Constructions of Even-variable Boolean Function with Optimum Algebraic Immunity

Yindong Chen and Peizhong Lu

Fudan University, Shanghai 200433, China
{chenyd, pzlu}@fudan.edu.cn

Abstract. This paper proposed an improved construction of even-variable Boolean function with optimum algebraic degree. Compared with those in [1], our Boolean functions are more balance. Specially, for $k=2t+1(t>1)$, the $2k$-variables Boolean function is balanced. Furthermore, we generalized it to a class of constructions, meaning there would be much more constructions.
Keywords: stream cipher, algebraic attacks, Boolean function, algebraic immunity

1 Introduction

Recently, algebraic attack has gained a lot of attention in cryptanalysis [1–8]. The main idea of algebraic attack is to deduce the security of a stream cipher to solve an over-defined system of multivariate nonlinear equations. To implement algebraic attack, attackers firstly construct equation system between the input bits (the secret key bits) and the output bits, then recover the input bits by solving the equation system with efficient methods such as Linearization, Relinearization, XL, Grönber bases, etc. [9–11].

Algebraic attack was firstly applied to LFSR (Linear Feedback Shift Register)-based stream cipher by Courtois and Meier [1] in 2003. By searching low degree annihilator, some LFSR-based stream ciphers such as Toyocrypt, LILI-128 [1], SFINKS [5], etc. were successfully attacked. The efficiency of algebraic attack is guaranteed by the existing of low degree multiple for any Boolean function [2]. That is, for any n-variable Boolean function, there exists multiple function with degree no more than $\lceil \frac{n}{2} \rceil$. The core of algebraic attack is to find out minimum degree nonzero annihilators of f or of $f+1$. This minimum degree is related to the complexity of algebraic attacks [2].

To resist algebraic attack, a new cryptographic property of Boolean functions which is known as *algebraic immunity* (AI) has been proposed by Meier *et al.* [2]. The AI of a Boolean function expresses its ability to resist standard algebraic attack. Thus the AI of Boolean function used in cryptosystem should be sufficiently high. Courtois and Meier [1, 2] showed that, for any n-variable Boolean function, its AI is bounded by $\lceil \frac{n}{2} \rceil$. If the bound is achieved, we say the Boolean function have optimum AI. Obviously, a Boolean function with optimum AI has strongest ability to resist standard algebraic attack. Therefore, the construction of Boolean functions with optimum AI is of great importance.

Dalai *et al.* [12, 13] presented Boolean functions with optimum AI in even variables by an recursive construction. It's a second order recursive construction. Further study [12] showed that the functions are not balanced (although it is possible to build balanced ones from them, but there would result in extra computation). Another class of constructions [14–16] contains symmetric functions. Being symmetric, they present a risk if attacks using this peculiarity can be found in the future. Moreover, they do not have high nonlinearities either [17]. Li [18–20] proposed a method to construct all $(2k+1)$-variable Boolean functions with optimum AI from one such given function. The construction has theoretical sense. But the computational complexity of the construction do not have been well studied. Carlet and Feng [21] proposed a well construction based on the Boolean functions' trace representation, recently. Their Boolean functions have not only optimum AI but also high nonlinearity. Furthermore, they also have a good behavior against fast algebraic attacks, at least for small values of the number of variables. The drawback of the construction is the high complexity of the computation for the value of $f(x)$.

In this paper, we proposed an improved construction of even-variable Boolean function with optimum algebraic degree. We'll show that the Boolean functions constructed in this paper compare favourably with those in [1], in respects of AI, algebraic degree, and nonlinearity. Specially, our Boolean functions are more balance. And furthermore, for $k{=}2t{+}1(t{>}1)$, the $2k$-variables Boolean function is balanced. In the end, we'll also generalize it to a class of constructions, thus there would be much more constructions of Boolean functions with optimum AI.

The organization of the paper is as follows. In the following section we give some preliminaries about Boolean functions. In Section 3, we recall the construction in [1], and then in Section 4, we present the improved construction of Boolean functions with optimum AI. Their cryptographic properties are studied in Section 5. We also generalize it to a class of constructions in Section 6. Section 7 concludes the paper.

2 Preliminaries

Let $\mathbb{F}_2 = \{0, 1\}$, be the finite field with two elements. Then a *Boolean function* in n variables is defined as mapping from $\mathbb{F}_2^n$ into $\mathbb{F}_2$. We denote by B_n the set of all n-variable Boolean functions. A basic representation of a Boolean function $f(x_1, \cdots, x_n)$ is by the output column of its *truth table*, i.e., a binary string of length 2^n,

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

Sometimes, we may use a binary string of length 2^n to represent a n-variable Boolean function.

For an n-variables Boolean function f, we define its *support* and *offset* as

$$\mathrm{supp}(f) = \{x \in \mathbb{F}_2^n | f(x) = 1\},$$
$$\mathrm{offset}(f) = \{x \in \mathbb{F}_2^n | f(x) = 0\}.$$

and denote them by 1_f and 0_f respectively. The *Hamming weight* $\mathrm{wt}(f)$ of f is the size of $\mathrm{supp}(f)$, i.e., $\mathrm{wt}(f) = |\mathrm{supp}(f)|$. It counts the number of 1's in the truth table of f. We say f is *balanced*, if the truth table contains an equal number of 1's and 0's, i.e., $\mathrm{supp}(f) = \mathrm{offset}(f)$, implying $\mathrm{wt}(f) = 2^{n-1}$. The Hamming distance between two Boolean functions, f and g, is denoted by $\mathrm{d}(f, g)$ and is the number of places where their truth tables differ. Note that $\mathrm{d}(f, g) = \mathrm{wt}(f + g)$ (by abuse of notation, we also use $+$ to denote the addition in $\mathbb{F}_2$, i.e., the XOR);

Any Boolean function has a unique representation as a multivariate polynomial over $\mathbb{F}_2$, called the *algebraic normal form* (ANF):

$$f(x_1, \cdots, x_n) = a_0 + \sum_{1 \le i \le n} a_i x_i + \sum_{1 \le i \le j \le n} a_{ij} x_i x_j + \cdots + a_{12\cdots n} x_1 x_2 \cdots x_n,$$

where the coefficients $a_0, a_i, a_{ij}, \cdots, a_{12\cdots n} \in \mathbb{F}_2$. The *algebraic degree* $\deg(f)$ of f is the number of variables in the highest order term with nonzero coefficient. A Boolean function is *affine* if it has algebraic degree at most 1 and we denote by A_n the set of all affine functions in n variables.

