraic immunity and analyse some existing constructions from this viewpoint.

Keywords: Algebraic Attacks, Algebraic Immunity, Annihilators, Boolean Functions, Correlation Immunity, Nonlinearity.

1 Introduction

Recent literature shows that algebraic attack has gained a lot of attention in cryptanalysing stream and block cipher systems. The attack uses overdefined systems of multivariate equations to recover the secret key [1, 2, 10–14, 18, 17]. Given a Boolean function f on n-variables, different kinds of scenarios related to low degree multiples of f have been studied in [13, 18]. The core of the analysis is to find out minimum (or low) degree annihilators of f and $1 + f$, i.e., to find out minimum (or low) degree functions g_1, g_2 such that $f * g_1 = 0$ and $(1 + f) * g_2 = 0$. To mount the algebraic attack, one needs only the low degree linearly independent annihilators [13, 18] of $f, 1 + f$.

So far very little attempt has been made to provide construction of Boolean functions that can resist algebraic attacks. In [15], some existing construction methods have been analysed that can provide Boolean functions with some other cryptographic properties to see how good they are in terms of algebraic immunity.

Algebraic immunity of certain constructions have also been studied in [3–5]. In [5], the authors have proved that the algebraic immunity of the n-variable functions constructed by Tarannikov's method [21, 19] attain $\mathbf{\Omega}(\sqrt{n})$ algebraic immunity. This presents a sharper result than what presented in [15] in terms of analysing Tarannikov's construction [21, 19]. Construction of cryptographically significant Boolean functions with improved algebraic immunity has also been presented in [7].

However, so far there is no existing construction method that can achieve maximum possible algebraic immunity. In this paper, for the first time, we provide a construction method where the algebraic immunity is the main concern. We show that given a Boolean function on $n - 2d$ variables having algebraic immunity 1, we can always construct a Boolean function on n variables with algebraic immunity at least $d + 1$. The construction is iterative in nature (a function with two more variables is constructed in each step) and we need to apply it d times to get an n-variable function from an $(n - 2d)$-variable initial function. We also show that the construction preserves the order of resiliency and increases the nonlinearity by more than 2^{2d} times in d-steps (as it can be seen as a direct sum of a function with good nonlinearity and resiliency with another function with good algebraic immunity). Also using our construction one can generate n-variable functions with highest possible algebraic immunity $\lceil \frac{n}{2} \rceil$ and good nonlinearity. For this one needs to start with 1 or 2-variable nonconstant functions.

Further, in a different direction, we show that if a Boolean function has low degree subfunctions then it is not good in terms of algebraic immunity. This result generalizes the analysis on Maiorana-McFarland type functions presented in [18]. Further our analysis answers some of the questions presented in [15] regarding the algebraic immunity of the functions presented in [20].

2 Preliminaries

A Boolean function on n variables may be viewed as a mapping from $\{0, 1\}^n$ into $\{0, 1\}$ and define B_n as the set of all n-variable Boolean functions. One of the standard representation of a Boolean function $f(x_1, \ldots, x_n)$ is by the output column of its *truth table*, i.e., a binary string of length 2^n,

$$f = [f(0, 0, \ldots, 0), f(1, 0, \ldots, 0), f(0, 1, \ldots, 0), \ldots, f(1, 1, \ldots, 1)].$$

The set of $x \in \{0, 1\}^n$ for which $f(x) = 1$ (respectively $f(x) = 0$) is called the onset (respectively offset), denoted by 1_f (respectively 0_f). We say that a Boolean function f is balanced if the truth table contains an equal number of 1's and 0's.

The Hamming weight of a binary string S is the number of ones in the string. This number is denoted by $wt(S)$. The Hamming distance between two strings,

S_1 and S_2 is denoted by $d(S_1, S_2)$ and is the number of places where S_1 and S_2 differ. Note that $d(S_1, S_2) = wt(S_1 + S_2)$ (by abuse of notation, we also use $+$ to denote the $GF(2)$ addition, i.e., the XOR). By $S_1||S_2$ we mean the concatenation of two strings. By $\overline{S}$ we mean the complement of the string S.

Any Boolean function has a unique representation as a multivariate polynomial over $GF(2)$, called the algebraic normal form (ANF),

$$f(x_1, \ldots, x_n) = a_0 + \sum_{1 \le i \le n} a_i x_i + \sum_{1 \le i < j \le n} a_{i,j} x_i x_j + \ldots + a_{1,2,\ldots,n} x_1 x_2 \ldots x_n,$$

where the coefficients $a_0, a_{i,j}, \ldots, a_{1,2,\ldots,n} \in \{0, 1\}$. The algebraic degree, $\deg(f)$, is the number of variables in the highest order term with non zero coefficient. A Boolean function is affine if there exists no term of degree > 1 in the ANF and the set of all affine functions is denoted A_n. An affine function with constant term equal to zero is called a linear function.

It is known that a Boolean function should be of high algebraic degree to be cryptographically secure [16]. Further, it has been identified recently, that it should not have a low degree multiple [13]. The algebraic attack (see [13, 18] and the references in these papers) is getting a lot of attention recently. To resist algebraic attacks, the Boolean functions used in the cryptosystems should be chosen properly. It is shown [13] that given any n-variable Boolean function f, it is always possible to get a Boolean function g with degree at most $\lceil \frac{n}{2} \rceil$ such that $f * g$ is of degree at most $\lceil \frac{n}{2} \rceil$. Here the functions are considered to be multivariate polynomials over $GF(2)$ and $f * g$ is the polynomial multiplication over $GF(2)$. Thus while choosing an f, the cryptosystem designer should be careful that it should not happen that degree of $f * g$ falls much below $\lceil \frac{n}{2} \rceil$.

