Characterizations on Algebraic Immunity for Multi-Output Boolean Functions

Xiao Zhong[1,2] and Mingsheng Wang[3]

1. Institute of Software, Chinese Academy of Sciences, Beijing 100190, China
2. Graduate School of Chinese Academy of Sciences, Beijing 100190, China
3. State Key Laboratory of Information Security, Institute of Information Engineering,
Chinese Academy of Sciences, Beijing, China
zhongxiao456@163.com
mingsheng_wang@yahoo.com.cn

Abstract. The general principle for algebraic attack for multi-output stream ciphers was proposed by Courtois [6]. Furthermore, Armknecht, and Krause gave a definition of algebraic immunity for multi-output Boolean functions in [2], and investigated some construction methods of multi-output Boolean functions with maximal algebraic immunity. In this note, several new characterizations of algebraic immunity for multi-output Boolean functions are given, and some related invariants and their relations are also investigated. Some examples are given to illustrate the usefulness of these results.

Key words: Algebraic attacks, Stream Ciphers, Multi-output boolean functions, Single-output boolean functions, Algebraic immunity, Augmented functions.

1 Introduction

Algebraic attacks are kinds of efficient methods on stream ciphers which adapt LFSR-based keystream generators with nonlinear filter generator or the combiner as one of their components. The basic idea is to recover the initial state of the LFSR by solving a system of nonlinear equations of low degree. Courtois and Meier made a major breakthrough in algebraic attacks on stream ciphers [4,5]. They proposed a powerful method to obtain low degree equations in the initial state bits by multiplying the filter function f by a well chosen boolean function g. Hence given some keystream bits, the attack tries to recover the initial state by solving systems of the overdefined polynomial equations. Further, the concept of algebraic immunity(AI) for boolean functions was introduced in [17] as a measure for stream ciphers based on linear feedbacks against algebraic attacks. In general it should be as high as possible and the maximum is $\lceil \frac{m}{2} \rceil$ where m is the number of variables of boolean functions.

Since then, a large body of literature has emerged on the field of algebraic immunity (AI) [10,11,12,13]. Carlet. C and Feng. K proposed an infinite class of

balanced boolean functions with optimum algebraic immunity [9]. This class of functions also achieves optimum algebraic degree and a much better nonlinearity than all the previously obtained infinite classes of functions.

On the other hand, with the development of the stream ciphers, some modern stream ciphers are multi-output, the encryption and decryption speed of which is much faster than the single-output ones. There are two basic designs for such stream ciphers based on the linear feedback shift register (LFSR) [16]. One kind is the filter function generator where m bits are extracted from one LFSR as inputs to the filter function $f(x)$ to produce n-bit ($n \geq 1$) keystream. The other one is the combiner generator consisting of m LFSRs and a multi-output boolean function $f(x)$. One bit is trapped from each LFSR as an input bit for $f(x)$ to produce n-bit keystream each clock.

Similar with the single-output stream ciphers, it is necessary to consider the algebraic attack on the multi-output case. Courtois firstly proposed a general algebraic attack on multi-output stream ciphers [6], which is of forward-looking significance. Armknecht and Krause presented a definition of the algebraic immunity for multi-output boolean functions in [2], and investigated some construction methods of multi-output boolean functions with maximal algebraic immunity. However, for the concept of algebraic immunity for multi-output Boolean functions, currently we still lack of good understanding. In this note, We study the general properties and various possible characterizations of the algebraic immunity of multi-output boolean functions, and get some interesting results.

The paper is organized as follows: In Section 2, we present some preliminaries which will be used throughout this paper, based on the work of Courtois and Meier [4,5,6,7,8], we propose several new invariants for describing low degree relations for a given multi-output function, and give related research problems. In Section 3, we give solutions to research problems proposed in Section 2, these results can be viewed as new characterizations of algebraic immunity of the multi-output boolean functions. In Section 4, we give some examples to show usefulness of results. Section 5 concludes the paper.