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

$$\mathrm{nl}(f) = \min_{g \in A_n} (\mathrm{d}(f, g)).$$

To be cryptographically secure [22, 23], Boolean functions used in cryptographic systems must be balanced to prevent the system from leaking statistical information on the plaintext when the ciphertext is known, have high algebraic degree to counter linear synthesis by Berlekamp-Massey algorithm, have high order of correlation immunity to counter correlation attacks, and have high nonlinearity to withstand linear attacks and correlation attacks.

Recently, it has been identified that any combining or filtering should not have a low-degree-multiple. More precisely, it is shown in [1] 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 $f \cdot g$ has degree at most $\lceil \frac{n}{2} \rceil$. Therefore, while choosing a Boolean function f, the cryptosystem designer should avoid that the degree of $f \cdot g$ falls much below $\lceil \frac{n}{2} \rceil$ with a nonzero Boolean function g whose degree is also much below $\lceil \frac{n}{2} \rceil$. Otherwise, resulting low degree multivariate relations between key bits and output bits of Boolean function f will allow a very efficient attack. As observed in [1, 2], it is necessary to check that f and $f + 1$ do not admit nonzero annihilators of low degrees.

Definition 1. *Given $f \in B_n$, we define*

$$\mathrm{Ann}(f) = \{g \in B_n | f \cdot g = 0\}.$$

Any function $g \in \mathrm{Ann}(f)$ is called an annihilator of f.

It's explicit that a function g is an annihilator of f if and only if g takes value 0 on $\mathrm{supp}(f)$, i.e.,

$$g \in \mathrm{Ann}(f) \Leftrightarrow 1_f \subseteq 0_g.$$

Definition 2. *Given $f \in B_n$, we define its algebraic immunity, denote by* $\mathrm{AI}_n(f)$, *as the minimum degree of all nonzero annihilators of f or $f+1$, i.e.,*

$$\mathrm{AI}_n(f) = \min\{\deg(g) | 0 \neq g \in \mathrm{Ann}(f) \cup \mathrm{Ann}(f+1)\}.$$

We usually denote $\mathrm{AI}_n(f)$ by $\mathrm{AI}(f)$ for short, when there is no confusion about the number of variables.

Note that $\mathrm{AI}(f) \leq \deg(f)$, since $f \cdot (f+1) = 0$. As f or $f+1$ must have an annihilator at an algebraic degree $\leq \lceil \frac{n}{2} \rceil$ [1], we have $\mathrm{AI}(f) \leq \lceil \frac{n}{2} \rceil$. If an n-variable Boolean function f satisfies that $\deg(f) = \lceil \frac{n}{2} \rceil$, we say it has optimum AI. The AI of a Boolean function expresses its ability to resist standard algebraic attack. So, Boolean functions with higher AI (even optimum AI) is preferred in cryptosystem. Note that although AI is not a property that can resist all kinds of algebraic attacks, but clearly still a necessary one.

3 The Construction of Boolean Function in [1]

From now on, we use a binary string of length 2^n to represent an n-variable Boolean function. As used in [1], we denote by "$\|$" the concatenation of binary strings. We also denote by "$^-$" the complement operation, i.e., $\bar{f}$ meaning the complement function of Boolean function f.

For example, let $s, t \in B_2$, and $s = x_1x_2 + x_2 + 1, t = x_1x_2 + x_2$. In the truth table representation, they are $s = 1101, t = 0010$. Let $u = s\|t$, then $u = 11010010 \in B_3$, i.e., $u = x_1x_2 + x_2 + x_3 + 1$. And $\bar{u} = 00101101$.

For the denotation "$\|$", the following proposition holds.

Proposition 1. *Given $f_1, f_2 \in B_n$, let $f = f_1\|f_2$, then*

i) $f = f_1 + x_{n+1}(f_1 + f_2) \in B_{n+1}$, *and* $\deg(f_1), \deg(f_2) \leq \deg(f)$;
ii) for any $g \in \mathrm{Ann}(f)$, decompose it as $g = g_1\|g_2$ where $g_1, g_2 \in B_n$, then $g_1 \in \mathrm{Ann}(f_1)$ and $g_2 \in \mathrm{Ann}(f_2)$.

The following proposition can be deduced from Proposition 1.

Proposition 2. *Given $f_1, f_2, f_3, f_4 \in B_n$, let $f = f_1\|f_2\|f_3\|f_4$, then*

i) $f \in B_{n+2}$, and $f = f_1 + x_{n+1}(f_1 + f_2) + x_{n+2}(f_1 + f_3) + x_{n+1}x_{n+2}(f_1 + f_2 + f_3 + f_4)$;
ii) for any $g \in \mathrm{Ann}(f)$, decompose it as $g = g_1\|g_2\|g_3\|g_4$ where $g_1, g_2, g_3, g_4 \in B_n$, then $g_1 \in \mathrm{Ann}(f_1), g_2 \in \mathrm{Ann}(f_2), g_3 \in \mathrm{Ann}(f_3)$ and $g_4 \in \mathrm{Ann}(f_4)$.