Towards defining algebraic immunity [13, 18, 15], one needs to consider the multiples of both f and $1 + f$.

1. Take $f, g, h \in B_n$. Assume that there exists a nonzero function g of low degree such that $f * g = h$ or $(1 + f) * g = h$, where h is a nonzero function of low degree and without loss of generality, $\deg(g) \le \deg(h)$. Among all such h's we denote the lowest degree h (may be more than one and then we take any one of them) by $ldgm_n(f)$.
2. Assume there exists a nonzero function g of low degree such that $f * g = 0$ or $(1 + f) * g = 0$. Among all such g's we denote the lowest degree g (may be more than one and then we take any one of them) by $ldga_n(f)$.

It can be checked that [18, 15] for $f \in B_n$, $\deg(ldgm_n(f)) = \deg(ldga_n(f))$ and in this line the following definition of algebraic immunity has been presented in [15].

Definition 1. *The algebraic immunity of an n-variable Boolean function f is denoted by $\mathcal{AI}_n(f)$ which is basically* $\deg(ldgm_n(f))$ *or* $\deg(ldga_n(f))$.

Later we also need the following definition related to the annihilator set of a function.

Definition 2. *Given $f \in B_n$, define $AN(f) = \{g \in B_n | g \text{ nonzero}, \; f * g = 0\}$.*

The nonlinearity of an n-variable function f is the minimum distance from the set of all n-variable affine functions, i.e.,

$$nl(f) = \min_{g \in A(n)} (d(f, g)).$$

Boolean functions used in crypto systems must have high nonlinearity to prevent linear attacks [16].

Many properties of Boolean functions can be described by the Walsh transform. Let $x = (x_1, \ldots, x_n)$ and $\omega = (\omega_1, \ldots, \omega_n)$ both belonging to $\{0,1\}^n$ and $x \cdot \omega = x_1\omega_1 + \ldots + x_n\omega_n$. Let $f(x)$ be a Boolean function on n variables. Then the *Walsh transform* of $f(x)$ is an integer valued function over $\{0,1\}^n$ which is defined as

$$W_f(\omega) = \sum_{x \in \{0,1\}^n} (-1)^{f(x) + x \cdot \omega}.$$

A Boolean function f is balanced iff $W_f(0) = 0$. The nonlinearity of f is given by $nl(f) = 2^{n-1} - \frac{1}{2} \max_{\omega \in \{0,1\}^n} |W_f(\omega)|$. Correlation immune functions and resilient functions are two important classes of Boolean functions. A function is m-resilient (respectively mth order correlation immune) iff its Walsh transform satisfies

$$W_f(\omega) = 0, \text{ for } 0 \le wt(\omega) \le m \text{ (respectively } 1 \le wt(\omega) \le m).$$

The paper is organized as follows. In the next section we present the construction and the following section discusses the analysis of algebraic immunity of a function in terms of the degree of its subfunctions.

3 Construction to get $\mathcal{AI}$ as required

In this section we present a construction to get Boolean function of $n+2$ variables with algebraic immunity $d+2 \le \lceil \frac{n+2}{2} \rceil$. The construction is iterative in nature and it starts from an initial function of $n+2-2(d+1) = n-2d$ variables having algebraic immunity 1 (the minimum possible value). In each step, 2 variables are added and algebraic immunity gets increased by 1. Let us now formalize the construction.

Construction 1 *Let $f \in B_n$ such that $f = E||F||G||H$ where $E, F, G, H \in B_{n-2}$. Let $n - 2d > 0$ and $d \ge 0$. Take an initial function $f_{n-2d} \in B_{n-2d}$ with $\mathcal{AI}_{n-2d}(f_{n-2d}) = 1$. Suppose after d-th step $f_n \in B_n$ has been constructed. The next function $f_{n+2} \in B_{n+2}$ is constructed in following manner:*

$$f_{n+2} = f_n||f_n||f_n||f_n^1, \text{ where } f^k = E^{k-1}||F^k||G^k||H^{k+1},$$

for any function f_j,

$$f_j^0 = f_j,$$

and for the initial function f_{n-2d},

$$f_{n-2d}^s = \overline{f}_{n-2d} \text{ (and } \overline{f}_{n-2d}^s = f_{n-2d}),$$

for $s > 0$.

To understand the recursion in the Construction 1, we present an example up to some depths.

- $f_n^1 = f_{n-2}||f_{n-2}^1||f_{n-2}^1||(f_{n-2}^1)^2$ as
  $f_n = f_{n-2}||f_{n-2}||f_{n-2}||f_{n-2}^1$,
- $(f_{n-2}^1)^2 = f_{n-4}^1||(f_{n-4}^1)^2||(f_{n-4}^1)^2||((f_{n-4}^1)^2)^3$ as
  $f_{n-2}^1 = f_{n-4}||f_{n-4}^1||f_{n-4}^1||(f_{n-4}^1)^2$,
- $((f_{n-4}^1)^2)^3 = (f_{n-6}^1)^2||((f_{n-6}^1)^2)^3||((f_{n-6}^1)^2)^3||(((f_{n-6}^1)^2)^3)^4$ as
  $(f_{n-4}^1)^2 = f_{n-6}^1||(f_{n-6}^1)^2||(f_{n-6}^1)^2||((f_{n-6}^1)^2)^3$.

This goes on unless we reach at the level of the $(n-2d)$-variable initial function. For $m \geq 2$, denote $((f^1)^{2\cdots})^m$ as $f^{1,m}$. As example, $((f_{n-6}^1)^2)^3 = f_{n-6}^{1,3}$. Also, for notational consistency, we take $(f^1)^1 = f^1$ and $f^{1,0} = f^0 = f$.