2 Preliminaries and Motivations to Research Problems

A boolean function on m variables is a mapping from $\mathbb{F}_2^m$ into $\mathbb{F}_2$. We denote the ring of boolean functions in m variables by $\mathbb{B}_m$. Let $X_1, \ldots, X_m$ be m indeterminates, then we may represent $\mathbb{B}_m$ as $\mathbb{B}_m = \mathbb{F}_2[X_1, \ldots, X_m]/I$, where $I = \mathrm{Ideal}(X_1^2 - X_1, \ldots, X_m^2 - X_m)$ is the ideal generated by $X_1^2 - X_1, \ldots, X_m^2 - X_m$ in $\mathbb{F}_2[X_1, \ldots, X_m]$.

Hence every $h \in \mathbb{B}_m$ can be represented as $h = h_1 \bmod I$, where $h_1 \in \mathbb{F}_2[X_1, \ldots, X_m]$. We denote $X_i \bmod I$ by x_i, thus an element in $\mathbb{B}_m$ is denoted by $h(x_1, \ldots, x_m)$, while its corresponding element in $\mathbb{F}_2[X_1, \ldots, X_m]$ is denoted by $h(X_1, \ldots, X_m)$.

Thus every boolean function f of m variables may be written as

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

Where $a_I \in \mathbb{F}_2$. The terms $\Pi_{i\in I} x_i$ are called monomials. The algebraic degree $\deg(f)$ of a boolean function f equals to the maximal degree of those monomials with nonzero coefficients.

Let $f \in \mathbb{B}_m$, we would like to introduce the following notations that will be used throughout the paper.

(i) For $g = (g_1, \ldots, g_n) \in \mathbb{B}_m^n$, $f = (f_1, \ldots, f_n) \in \mathbb{B}_m^n$, $\langle g, f\rangle$ denotes $\sum_{i=1}^n g_i f_i$, which is in $\mathbb{B}_m$. In particular, for $c = (c_1, \ldots, c_n) \in \mathbb{F}_2^n$, $\langle c, g\rangle = \sum_{i=1}^n c_i g_i$.

(ii) Let $b = (b_1, \ldots, b_n) \in \mathbb{F}_2^n$, and $f = (f_1, \ldots, f_n) \in \mathbb{B}_m^n$, $\mathrm{Ideal}(f + b)$ denotes the ideal generated by $f_1 + b_1, \ldots, f_n + b_n$ in $\mathbb{B}_m$, and $\mathrm{Ann}(f + b) = \{g \in \mathbb{B}_m^n | g \circ (f + b) = \sum_{i=1}^n (f_i + b_i) g_i = 0\}$.

(iii) For a nonzero $g = (g_1, \ldots, g_n) \in \mathbb{B}_m^n$, define $\deg_1(g) = \max_{0\neq c\in\mathbb{F}_2^n}\{\deg(\langle c, g\rangle)|\langle c, g\rangle \neq 0\}$, $\deg_2(g) = \min_{0\neq c\in\mathbb{F}_2^n}\{\deg(, g\rangle)|\langle c, g\rangle \neq 0\}$. It is easy to know $\deg_1(g) = \max_i\{\deg(g_i)|g_i \neq 0\}$.

2.1 Algebraic Attack on LFSR-Based Single-Output Stream Ciphers

Courtois and Meier [4,5] presented algebraic attack on stream ciphers with LFSR, and they focused on the single-output cases. The combination of the work in [4,5,7] indicates the following important algebraic attack scenarios on single-output stream ciphers:

For a boolean function $f \in \mathbb{B}_m$, the following cases may arise:

S1: f has a low algebraic degree D (classical criterion).

S2: f can be approximated by a low-degree non-linear multivariate function with probability $1 - \epsilon$ for some small ϵ.

S3: f has some multiple fg of low degree d, with g being some non-zero Boolean polynomial.

S4: f has some multiple fg, such that f can be approximated by a function of low degree with probability $1 - \epsilon$ for some small ϵ.

S5: There exists a non-trivial multivariate relation of low degree that relates(only) the key bits and several output bits of the cipher.

S6: There exists a non-trivial multivariate relation of low degree true with probability $1 - \epsilon$ for some small ϵ that relates (only) the key bits and several output bits of the cipher.

Particularly, there is a more general attack **S5** promoted by Courtois, Meier, Armknecht and Krause [8].