For the denotation "$^-$", the following proposition holds.

Proposition 3. *Given $f \in B_{n+2}$, decompose it as $f = f_1\|f_2\|f_3\|f_4$, where $f_1, f_2, f_3, f_4 \in B_n$, then*

i) $\bar{f} = f + 1$;
ii) $\bar{\bar{f}} = f$;

iii) $\bar{f} = \bar{f}_1 \| \bar{f}_2 \| \bar{f}_3 \| \bar{f}_4$;
iv) $\mathrm{AI}(\bar{f}) = \mathrm{AI}(f)$.

Dalai et.al. firstly proposed a recursive construction of Boolean functions:

Construction 1.

$$\begin{cases} \phi_{2k+2} = \phi_{2k} \| \phi_{2k} \| \phi_{2k} \| \phi_{2k}^1, \\ \phi_{2j}^i = \phi_{2j-2}^{i-1} \| \phi_{2j-2}^i \| \phi_{2j-2}^i \| \phi_{2j-2}^{i+1}, \end{cases} \tag{1}$$

with base step $\phi_k^0 = \phi_k, \phi_0^i = x_1 + (j \mod 2)$, $i, n \geq 1, j > 0$.

They proved that the constructed Boolean function ϕ_{2k} has optimum AI. Based on the construction, we are to propose another construction for Boolean functions with optimum AI.

4 An Improved Construction of Boolean Function

Our construction is:

Construction 2.

$$\begin{cases} \phi_{2k+2} = \bar{\phi}_{2k} \| \bar{\phi}_{2k} \| \phi_{2k} \| \phi_{2k}^1, \\ \phi_{2j}^i = \bar{\phi}_{2j-2}^{i-1} \| \bar{\phi}_{2j-2}^i \| \phi_{2j-2}^i \| \phi_{2j-2}^{i+1}, \end{cases} \tag{2}$$

with base step $\phi_k^0 = \phi_k, \phi_0^i = x_1 + (j \mod 2)$, $i, n \geq 1, j > 0$.

From now on, we denote by ϕ the Boolean function defined by Construction 2. To prove that ϕ_{2k} has optimum AI, we need intermediate results. For technical reasons, during our proofs, we will encounter certain situations when the degree of a function is negative. As such functions do not exist, we will replace them by function 0.

Lemma 1. *Assume that the function $\phi_{2t} \in B_{2t}$ has been generated by Construction 2 and $\mathrm{AI}(\phi_t) = t$ for $1 \leq t \leq k$. If, for some $i > 0$, there exists $g, h \in B_{2t}$ such that*

i) $g \in \mathrm{Ann}(\phi_{2t}^i), h \in \mathrm{Ann}(\phi_{2t}^{i+1})$ *and* $\deg(g+h) \leq t-2-i$, *or*
ii) $g \in \mathrm{Ann}(\bar{\phi}_{2t}^i), h \in \mathrm{Ann}(\bar{\phi}_{2t}^{i+1})$ *and* $\deg(g+h) \leq t-2-i$,

then $g = h$.

Proof. We prove it by induction on t.

For the base step $t = 0$, $\deg(g+h) \leq 0-2-i \leq -2$ implies that functions in the assumption cannot exist, i.e., $g = h = 0$.

Now we prove the inductive step. Assume that, for $t < k$, the induction assumption holds (for every $i \geq 0$). We show it for $t = k$ and for every $i \geq 0$.

We rewrite g and h as

$$\begin{cases} g = g_1 \| g_2 \| g_3 \| g_4, \\ h = h_1 \| h_2 \| h_3 \| h_4, \end{cases}$$

where $g_1, g_2, g_3, g_4, h_1, h_2, h_3, h_4 \in B_{2k-2}$.
By Proposition 2, there is

$$\begin{aligned} g + h =& (g_1 + h_1) + x_{2k-1}(g_1 + h_1 + g_2 + h_2) + x_{2k}(g_1 + h_1 + g_3 + h_3) \\ &+ x_{2k-1}x_{2k}(g_1 + h_1 + g_2 + h_2 + g_3 + h_3 + g_4 + h_4). \end{aligned}$$

1) If $g \in \mathrm{Ann}(\phi_{2t}^i), h \in \mathrm{Ann}(\phi_{2t}^{i+1})$, such that $\deg(g + h) \le t - 2 - i$.
By the recursion (2), there is

$$\begin{cases} \phi_{2k}^{i+1} = \bar{\phi}_{2(k-1)}^i \| \bar{\phi}_{2(k-1)}^{i+1} \| \phi_{2(k-1)}^{i+1} \| \phi_{2(k-1)}^{i+2} & , \\ \phi_{2k}^i = \bar{\phi}_{2(k-1)}^{i-1} \| \bar{\phi}_{2(k-1)}^i \| \phi_{2(k-1)}^i \| \phi_{2(k-1)}^{i+1} & i > 0, \\ \phi_{2k} = \bar{\phi}_{2(k-1)} \| \bar{\phi}_{2(k-1)} \| \phi_{2(k-1)} \| \phi_{2(k-1)}^1 & i = 0. \end{cases}$$