Take an initial function f_l (the 2^l length binary string which is the truth table of the function) on l variables. Below we present the construction idea as truth table concatenation.

Step 1: $f_{l+2} = f_l f_l f_l \overline{f}_l$

Step 2: $f_{l+4} = f_{l+2} f_{l+2} f_{l+2} f_l \overline{f}_l \overline{f}_l f_l = f_l f_l f_l \overline{f}_l f_l f_l f_l \overline{f}_l f_l f_l f_l \overline{f}_l f_l \overline{f}_l \overline{f}_l f_l$

Step 3: $f_{l+6} = f_{l+4} f_{l+4} f_{l+4} f_{l+2} f_l \overline{f}_l \overline{f}_l f_l f_l \overline{f}_l \overline{f}_l f_l \overline{f}_l f_l f_l \overline{f}_l$

Step 4: $f_{l+8} = f_{l+6} f_{l+6} f_{l+6} f_{l+4} f_{l+2} f_l \overline{f}_l \overline{f}_l f_l f_l \overline{f}_l \overline{f}_l f_l \overline{f}_l f_l f_l \overline{f}_l f_{l+2} f_l \overline{f}_l \overline{f}_l f_l$
$f_l \overline{f}_l \overline{f}_l f_l \overline{f}_l f_l f_l \overline{f}_l f_l \overline{f}_l \overline{f}_l f_l \overline{f}_l f_l f_l \overline{f}_l \overline{f}_l f_l f_l \overline{f}_l f_l \overline{f}_l \overline{f}_l f_l$

Thus after the k-th step, the function f_{l+2k} is the concatenation of 2^{2k} numbers of f_l and $\overline{f_l} = 1 + f_l$. That is, the subfunctions of f_{l+2k} at $2k$-depth are only f_l and $\overline{f_l}$. That is, f_{l+2k} can be seen as direct sum of f_l and a $2k$-variable function.

To prove the main theorem we first present the following results. In the proofs we will use the fact that for any $f \in B_n$ and any subset $V \in \{0,1\}^n$, the restriction of any annihilator g of f to V is an annihilator of the restriction of f to V. For technical reasons (see also Remark 1 after the proof of Lemma 2), during our proofs we will encounter certain situations when degree of a function is < 0. As such functions cannot exist, we will replace those functions by 0 (function).

Lemma 1. *Consider that the function $f_{n+2} \in B_{n+2}$ has been generated by Construction 1 after $(d+1)$ many steps, $d \geq 1$, taking f_{n-2d} as the initial function. Take $g, h \in B_{n-2}$. We assume that if $g' \in AN(f_{n-2d+2j})$, $h' \in AN(f_{n-2d+2j}^1)$ for $0 \leq j \leq d-1$ and $\deg(g'+h') \leq j-1$, then $g' = h' = 0$. If*

1. $g \in AN(f_{n-2}^{1,i})$ *and* $h \in AN(f_{n-2}^{1,i+1})$ *for any* i, $i \geq 1$ *and*
2. $\deg(g+h) \leq d-2-i$,

then $g = h = 0$.

Proof. We prove it by induction. For the base step $d = 1$. Here $\deg(g+h) \leq 1-2-i \leq -2$ implies such a function cannot exist (see also Remark 1), i.e., $g+h$ is identically 0, which gives $g = h$.

Now $g \in AN(f_{n-2d}^{1,i})$ and $h \in AN(f_{n-2d}^{1,i+1})$. Since f_{n-2d} is the initial function, by Construction 1, $f_{n-2d}^{1,i+1} = (f_{n-2d}^{1,i})^{i+1} = \overline{f_{n-2d}^{1,i}}$. Hence $g \in AN(f_{n-2d}^{1,i})$ and

$h \in AN(\overline{f^{1,i}_{n-2d}})$. Thus g, h, being nonzero, cannot be same. So $g = h = 0$. This proves the base step.

Now we prove the inductive step. Consider that the function $f_n \in B_n$ has been generated by Construction 1 after d many steps, $d \geq 1$, taking f_{n-2d} as the initial function. For any $g', h' \in B_{n-4}$ with $g' \in AN(f^{1,i}_{n-4})$ and $h' \in AN(f^{1,i+1}_{n-4})$ and for any i, $i \geq 1$, if $\deg(g' + h') \leq (d-1) - 2 - i$, then $g' = h' = 0$.

Suppose that f_{n+2} is constructed by Construction 1 and there exists $g \in AN(f^{1,i}_{n-2})$ and $h \in AN(f^{1,i+1}_{n-2})$ with $\deg(g + h) \leq d - 2 - i$. By construction, we have

$$f^{1,i}_{n-2} = f^{1,i-1}_{n-4}||f^{1,i}_{n-4}||f^{1,i}_{n-4}||f^{1,i+1}_{n-4},$$
$$f^{1,i+1}_{n-2} = f^{1,i}_{n-4}||f^{1,i+1}_{n-4}||f^{1,i+1}_{n-4}||f^{1,i+2}_{n-4}.$$

Take,

$$g = v_1||v_2||v_3||v_4,$$
$$h = v_5||v_6||v_7||v_8,$$

This gives, $v_1 \in AN(f^{1,i-1}_{n-4}), v_2, v_3, v_5 \in AN(f^{1,i}_{n-4}), v_4, v_6, v_7 \in AN(f^{1,i+1}_{n-4})$ and $v_8 \in AN(f^{1,i+2}_{n-4})$. Since $\deg(g + h) \leq d - 2 - i$, from ANF of $g + h = (v_1 + v_5) + x_{n-3}(v_1 + v_5 + v_2 + v_6) + x_{n-2}(v_1 + v_5 + v_3 + v_7) + x_{n-3}x_{n-2}(v_1 + \cdots + v_8)$ we deduce the following.