   a) $\deg(g_1 + h_1) \le k - 2 - i = (k-1) - 2 - (i-1)$.
   If $i > 0$, then $g_1 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)}^{i-1}), h_1 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)}^i)$ implies that $g_1 = h_1$, according to the induction assumption.
   If $i = 0$, then $g_1, h_1 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)})$, and therefore $g_1 + h_1 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)})$. By hypothesis and Proposition 3, $\mathrm{AI}_{2(k-1)}(\bar{\phi}_{2(k-1)}) = \mathrm{AI}_{2(k-1)}(\phi_{2(k-1)})) = k - 1$.
   Since $\deg(g_1 + h_1) \le k - 2$, we have $g_1 + h_1 = 0$, i.e., $g_1 = h_1$.
   b) $\deg(g_2 + h_2) \le k - 2 - i - 1 = (k-1) - 2 - i$, and $g_2 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)}^i), h_2 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)}^{i+1})$. Then by the induction assumption, we have $g_2 = h_2$.
   c) $\deg(g_3 + h_3) \le k - 2 - i - 1 = (k-1) - 2 - i$, and $g_3 \in \mathrm{Ann}(\phi_{2(k-1)}^i), h_3 \in \mathrm{Ann}(\phi_{2(k-1)}^{i+1})$. Then by the induction assumption, we have $g_3 = h_3$.
   d) $\deg(g_4 + h_4) \le k - 2 - i - 2 = (k-1) - 2 - (i+1)$, and $g_4 \in \mathrm{Ann}(\phi_{2(k-1)}^{i+1}), h_4 \in \mathrm{Ann}(\phi_{2(k-1)}^{i+2})$. Then by the induction assumption, we have $g_4 = h_4$.

2) If $g \in \mathrm{Ann}(\bar{\phi}_{2k}^i), h \in \mathrm{Ann}(\bar{\phi}_{2k}^{i+1})$, such that $\deg(g + h) \le k - 2 - i$, then by the recursion (2), there is

$$\begin{cases} \bar{\phi}_{2k}^{i+1} = \phi_{2(k-1)}^i \| \phi_{2(k-1)}^{i+1} \| \bar{\phi}_{2(k-1)}^{i+1} \| \bar{\phi}_{2(k-1)}^{i+2} & , \\ \bar{\phi}_{2k}^i = \phi_{2(k-1)}^{i-1} \| \phi_{2(k-1)}^i \| \bar{\phi}_{2(k-1)}^i \| \bar{\phi}_{2(k-1)}^{i+1} & i > 0, \\ \bar{\phi}_{2k} = \phi_{2(k-1)} \| \phi_{2(k-1)} \| \bar{\phi}_{2(k-1)} \| \bar{\phi}_{2(k-1)}^1 & i = 0. \end{cases}$$

   Similar to 1), we can prove that $g_1 = h_1, g_2 = h_2, g_3 = h_3$ and $g_4 = h_4$.

Hence we get $g + h = 0$, i.e., $g = h$ which finishes the proof. □

Lemma 2. *Assume that the function $\phi_{2t} \in B_{2t}$ has been generated by Construction 2 and $\mathrm{AI}(\phi_t) = t$ for $1 \le t \le k$. If there exists $g, h \in B_{2t}$ such that*

i) $g \in \mathrm{Ann}(\bar{\phi}_{2t}^{i}), h \in \mathrm{Ann}(\phi_{2t}^{i+1})$ *and* $\deg(g+h) \leq t-1-i$, $i>0$, *or*
ii) $g \in \mathrm{Ann}(\bar{\phi}_{2t}), h \in \mathrm{Ann}(\phi_{2t})$ *and* $\deg(g+h) \leq t$, $i=0$,

then $g=h=0$.

Proof. We prove it by induction on t.

For the base step $t=0$, $\deg(g+h) \leq t-2-i = 0-2-i \leq -2$, implies that functions in the assumption cannot exist, i.e., $g+h=0, g=h$.

Now we prove the inductive step. Assume that, for $t<k$, the induction assumption holds. We show it for $t=k$.

We rewrite g and h as

$$\begin{cases} g = g_1 \| g_2 \| g_3 \| g_4, \\ h = h_1 \| h_2 \| h_3 \| h_4, \end{cases}$$

where $g_1, g_2, g_3, g_4, h_1, h_2, h_3, h_4 \in B_{2k-2}$.
By Proposition 2, there is

$$\begin{aligned} g+h =& (g_1+h_1) + x_{2k-1}(g_1+h_1+g_2+h_2) + x_{2k}(g_1+h_1+g_3+h_3) \\ &+ x_{2k-1}x_{2k}(g_1+h_1+g_2+h_2+g_3+h_3+g_4+h_4). \end{aligned}$$

1) If $g \in \mathrm{Ann}(\bar{\phi}_{2k}^{i}), h \in \mathrm{Ann}(\phi_{2k}^{i+1})$ and $\deg(g+h) \leq k-1-i$, $i>0$then by the recursion 2, there is

$$\begin{cases} \bar{\phi}_{2k}^{i} = \phi_{2(k-1)}^{i-1} \| \phi_{2(k-1)}^{i} \| \bar{\phi}_{2(k-1)}^{i} \| \bar{\phi}_{2(k-1)}^{i+1}, \\ \phi_{2k}^{i} = \bar{\phi}_{2(k-1)}^{i-1} \| \bar{\phi}_{2(k-1)}^{i} \| \phi_{2(k-1)}^{i} \| \phi_{2(k-1)}^{i+1}, \end{cases}$$

By Proposition 2, we have

a) $\deg(g_1+h_1) \leq k-1-i = (k-1)-1-(i-1)$, and $g_1 \in \mathrm{Ann}(\phi_{2(k-1)}^{i-1}), h_1 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)}^{i-1})$.
If $i-1>0$, then $g_1=h_1$, according to the induction assumption i).
If $i-1=0$, then $g_1=h_1$, according to the induction assumption ii).

b) $\deg(g_2+h_2) \leq k-1-i-1 = (k-1)-1-i$, and $g_2 \in \mathrm{Ann}(\phi_{2(k-1)}^{i}), h_2 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)}^{i+1})$. Then by the induction assumption i), we have $g_2=h_2$.

c) $\deg(g_3+h_3) \leq k-1-i-1 = (k-1)-1-i$, and $g_3 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)}^{i}), h_3 \in \mathrm{Ann}(\phi_{2(k-1)}^{i+1})$. Then by the induction assumption i), we have $g_3=h_3$.

d) $\deg(g_4+h_4) \leq k-1-i-2 = (k-1)-1-(i+1)$, and $g_4 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)}^{i+1}), h_4 \in \mathrm{Ann}(\phi_{2(k-1)}^{i+2})$. Then by the induction assumption i), we have $g_4=h_4$.

Thus, $g_1=h_1, g_2=h_2, g_3=h_3, g_4=h_4$, i.e., $g=h$.
And then $g=h=0$, since $g=h \in \mathrm{Ann}(\bar{\phi}_{2k}^{i}) \cap \mathrm{Ann}(\phi_{2k}^{i+1})$.

2) If $g \in \mathrm{Ann}(\bar{\phi}_{2k}), h \in \mathrm{Ann}(\phi_{2k})$, such that $\deg(g+h) \leq k, i=0$, then by the recursion 2, there is

$$\begin{cases} \bar{\phi}_{2k} = \phi_{2(k-1)} \| \phi_{2(k-1)} \| \bar{\phi}_{2(k-1)} \| \bar{\phi}_{2(k-1)}^{1}, \\ \phi_{2k} = \bar{\phi}_{2(k-1)} \| \bar{\phi}_{2(k-1)} \| \phi_{2(k-1)} \| \phi_{2(k-1)}^{1}, \end{cases}$$

By Proposition 2, we have

a) $g_1 \in \mathrm{Ann}(\phi_{2(k-1)}), h_1 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)})$, and $\deg(g_1), \deg h_1 \leq k-1$ implying $\deg(g_1+h_1) \leq k-1$. Then $g_1 = h_1$, according to the induction assumption ii).
b) $g_2 \in \mathrm{Ann}(\phi_{2(k-1)}), h_2 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)})$, and $\deg(g_2+h_2) \leq k-1$. Then $g_2 = h_2$, according to the induction assumption ii).
c) $g_3 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)}), h_3 \in \mathrm{Ann}(\phi_{2(k-1)})$, and $\deg(g_3+h_3) \leq k-1$. Then $g_3 = h_3$, according to the induction assumption ii).
d) $g_4 \in \mathrm{Ann}(\bar{\phi}_{2(k-1)}), h_4 \in \mathrm{Ann}(\phi_{2(k-1)})$, and $\deg(g_4+h_4) \leq k-1$. Then $g_4 = h_4$, according to the induction assumption ii).

Thus, $g_1 = h_1, g_2 = h_2, g_3 = h_3, g_4 = h_4$, i.e., $g = h$.
And then $g = h = 0$, since $g = h \in \mathrm{Ann}(\bar{\phi}_{2k}) \cap \mathrm{Ann}(\phi_{2k})$.

Here we finish the proof. □

Theorem 1. *The function ϕ_{2k} obtained in Construction 2 has optimum algebraic immunity, for every $k \geq 1$, i.e.,*

$$\mathrm{AI}(\phi_{2k}) = k.$$

Proof. We prove it by induction on k.

For the base step $k = 1$, it can easily be checked.

Now we prove the inductive step. Assume that, for $k < t$, the induction assumption holds. We show it for $k = t$.

We have to prove that any nonzero function g such that $g \cdot \phi_{2t} = 0$ has degree at least t (proving that any nonzero function g such that $g \cdot (\phi_{2t} + 1) = 0$ has degree at least t is similar). Suppose that such a function $g \in B_{2t}$ with $\deg(g) < t$ exists. Then, g can be rewritten as $g = g_1 \| g_2 \| g_3 \| g_4$ where $g_1, g_2, g_3, g_4 \in B_{2(t-1)}$. By Proposition 2, we have

$$\begin{aligned} g(x_1, x_2, \cdots, x_{2t-1}, x_{2t}) =& g_1 + x_{2t-1}(g_1 + g_2) + x_{2t}(g_1 + g_3) \\ &+ x_{2t-1}x_{2t}(g_1 + g_2 + g_3 + g_4). \end{aligned} \tag{3}$$

Since $\phi_{2tk} = \bar{\phi}_{2(t-1)} \| \bar{\phi}_{2(t-1)} \| \phi_{2(t-1)} \| \phi^1_{2(t-1)}$ and $\deg(g) < t$, by Proposition 2, we have

$$\begin{aligned} &\deg(g_1 + g_2), \deg(g_1 + g_3) \leq t-2, \deg(g_1 + g_2 + g_3 + g_4) \leq t-3, \\ &g_1, g_2 \in \mathrm{Ann}(\bar{\phi}_{2(t-1)}), g_3 \in \mathrm{Ann}(\phi_{2(t-1)}), g_4 \in \mathrm{Ann}(\phi^1_{2(t-1)}). \end{aligned} \tag{4}$$

Then $g_1 + g_2 \in \mathrm{Ann}(\bar{\phi}_{2(t-1)})$. By the induction assumption, there is $\mathrm{AI}(\bar{\phi}_{2(t-1)}) = \mathrm{AI}(\phi_{2(t-1)}) = t-1$.
Thus, $g_1 + g_2 = 0$, since $\deg(g_1 + g_2) \leq t-2$.
So, (3) becomes

$$g = g_1 + x_{2t}(g_1 + g_3) + x_{2t-1}x_{2t}(g_3 + g_4), \tag{5}$$

and then

$$\deg(g_1 + g_3) \leq t-2 < t-1, \quad \deg(g_3 + g_4) \leq t-3 = (t-1)-2.$$

By Lemma 1, we have $g_3 = g_4$.
By Lemma 2, we have $g_1 = g_3 = 0$.
Thus, $g_4 = 0$, too.
Hence $g = 0$, which completes the proof. □

5 The analysis of other cryptographic properties

In this section, we will analyze other cryptographic properties of the constructed Boolean functions. In the analysis, we lay emphasis on their balance, algebraic degree and nonlinearity.