- $\deg(v_1 + v_5) \leq d - 2 - i = (d-1) - 2 - (i-1)$, implying that $v_1 = v_5 = 0$, for $i \geq 2$. For $i = 1$, we have $v_1 \in AN(f_{n-4})$, $v_5 \in AN(f^1_{n-4})$ with $\deg(v_1 + v_5) \leq d - 3$. Following the assumption in the statement of the lemma, we get $v_1 = v_5 = 0$.
- $\deg(v_2 + v_6) \leq d - 3 - i = (d-1) - 2 - i$, implying that $v_2 = v_6 = 0$.
- $\deg(v_3 + v_7) \leq d - 3 - i = (d-1) - 2 - i$, implying that $v_3 = v_7 = 0$.
- $\deg(v_4 + v_8) \leq d - 4 - i = (d-1) - 2 - (i+1)$, implying that $v_4 = v_8 = 0$.

Hence we get $g = h = 0$ for $i \geq 1$. □

Lemma 2. *Consider that $f_{n+2} \in B_{n+2}$ has been generated using the Construction 1 after $(d+1)$-th step with initial function f_{n-2d}. Let $\mathcal{AI}(f_{n-2d+2i}) = i+1$ for $0 \leq i \leq d$. Consider that*

1. *$g_{n+2} \in AN(f_{n+2})$,*
2. *$\deg(g_{n+2}) \leq d + 1$, and*
3. *g_{n+2} is of the form $g_{n+2} = g_n + x_{n+2}x_{n+1}(g_n + h_n)$, where $g_n \in AN(f_n), h_n \in AN(f^1_n)$.*

If $\deg(g_n + h_n) \leq d - 1$, then $g_n = h_n = 0$.

Proof. We will use induction on d. For the base step (i.e., $d = 0$) we have $f^1_n = \overline{f}_n$ as f_n the initial function when $d = 0$. Here g_n and h_n are annihilators of f_n and $f^1_n = \overline{f}_n$ respectively. Since $\deg(g_{n+2}) \leq 1$, and $g_{n+2} = g_n + x_{n+2}x_{n+1}(g_n + h_n)$,

$g_n + h_n = 0$, which gives $g_n = h_n$. Since $g_n \in AN(f_n)$ and $h_n \in AN(\overline{f}_n)$, being non zero functions, they cannot be same, i.e., $g_n = h_n = 0$. Then $g_{n+2} = 0$.

Now we prove the inductive step. Assume the induction assumption holds till d steps, $d \geq 0$. Now we will prove the lemma statement at $(d+1)$-th step. That is $f_{n+2} \in B_{n+2}$ has been generated by Construction 1 after $(d+1)$-th step with $\mathcal{AI}(f_{n+2}) \leq d+1$ and $\mathcal{AI}(f_{n-2d+2i}) = i+1$ for $0 \leq i \leq d$. Here $g_{n+2} = g_n + x_{n+2}x_{n+1}(g_n + h_n) \in AN(f_{n+2})$ of degree $\leq d+1$, where $g_n \in AN(f_n)$ and $h_n \in AN(f_n^1)$. Suppose $\deg(g_n + h_n) \leq d-1$. Then here, we will prove that $g_n = h_n = 0$. Here

$$
\begin{aligned}
f_n &= f_{n-2}||f_{n-2}||f_{n-2}||f_{n-2}^1, \\
f_n^1 &= f_{n-2}||f_{n-2}^1||f_{n-2}^1||(f_{n-2}^1)^2.
\end{aligned}
$$

Let

$$
\begin{aligned}
g_n &= A||B||C||D, \\
h_n &= E||F||G||H, \text{ where}
\end{aligned}
$$

$A, B, C, E \in AN(f_{n-2})$, $D, F, G \in AN(f_{n-2}^1)$ and $H \in AN((f_{n-2}^1)^2)$. Since $A, E \in AN(f_{n-2})$, we have $A + E \in AN(f_{n-2})$ and hence $\deg(A+E) \geq d$ or $A + E = 0$. Since $\deg(g_n + h_n) \leq d-1$, $A + E = 0$. Then $\deg(B+F) \leq d-2$ and $\deg(C+G) \leq d-2$. Thus, using the induction hypothesis we have $B = C = F = G = 0$. So,

$$
\begin{aligned}
g_n &= A||0||0||D, \\
h_n &= E||0||0||H.
\end{aligned}
$$

So, $\deg(D+H) \leq d-3 = d-2-1$.

We have assumed the inductive steps upto d-th step. That gives that if $g_{n-2d+2i} \in AN(f_{n-2d+2i})$, $h_{n-2d+2i} \in AN(f_{n-2d+2i}^1)$ for $0 \leq i \leq d-1$ and $\deg(g_{n-2d+2i} + h_{n-2d+2i}) \leq i$, then $g_{n-2d+2i} = h_{n-2d+2i} = 0$. Note that, this satisfies the assumption considered in the statement of Lemma 1 and now we can apply it.