5.1 Balance

For an n-variable Boolean function f, we can use the value

$$\mathrm{b}(f) = |2^{n-1} - \mathrm{wt}(f)|$$

to measure its balance. Obviously, the less the value $\mathrm{b}(f)$ is , the more balance the function f is. If $\mathrm{b}(f) = 0$, then we say the Boolean function f is *balanced*.

Let $w_{2k}^i = \mathrm{wt}(\phi_{2k}^i), w_{2k} = \mathrm{wt}(\phi_{2k}) = w_{2k}^0 = \mathrm{wt}(\phi_{2k}^0)$. By the recursion in Construction 2, there is

$$\begin{cases} w_{2k}^i = 2^{2k-1} - w_{2k-2}^{i-1} + w_{2k-2}^{i+1}, \\ w_{2k} = 2^{2k-1} - w_{2k-2} + w_{2k-2}^1, \end{cases} \tag{6}$$

where $k, i \geq 1$, and $w_0^j = j \mod 2 (j \geq 0)$.

Denote $b_{2k}^i = 2^{2k-1} - w_{2k}^i$, then $|b_{2k}^i|$ measures the balance of ϕ_{2k}^i. By (6), we have

$$\begin{cases} b_{2k}^i = b_{2k-2}^{i+1} - b_{2k-2}^{i-1}, \\ b_{2k}^0 = b_{2k-2}^1 - b_{2k-2}^0, \end{cases} \tag{7}$$

where $k, i \geq 1$, and $b_0^j = j \mod 2 (j \geq 0)$.

By induction on k, the following property can easily be proved.

Property 1. Let $b_{2k}^i = 2^{2k-1} - \mathrm{wt}(\phi_{2k}^i), (k \geq 2, i \geq 0)$, then

i) if k is even, then $b_{2k}^{2i} = b_{2k}^{2i+1}$;
ii) if k is odd, then $b_{2k}^0 = 0$, $b_{2k}^{2i+1} = b_{2k}^{2i+2}$.

Property 1 shows that if $k > 1$ is odd number, then the Boolean function ϕ_{2k} obtained from Construction 2 is balanced. Note that none of the Boolean functions from Construction 1 is balanced. Although it is possible to build balanced ones from them, but there would result in extra computation. The balance of Boolean functions between the two construction are compared in Table 1 . It's clearly that the balance of Boolean functions from Construction 2 is better than that from Construction 1.

Table 1: The balance of boolean function in both constructions

	$2k$=2	$2k$=4	$2k$=6	$2k$=8	$2k$=10	$2k$=12	$2k$=14	$2k$=16	$2k$=18	$2k$=20	$2k$=22	$2k$=24
$\mathrm{b}(\phi'_{2k})$	1	3	10	35	126	462	1716	6435	24310	92378	352716	1352078
$\mathrm{b}(\phi_{2k})$	1	1	0	1	0	2	0	5	0	14	0	42

Note 1. ϕ'_{2k} is the Boolean function in Construction 1, and ϕ_{2k} in Construction 2.

5.2 Nonlinearity

Property 2. For the function $\phi_{2k}(k > 0)$ obtained in Construction 2, there is

$$\mathrm{nl}(\phi_{2k}) \geq 2^{2k-1} - \binom{2k-1}{k-1}. \tag{8}$$

Proof. Consider the $(2k+1)$-variable Boolean function

$$\varphi_{2k+1} = x_{2k+1} + \phi_{2k}.$$

By Theorem 1, we have $\mathrm{AI}(\phi_{2k}) = k$.

By Lemma 2, there does not exists $g, h \in B_{2k}$, such that

$$0 \neq g \in \mathrm{Ann}(\phi_{2k}), 0 \neq h \in \mathrm{Ann}(\bar{\phi}_{2k}), \text{and } \deg(g), \deg(h), \deg(g+h) < k.$$

Thus $\mathrm{AI}(\varphi_{2k+1}) = k+1$, according to [26].

And then $\mathrm{nl}(\varphi 2k+1) = nl(x_{2k+1} + \phi_{2k}) \geq 2^{2k} - \binom{2k}{k}$.

Note that for a k-variable Boolean function f, there has $\mathrm{nl}(x_{2k+1} + f) = 2\,\mathrm{nl}(f)$. Hence, $\mathrm{nl}(\phi_{2k}) \geq 2^{2k-1} - \binom{2k-1}{k-1}$, which finishes the proof. □

The nonlinearity of Boolean function ϕ_{2k} in Construction 1 is $\mathrm{nl}(\phi_{2k}) = 2^{2k-1} - \binom{2k-1}{k-1}$. Therefor, the Boolean functions in Construction 2 is not worse then those in Construction 1, in the aspect of nonlinearity.

5.3 Algebraic Degree

To avoid confusion, we denote by $\phi'^{i}_{2k}, \phi^{i}_{2k}$ the Boolean functions obtained in Construction 1 and Construction 2, respectively. Let $\gamma^i_k = \phi^i_{2k} + \phi^{i+1}_{2k}, \delta^i_k = \phi^i_{2k} + \phi^{i+2}_{2k}, \gamma'^{i}_{k} = \phi'^{i}_{2k} + \phi'^{i+1}_{2k}, \delta'^{i}_{k} = \phi'^{i}_{2k} + \phi'^{i+2}_{2k}$. By induction on k, it can be easily proved that: For any $k \geq i \geq 0$, there has $\deg(\phi^i_{2k}) = \deg(\phi'^{i}_{2k}), \deg(\gamma^i_k) = \deg(\gamma'^{i}_{k}), \deg(\delta^i_k) = \deg(\delta'^{i}_{k})$. Sequentially, the following property holds.

Property 3. $\phi'^{i}_{2k}, \phi^{i}_{2k}$ are the Boolean functions obtained in Construction 1 and Construction 2 respectively, then $\deg(\phi^i_{2k}) = \deg(\phi'^{i}_{2k})$, i.e., [1]

$$\begin{cases} \deg(\phi_{2k}) = 2k & k = 2^t, \\ 2k-3 \leq \deg(\phi_{2k}) \leq 2k-1 & k = 2^t - 1, \\ \deg(\phi_{2k}) = 2k-1 & \text{others.} \end{cases}$$

Property 3 shows that the Boolean functions obtained in Construction 1 and Construction 2 are identical, in the aspect of algebraic degree.

6 Generalization

6.1 More constructions

Comparing Construction 2 with Construction 1, the former one just change the fist two concatenated functions (of the latter one in the recursive formula) into their complement functions. We can express the transformation in a vector $(a_1, a_2, a_3, a_4) \in \mathbb{F}_2^4$ as follows: If $a_i = 1 (1 \leq i \leq 4)$ then change the i-th concatenated function into its complement function.

In this way, the Construction 2 can be expressed as the vector $(1, 1, 0, 0)$, and Construction 1, $(0, 0, 0, 0)$.

Actually, we found more constructions by exhaust all the vectors in $\mathbb{F}_2^4$. We list the constructions that can produce Boolean functions with optimum AI in Table 2. All the constructions are with the same base step as Construction 1 and Construction 2.

Table 2: More constructions of Boolean functions with optimum AI

Cons.1	Cons.2	Cons.3	Cons.4	Cons.5	Cons.6	Cons.7	Cons.8
(0,0,0,0)	(1,1,0,0)	(1,0,1,0)	(1,0,0,1)	(0,1,1,0)	(0,1,0,1)	(0,0,1,1,)	(1,1,1,1)

Observing that the AI, algebraic degree, and nonlinearity of Boolean functions in Construction 2 are identical with those in Construction 1, we doubt that whether Boolean functions in Construction 2 are affinely equivalent to those in Construction 1. We firstly check that whether it's a affine transformation that changing any one of the four concatenated functions into its complement function. Since any of the transformation is combined by some of the four base one: $(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)$. For example, $(1, 1, 0, 0)$ is the combined by $(1, 0, 0, 0)$ and $(0, 1, 0, 0)$. If the four base transformation are affine ones, than all the transformation in $\mathbb{F}_2^4$ are affine, too. So we focus our attention on the four base transformation. Actually, we compute the AI's of Boolean functions in the four base transformation, and find that they are not equal to those of Construction 1. That means the four base transformation are not affine ones, since AI is a invariable in affine transformation.

Table 3: Constructions of Boolean functions have no optimum AI

Cons.9	Cons.10	Cons.11	Cons.12	Cons.13	Cons.14	Cons.15	Cons.16
(1,0,0,0)	(0,1,0,0)	(0,0,1,0)	(0,0,0,1)	(0,1,1,1)	(1,0,1,1)	(1,1,0,1,)	(1,1,1,0)

6.2 Further discussion

In further research, we find that the Boolean functions still has optimum AI by replacing the "$^-$" operation by affine transformation. That is

Theorem 2. *$\tau_2, \tau_4, \cdots, \tau_{2k} (k > 1)$ are affine transformation, and*

$$\begin{cases} \phi_{2k+2} = \tau_{2k}(\phi_{2k}) \| \tau_{2k}(\phi_{2k}) \| \phi_{2k} \| \phi_{2k}^1, \\ \phi_{2j}^i = \tau_{2j}(\phi_{2j-2}^{i-1}) \| \tau_{2j}(\phi_{2j-2}^i) \| \phi_{2j-2}^i \| \phi_{2j-2}^{i+1}, \end{cases} \tag{9}$$

with base step $\phi_k^0 = \phi_k, \phi_0^i = x_1 + (j \mod 2)$, $i, n \geq 1, j > 0$. Then the Boolean functions ϕ_{2k} in the upper construction have optimum AI.

Consider the reversed operation "$\leftarrow$", which means reversing the Boolean function's truth table (e.g., $\overleftarrow{1101} = 1011$). We find that, after replacing the "$^-$" operation by "$\leftarrow$", ϕ_{2k} in Construction 2 still has optimum AI. In fact, the reversed operation "$\leftarrow$" can be represented as $\overleftarrow{f}(x_1, x_2, \cdots, x_n) = f(x_1 + 1, x_2 + 1, \cdots, x_n + 1)$, obviously an affine transformation. It's consistent with Theorem 2. As another application of Theorem 2, we can assert that the following construction produces Boolean functions with optimum AI.

$$\begin{cases} \phi_{2k}^i = \bar{\phi}_{2k-2}^{i-1} \| \bar{\phi}_{2k-2}^i \| \phi_{2k-2}^i \| \phi_{2k-2}^{i+1} & k \bmod 2 = 1, \\ \phi_{2k}^i = \overleftarrow{\phi}_{2k-2}^{i-1} \| \overleftarrow{\phi}_{2k-2}^i \| \phi_{2k-2}^i \| \phi_{2k-2}^{i+1}, & k \bmod 2 = 0. \end{cases} \tag{10}$$

What need to point out is that, although for any affine transformation in Theorem 2 the Boolean functions' AI maintain, but other cryptographic properties may change. How to select suitable transformation in Theorem 2 to make sure the Boolean functions have good cryptographic properties is an open problem.