Since $D \in AN(f_{n-2}^1)$ and $H \in AN((f_{n-2}^1)^2)$ with $\deg(D+H) \leq d-3$, following Lemma 1 we get $D = H = 0$. So we have $g_n = A||0||0||0$, $h_n = E||0||0||0$, and hence

$$
\begin{aligned}
g_{n+2} &= g_n + x_{n+2}x_{n+1}(g_n + h_n) \\
&= (1 + x_{n-1} + x_n + x_{n-1}x_n)A \\
&\quad + x_{n+1}x_{n+2}((1 + x_{n-1} + x_n + x_{n-1}x_n)(A+E)),
\end{aligned}
$$

i.e., $g_{n+2} = (1 + x_{n-1} + x_n + x_{n-1}x_n)A$, since $A + E = 0$. Then $\deg(g_{n+2}) \geq d+2$, since $\deg(A) \geq d$ as $A \in AN(f_{n-2})$. As $\deg(g_{n+2}) \leq d+1$, we have $A = 0$. This gives the proof. □

Remark 1. In the proof of Lemma 2 above, if $\deg(D+H) \leq d-3 < 0$, we have $(D+H) = 0$, because here $g_{n+2} = g_n + x_{n+2}x_{n+1}x_nx_{n-1}(D+H)$ and $\deg(g_{n+2}) \leq d+1$. Since there is no negative degree function, we have to take the term $x_{n+2}x_{n+1}x_nx_{n-1}(D+H)$ as 0.

Now we present the main result.

Theorem 1. *Refer to Construction 1. Let the algebraic immunity of the initial function f_{n-2d} be 1. Then after $(d+1)$-th step the algebraic immunity of the constructed function f_{n+2} is $d+2$.*

Proof. We have to prove that any nonzero function g_{n+2} such that $g_{n+2}f_{n+2} = 0$ has degree at least $d+2$. Suppose that such a function g_{n+2} with degree at most $d+1$ exists. Then, g_{n+2} can be decomposed as

$$g_{n+2} = g_n || g'_n || g''_n || h_n,$$

where $g_n, g'_n, g''_n \in AN(f_n)$, and $h_n \in AN(f^1_n)$. The algebraic normal form of g_{n+2} is then

$$\begin{aligned} g_{n+2}(x_1, \ldots, x_{n+2}) = g_n &+ x_{n+1}(g_n + g'_n) + x_{n+2}(g_n + g''_n) \\ &+ x_{n+1}x_{n+2}(g_n + g'_n + g''_n + h_n) \ . \end{aligned}$$

If g_{n+2} has degree at most $d+1$, then $(g_n + g'_n)$ and $(g_n + g''_n)$ have degree at most d. Because both functions lie in $AN(f_n)$ and $\mathcal{AI}(f_n) = d+1$, we deduce that $g_n + g'_n = 0$ and $g_n + g''_n = 0$, which give, $g_n = g'_n = g''_n$. Therefore, $g_{n+2} = g_n + x_{n+1}x_{n+2}(g_n + h_n)$. So, $\deg(g_n + h_n) \leq d-1$. Now following the Lemma 2 we have $g_n = h_n = 0$, that gives, $g_{n+2} = 0$.

Similarly one can check that there cannot be any nonzero annihilator of $1 + f_{n+2}$ having degree $\leq d+1$. This completes the proof. $\square$

Using this Construction 1, one can generate a function on n variables whose algebraic immunity is the highest possible, i.e., $\lceil \frac{n}{2} \rceil$. In this case one has to start from 1 or 2-variable nonconstant function. Then after each step we will get a function on two more variables and the algebraic immunity will increase by 1.

Example 1. First we present the case for odd n. One can start from $f_1 = x_1$.
Step 1: $f_1 = 01$
Step 2: $f_3 = f_1f_1f_1\overline{f}_1 = 01010110$
Step 3: $f_5 = f_3f_3f_3 01101001 = 01010110010101100101011001101001$
Step 4: $f_7 = f_5f_5f_5 01010110011010010110100110010110$
Step 5: $f_9 = f_7f_7f_7f_5 0101011001101001011010011001011001010110$
$01101001011010011001011001101001100101101001011001101001$

Then we present the case for even n. One can start from nonlinear function $f_2 = x_1x_2$ as the initial function.
Step 1: $f_2 = 0001$
Step 2: $f_4 = f_2f_2f_2\overline{f}_2 = 0001000100011110$
Step 3: $f_6 = f_4f_4f_4f_2 111011100001$
Step 4: $f_8 = f_6f_6f_6f_4f_2 111011100001 f_2 1110111000011110000100011110$
Step 5: $f_{10} = f_8f_8f_8f_6f_4f_2 111011100001 f_2 1110111000011110000100011110$
$f_4f_2 111011100001 f_2 1110111000011110000100011110 f_2 11101110$
$0001111000010001111011100001000111100001111011100001$

Note that the algebraic immunity stays the same if a function is subjected to linear transformation on input variables. Thus, taking any function presented in the above example, one can apply linear transformation to get number of functions. Further the nonlinearity and algebraic degree also stays same after linear transformation.

Now we will discuss some other cryptographic properties of the functions generated using Construction 1 after k-th step.

Corollary 1. *Let $f_{d+2k} \in B_{d+2k}$ is constructed by Construction 1 taking $f_d \in B_d$ as the initial function, i.e., $f_{d+2k} = f_d + \phi_{2k}$, the direct sum.*

1. $nl(f_{d+2k}) = 2^d nl(\phi_{2k}) + 2^{2k} nl(f_d) - 2nl(\phi_{2k}) nl(f_d) > 4^k nl(f_d)$.
2. *Let f_d be an r-resilient function. Then f_{d+2k} is also r-resilient.*
3. $\deg(f_{d+2k}) = \max\{\deg(f_d), \deg(\phi_{2k})\}$.