7 Conclusion

Based on [1], this paper proposed an improved recursive construction of even-variable Boolean function with optimum algebraic immunity. We showed that the Boolean functions constructed in this paper compare favourably with those in [1], in respects of AI, algebraic degree, and nonlinearity. In the case of balance, our Boolean functions superior to those in [1]. Furthermore, when $k > 3$ is odd, our construction produce balanced Boolean functions. In the end, we also generalized it to a class of constructions, thus there would be much more constructions of Boolean functions with optimum AI.

References

1. N. Courtois and W. Meier. Algebraic Attacks on Stream Ciphers with Linear Feedback[A]. Advances in Cryptology-Eurocrypt 2003[C], Berlin: Springer-Verlag, 2003, 345-359

2. W. Meier, E. Pasalic, and C. Carlet. Algebraic attacks and decomposition of Boolean functions[A]. Advances in Cryptology-Eurocrypt 2004[C], Berlin: Springer-Verlag, 2004, 474-491
3. F. Armknecht and M. Krause. Algebraic Attacks on Combiners with Memory[A]. Advances in Cryptology-Crypto 2003[C], Berlin: Springer-Verlag, 2003, 162-175
4. N. Courtois. Algebraic Attacks on Combiners with Memory and Several Outputs[A]. Information security and cryptology 2004 (ICISC 2004), LNCS 3506, 2005, 3-20
5. N. Courtois. Cryptanalysis of SFINKS[A]. Information Security and Cryptology 2005 (ICISC 2005)[C]. Berlin: Springer-Verlag, 2006, 261-269
6. L. M. Batten. Algebraic Attacks over GF(q) [A]. Progress in Cryptology-Indocrypt 2004[C], Berlin: Springer-Verlag, 2004, 84-91
7. J. Faugère and A. Joux. Algebraic Cryptanalysis of Hidden Field Equation (HFE) Cryptosystems Using Gröbner Bases[A]. Advances in Cryptology-Crypto 2003[C], Berlin: Springer-Verlag, 2003, 44-60
8. F. Armknecht. On the Existence of low-degree Equations for Algebraic Attacks[EB/OL]. http://eprint.iacr.org/2004/185
9. N. Courtois, A. Klimov, J. Patarin, *et al.* Efficient algorithms for solving overdefined systems of multivariate polynomial equations[A]. Advances in Cryptology-Eurocrypt 2000[C], Berlin: Springer-Verlag, 2000, 392-407
10. A. Kipnis and A. Shamir. Cryptanalysis of the HFE public key cryptosystem by relinearization[A]. Advances in Cryptology-Crypto'99, Berlin: Springer-Verlag, 1999, 19-30
11. William W Adams, Philippe Loustaunau. An introduction to gröbner bases[M]. USA: AMS, 1994
12. C. Carlet, D. K. Dalai, K. C. Gupta, *et al.* Algebraic Immunity for Cryptographically Significant Boolean Functions: Analysis and Construction[J]. IEEE Transactions on Information Theory, 2006, 52(7): 3105-3121
13. D. K. Dalai, K. C. Gupta and S. Maitra. Cryptographically Significant Boolean functions: Construction and Analysis in terms of Algebraic Immunity[A]. Fast Software Encryption 2005 (FSE05) [C], Paris, France, 2005, 98-111
14. A. Braeken and B. Preneel. On the algebraic immunity of symmetric Boolean functions[A]. Progress in Cryptology-Indocrypt 2005[C], Berlin: Springer-Verlag, 2005, 35-48
15. D. K. Dalai, S. Maitra and S. Sarkar. Basic Theory in Construction of Boolean Functions with Maximum Possible Annihilator Immunity[J]. Design, Codes and Cryptography, 2006, 40(1): 41-58
16. C. Carlet. A method of construction of balanced functions with optimum algebraic immunity[EB/OL]. http://eprint.iacr.org/2006/149
17. C. Carlet, X. Zeng, C. Li, *et al.* Further properties of several classes of Boolean functions with optimum algebraic immunity[EB/OL]. http://eprint.iacr.org/2007/370
18. N. Li and W. Qi. Construction and analysis of Boolean functions of $2t+1$ variables with maximum algebraic immunity[A]. Advances in Cryptology-Asiacrypt 2006[C], Berlin: Springer-Verlag, 2006, 84-98
19. N. Li and W. Qi. Boolean function of an odd number of variables with maximum algebraic immunity[J]. Science in China, Ser. F, 2007, 50(3): 307-317
20. N. Li, L. Qu, W. Qi, *et al.* On the construction of Boolean functions with optimal algebraic immunity[J]. IEEE Transactions on Information Theory, 2008, 54(3): 1330-1334
21. C. Carlet and K. Feng. An Infinite Class of Balanced Functions with Optimal Algebraic Immunity, Good Immunity to Fast Algebraic Attacks and Good Nonlin-

earity[A]. Advances in Cryptology-Asiacrypt 2008[C], Berlin: Springer-Verlag, 2008, 425-440
22. A. Canteaut and M. Trabbia. Improved fast correlation attacks using parity-check equations of weight 4 and 5[A]. Advances in Cryptology-Eurocrypt 2000[C]. Berlin: Springer-Verlag, 2000, 573-588
23. C. Ding, G. Xiao, and W. Shan. The Stability Theory of Stream Ciphers[M]. Lecture Notes in Computer Science (vol.561). Berlin: Springer-Verlag, 1991
24. M. Lobanov. Tight bound between nonlinearity and algebraic immunity[EB/OL]. http://eprint.iacr.org/2005/441