Proof. The proof of item 1 follows from [20, Proposition 1(d)] and the proof of item 2 follows from [20, Proposition 1(c)]. The result related to algebraic degree is also easy to see. □

In Item 1 of Corollary 1 we are using $nl(\phi_{2k})$. We have observed that $nl(\phi_{2k})$ is equal to the number of 1's in its truth table. We have checked that the values of $nl(\phi_{2k})$ are 1, 5, 22, 93, 386, 1586, 6476, 26333 for $k = 1, \ldots, 8$. Using this, here we present the nonlinearity of the functions given in Example 1. The initial function is the $f_1 = x_1$ which is a linear function. So, $nl(f_1) = 0$. Therefore, $nl(f_3) = 2, nl(f_5) = 10, nl(f_7) = 44, nl(f_9) = 186, nl(f_{11}) = 772, nl(f_{13}) = 3172, nl(f_{15}) = 12952, nl(f_{17}) = 52666$. Similarly if one starts with a 5-variable 1-resilient function with nonlinearity 12, one gets a 7-variable 1-resilient function with nonlinearity 56 (as $nl(\phi_2) = 1$), then a 9-variable 1-resilient function with nonlinearity 232 (as $nl(\phi_4) = 5$) and so on. We like to point out once again that the nonlinearity remains very good in this construction and the order of resiliency is also not disturbed as it is a direct sum construction of a function f_d with good properties in terms of nonlinearity and resiliency and a function ϕ_{2k} which is good in terms of algebraic immunity. When the weight (also nonlinearity) of the function ϕ_{2k} is odd, then clearly its algebraic degree is $2k$. We have also checked upto $k = 6$, that when the weight (also nonlinearity) is even then the algebraic degree is $2k - 1$. The exact nonlinearity and algebraic degree of ϕ_{2k} is still open at this stage and we are working on it. Certain ideas in this area have also been provided by Carlet [9].

Note that if one starts with an initial function $f_{n-2d} \in B_{n-2d}$ having algebraic immunity D, it is not guaranteed that after d steps f_n will have algebraic immunity $d + D$; the only guarantee is that it will be $\geq d + 1$ (following similar arguement as in the proof of Theorem 1). It will be interesting to see what is the exact algebraic immunity of f_n.

4 Functions with low degree subfunctions

In this section we discuss why a Boolean function with low degree subfunction is not good in terms of algebraic immunity. This result is a generalization of the result presented in [18], where the authors have shown that certain kind of Maiorana-McFarland constructions are not good in terms of algebraic immunity.

Proposition 1. *Let $f \in B_n$. Let $g \in B_{n-r}$ be a subfunction of $f(x_1, \ldots, x_n)$ after fixing r many distinct inputs $x_{i_1}, \ldots, x_{i_r} \in \{x_1, \ldots, x_n\}$. If the algebraic degree of g is d, then $\mathcal{AI}_n(f) \leq d + r$.*

Proof. Let $x_{i_1}, \ldots, x_{i_r}$ are fixed at the values $a_{i_1}, \ldots, a_{i_r} \in \{0, 1\}$. Thus g is a function on the variables $\{x_1, \ldots, x_n\} \setminus \{x_{i_1}, \ldots, x_{i_r}\}$. It can be checked that $(1 + a_{i_1} + x_{i_1}) \ldots (1 + a_{i_r} + x_{i_r})(1 + g)$ is an annihilator of f. The algebraic degree of $(1 + a_{i_1} + x_{i_1}) \ldots (1 + a_{i_r} + x_{i_r})(1 + g)$ is $d + r$. Thus the result. $\square$

The Maiorana-McFarland construction can be seen as concatenation of affine functions on $n-r$ variables to construct an n-variable functions. Clearly we have affine subfunctions of the constructed function in this case and hence $\deg(g) = 1$ following the notation of Proposition 1. Thus there will be annihilators of degree $1 + r$. Note that if r is small, then one can get annihilators at low degree [18, Theorem 2, Example 1]. This situation for Maiorana-McFarland construction is only a subcase of our proposition. Our result works on any function, it need not be of Maiorana-McFarland type only. We present an example below.

Example 2. Let us consider a 20-variable function, with a subfunction of degree 2 on 17-variables, i.e., we fix 3 inputs. In that case the 20-variable function will have an annihilator at degree $2 + 3 = 5$.

Maiorana-McFarland type of constructions are used in design of resilient functions. One idea in this direction is to concatenate k-variable affine functions (repetition may be allowed) non degenerate on at least $m+1$ variables to generate an m-resilient function f on n-variables. For such a function f, it is easy to find an annihilator g of degree $n - k + 1$ as described in [18]. However, it should be noted that in construction of resilient functions, there are lot of techniques [20] that use concatenation of k-variable affine functions where $k < \frac{n}{2}$. In such a case, the annihilators described in [18, Theorem 2] will be of degree greater than $\frac{n}{2}$ and will not be of practical use as there are other annihilators of degree $\leq \frac{n}{2}$ which are not of the form given in [18, Theorem 2]. We will show that even in such a case, Proposition 1 can provide further insight. We will show that a well known construction of resilient function [20, Theorem 10(b)] on n-variables (n odd) can never achieve the algebraic immunity $\lceil \frac{n}{2} \rceil$. At the best, it can only achieve the value $\lfloor \frac{n}{2} \rfloor$. To explain this construction we briefly present some notations from [20].

Take a bit b and a bit string $s = s_0 \ldots s_{n-1}$. Then the string b AND $s = s'_0 \ldots s'_{n-1}$, where $s'_i = b$ AND s_i. Take two bit strings $x = x_0 \ldots x_{n-1}$ and $y = y_0 \ldots y_{m-1}$. The Kronecker product $x \otimes y = (x_0 \text{ AND } y) \ldots (x_{n-1} \text{ AND } y)$, which is a string of length nm. The direct sum of two bit strings x, y is $x\$y = (x \otimes y^c) \oplus (x^c \otimes y)$, where x^c, y^c are bitwise complement of x, y respectively. As an example presented in [20], if $f = 01$, and $g = 0110$, then $f\$g = 01101001$. Now we present the construction for $(2p+1, 1, 2p-1, 2^{2p} - 2^p)$ function as presented in [20] for $p \geq 4$.

Construction 2 *[20, Theorem 10(b)] Let $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ be the 3-variable linear functions non degenerate on two variables (i.e., the functions $x_1 + x_2, x_2 + x_3, x_1 + x_3, x_1 + x_2 + x_3$) and μ_1, μ_2, μ_3 be the 3-variable linear functions non degenerate on 1 variable (i.e., the functions x_1, x_2, x_3). Let g_i be the concatenation of the 3-variable function μ_i and its complement μ_i^c, for $1 \leq i \leq 3$. That is g_i's are basically 4-variable functions. Let h_1, h_2 be bent functions on $2p - 4$ variables, and h_3, h_4, h_5 be bent functions of $2p - 6$ variables and h_6, h_7 be two strings of lengths $2^{2p-6} + 1$ and $2^{2p-6} - 1$ which are prepared by properly adding and removing 1 bit from the truth table of $(2p - 6)$-variable bent functions respectively. Let f be a concatenation of the following sequence of functions. $h_1\$\lambda_1, h_2\$\lambda_2, h_3\$g_1, h_4\$g_2, h_5\$g_3, h_6\$\lambda_3, h_7\$\lambda_4$. This is a $(2p+1, 1, 2p-1, 2^{2p} - 2^p)$ function.*

Proposition 2. *The $(2p+1)$-variable function presented in Construction 2 has a subfunction of degree at most $p - 1$ when $x_{2p+1} = 0$.*

Proof. Consider the subfunction when $x_{2p+1} = 0$. The subfunction (call it g) in concatenation form is $h_1\$\lambda_1, h_2\$\lambda_2$. Since h_1, h_2 are bent functions on $2p - 4$ variables, they can have algebraic degree at most $p - 2$. Further λ_1, λ_2 are 3-variable linear functions. The algebraic normal form of g is $(1 + x_{2p})(h_1 + \lambda_1) + x_{2p}(h_2 + \lambda_2)$. So the degree of g is $\leq 1 + (p - 2) = p - 1$. $\square$

Theorem 2. *For a function $f \in B_n$ (n odd) generated out of Construction 2, $\mathcal{AI}_n(f) \leq \lfloor \frac{n}{2} \rfloor$.*

Proof. Here $n = 2p+1$. We take $g \in B_{n-1}$, i.e., $r = 1$ according to Proposition 1. Further from Proposition 2, $\deg(g) \leq p - 1 = \frac{n-1}{2} - 1$. Thus, $\mathcal{AI}_n(f) \leq \frac{n-1}{2} - 1 + 1 = \lfloor \frac{n}{2} \rfloor$. $\square$

Thus using our technique we can show that the construction proposed in [20, Theorem 10(b)] can not achieve the maximum possible algebraic immunity $\lceil \frac{n}{2} \rceil$. The maximum value it can achieve is $\leq \lfloor \frac{n}{2} \rfloor$. This can be seen only by Proposition 1 which generalizes the result of [18, Theorem 2, Example 1]. This also answers a question presented in [15, Example 2] for $n = 9$. There Construction 2 has been exploited for $p = 4$ and the functions constructed are as follows.

1. $h_1 = 0000010100110110, h_2 = 0000010100110110, h_3 = 0001, h_4 = 0001, h_5 = 0001, h_6 = 00010, h_7 = 001$. In this case, one gets a $(9, 1, 7, 240)$ function f_1 with $\mathcal{AI}_9(f_1) = 3$.
2. If one changes $h_2 = 0000010100110110$ by $h_2 = 0000010100111001$, then we get a $(9, 1, 7, 240)$ function f_2 with $\mathcal{AI}_9(f_2) = 4$.

The question raised in [15] was why the algebraic immunity of these two function are different? The reason is in the first case the functions h_1, h_2 are same with the ANF $x_1x_3+x_2x_4$. Thus the subfunction g (i.e., $h_1\$\lambda_1, h_2\$\lambda_2$) is a degree 2 function. So the maximum algebraic immunity, according to Proposition 1 can be $2 + 1 = 3$. That is the value achieved in [15]. In the second case, h_1 is different from h_2 and the algebraic degree of g (i.e., $h_1\$\lambda_1, h_2\$\lambda_2$) becomes 3 and it achieves the value $3 + 1 = 4$. Thus Proposition 1 helps in answering this question. It is important to note that this technique can be employed to study the upper bound of algebraic immunity for various constructions by analysing their subfunctions and in particular, directly for the constructions proposed in [20, 6].

It should be noted that the converse of Proposition 1 is not always true. That is, a function having low degree annihilator does not imply it always has some low degree subfunction by fixing a few variables. As example, one may refer to the 5-variable function $f = x_1+x_2+x_2x_4+x_3x_4+(x_2+x_3+x_1x_4+x_2x_4+x_3x_4)x_5$. This function has algebraic immunity 2 and the only annihilator of degree 2 is $1+x_1+x_2+x_1x_4+x_3x_4+(x_2+x_3+x_4)x_5$. If one verifies all possible subfunctions of of f after fixing 1 and 2 variables, it is not possible to get subfunctions of degree 1 and 0 respectively.

It will be interesting to extend our idea on the Boolean functions that can be seen as concatenation of indicators of flats [8].

5 Conclusion

In this paper we study the algebraic immunity of Boolean functions since the property becomes a necessary requirement in Boolean functions to be used as cryptographic primitives. For the first time we present a construction where one can get Boolean functions with maximum possible algebraic immunity. Also the construction can be used in conjunction with Boolean functions with other cryptographic properties to have functions which are suitable for different cryptographic applications. Further we also point out that functions having low degree subfunctions are not good in terms of algebraic immunity and study some well known existing constructions from this approach.

Acknowledgment: The authors like to thank the anonymous reviewers for their excellent comments that improved both the technical and editorial quality of this paper. In particular, the current proof of Theorem 1 has been outlined by one of the reviewers and it looks more compact than our long initial proof using systems of homogeneous linear equations. We also like to thank Prof. Claude Carlet for carefully reading our construction and providing a correction in the expression of f_{l+8}.

References

1. F. Armknecht. Improving Fast Algebraic Attacks In *FSE 2004*, number 3017 in Lecture Notes in Computer Science, pages 65–82. Springer Verlag, 2004.
2. L. M. Batten. Algebraic Attacks over GF(q). In *Progress in Cryptology - INDOCRYPT 2004*, pages 84–91, number 3348, Lecture Notes in Computer Science, Springer-Verlag.
3. A. Botev. On algebraic immunity of some recursively given sequence of correlation immune functions. In Proceedings of *XV international workshop on Synthesis*

and complexity of control systems, Novosibirsk, October 18-23, 2004, pages 8-12 (in Russian).
4. A. Botev. On algebraic immunity of new constructions of filters with high nonlinearity. In Proceedings of *VI international conference on Discrete models in the theory of control systems*, Moscow, December 7-11, 2004, pages 227-230 (in Russian).
5. A. Botev and Y. Tarannikov. Lower bounds on algebraic immunity for recursive constructions of nonlinear filters. Preprint 2004.
6. C. Carlet. A larger class of cryptographic Boolean functions via a study of the Maiorana-McFarland construction. In *CRYPTO 2002*, number 2442 in Lecture Notes in Computer Science, pages 549–564. Springer Verlag, 2002.
7. C. Carlet. Improving the algebraic immunity of resilient and nonlinear functions and constructing bent functions. IACR ePrint server, http://eprint.iacr.org, 2004/276.
8. C. Carlet. Concatenating indicators of flats for designing cryptographic functions. To appear in *Design, Codes and Cryptography.*
9. C. Carlet. Personal communications, 2005.
10. J. H. Cheon and D. H. Lee. Resistance of S-boxes against Algebraic Attacks. In *FSE 2004*, number 3017 in Lecture Notes in Computer Science, pages 83–94. Springer Verlag, 2004.
11. J. Y. Cho and J. Pieprzyk. Algebraic Attacks on SOBER-t32 and SOBER-128. In *FSE 2004*, number 3017 in Lecture Notes in Computer Science, pages 49–64. Springer Verlag, 2004.
12. N. Courtois and J. Pieprzyk. Cryptanalysis of block ciphers with overdefined systems of equations. In *Advances in Cryptology - ASIACRYPT 2002*, number 2501 in Lecture Notes in Computer Science, pages 267–287. Springer Verlag, 2002.
13. N. Courtois and W. Meier. Algebraic attacks on stream ciphers with linear feedback. In *Advances in Cryptology - EUROCRYPT 2003*, number 2656 in Lecture Notes in Computer Science, pages 345–359. Springer Verlag, 2003.
14. N. Courtois. Fast algebraic attacks on stream ciphers with linear feedback. In *Advances in Cryptology - CRYPTO 2003*, number 2729 in Lecture Notes in Computer Science, pages 176–194. Springer Verlag, 2003.
15. D. K. Dalai, K. C. Gupta and S. Maitra. Results on Algebraic Immunity for Cryptographically Significant Boolean Functions. In *INDOCRYPT 2004*, pages 92–106, number 3348, Lecture Notes in Computer Science, Springer-Verlag.
16. C. Ding, G. Xiao, and W. Shan. *The Stability Theory of Stream Ciphers*. Number 561 in Lecture Notes in Computer Science. Springer-Verlag, 1991.
17. D. H. Lee, J. Kim, J. Hong, J. W. Han and D. Moon. Algebraic Attacks on Summation Generators. In *FSE 2004*, number 3017 in Lecture Notes in Computer Science, pages 34–48. Springer Verlag, 2004.
18. W. Meier, E. Pasalic and C. Carlet. Algebraic attacks and decomposition of Boolean functions. In *Advances in Cryptology - EUROCRYPT 2004*, number 3027 in Lecture Notes in Computer Science, pages 474–491. Springer Verlag, 2004.
19. E. Pasalic, S. Maitra, T. Johansson and P. Sarkar. New constructions of resilient and correlation immune Boolean functions achieving upper bounds on nonlinearity. In *Workshop on Coding and Cryptography - WCC 2001*, Paris, January 8–12, 2001. Electronic Notes in Discrete Mathematics, Volume 6, Elsevier Science, 2001.
20. P. Sarkar and S. Maitra. Construction of nonlinear Boolean functions with important cryptographic properties. In *EUROCRYPT 2000*, number 1807 in Lecture Notes in Computer Science, pages 485–506. Springer Verlag, May 2000.
21. Y. V. Tarannikov. On resilient Boolean functions with maximum possible nonlinearity. In *Progress in Cryptology - INDOCRYPT 2000*, number 1977 in Lecture Notes in Computer Science, pages 19–30. Springer Verlag, 2000.