The above scenarios can be used to mount algebraic attack on LFSR-based stream ciphers, such as LiLi-128, Toyocrypt, and E_0 [4,1]. The algebraic immunity of boolean functions quantifies the resistance to the algebraic attack of stream ciphers based on LFSR filtered by a boolean function. Let us recall the following important definition:

Definition 1. *[17] Let $f \in \mathbb{B}_m$, $AI(f) = \min\{deg(g)|g \neq 0, gf = 0$ or $g(f+1) = 0\}$ is called algebraic immunity of f.*

By Courtois and Meier's theorem [4], $AI(f) \leq \lceil \frac{m}{2} \rceil$. In general $AI(f)$ should be as large as possible in order to resist algebraic attack.

2.2 Research Problems

Let S be a subset of $\mathbb{F}_2^m$, define $I(S) = \{g \in \mathbb{B}_m | g(s) = 0, \forall s \in S\}$.

Let $f : \mathbb{F}_2^m \to \mathbb{F}_2^n$, Armknecht and Krause [2] introduced the following definition:

Definition 2. *Let $f : \mathbb{F}_2^m \to \mathbb{F}_2^n$. $AI(f) = \min\{deg(g)|0 \neq g \in I(f^{-1}(a)), a \in \mathbb{F}_2^n$ with $f^{-1}(a) \neq \emptyset\}$ is called algebraic immunity of f.*

Let d be the least integer such that $\sum_{i=0}^{d} \binom{m}{d} > 2^{m-n}$, then $AI(f) \leq d$ [2]. Feng. K et al. proved that d can be reached by suitable boolean functions [13].

Similar with the conventional algebraic attack in [4], we consider only synchronous stream ciphers. The target cipher systems are multi-output stream ciphers.

Let the length of the linear feedback shift register be m. L is the "connection function" of the LFSR, and it is linear. The LFSR generator polynomial is a primitive polynomial $p(x) = p_0 + p_1 x + ... + p_{m-1}x^{m-1} + x^m$. It generates an m-sequence which is filtered by a nonlinear multi-output boolean function $f : \mathbb{F}_2^m \to \mathbb{F}_2^n$. Let the initial state of the LFSR is $s^0 = (s_0, s_1, ..., s_{m-1})$, then the state of the LFSR at time t is

$$s^t = (s_t, s_{t+1}, ..., s_{t+m-1}) = L^t(s_0, s_1, ..., s_{m-1}).$$

Denote the output of the filter generator by $C_0, C_1, C_2, ...$, where $C_i \in \mathbb{F}_2^n$, then we can get the following equation system:

$$\begin{cases} C_0 = f \quad (s_0, s_1, ..., s_{m-1}) \\ C_1 = f(L\,(s_0, s_1, ..., s_{m-1})) \\ C_2 = f(L^2(s_0, s_1, ..., s_{m-1})) \\ \vdots \end{cases}$$

The problem is to recover the initial state $s^0 = (s_0, s_1, ..., s_{m-1})$ from some keystream bits $C_0, C_1, C_2, ...$, where $C_i \in \mathbb{F}_2^n$.

Courtois proposed a general algebraic attack on such stream ciphers [6]. It works as follows:

Find (by some method that is very different for each cipher) one (at least, but one is enough) multivariate relation Q of low degree between the LFSR state bits and intermediate output bits, for example:

$$Q(s_0, s_1, ..., s_{m-1}, C_0, C_1, C_2, ..., C_{n-1}) = 0.$$

This principle is very significance for studying algebraic attacks of multi-output stream ciphers.

Inspired by the work of Courtois and Meier [4,5,6,7,8] on the algebraic attack on stream ciphers, for $f : \mathbb{F}_2^m \to \mathbb{F}_2^n$, considering the following situations:

T1: There exists low degree $g = (g_1, \ldots, g_n) \in \mathrm{Ann}(f)$, that is, $\langle f, g \rangle = \sum_{i=1}^{n} f_i g_i = 0$.

T2: There exists $g = (g_1, \ldots, g_n) \in \mathbb{B}_m^n$ such that $\langle f, g\rangle = \sum_{i=1}^n f_i g_i = h$ is of low degree.

T3: There exists low degree $g = (g_1, \ldots, g_n) \in \mathbb{B}_m^n$ such that $g \in \mathrm{Ann}(f + b)$ for some nonzero $b \in F_2^n$.

T4: There exists low degree $h \in \mathrm{Ideal}(f + b)$ for some nonzero $b \in F_2^n$.

There are three cases that one can mount algebraic attack by using $T1$, $T2$, $T3$ and $T4$.

Case 1: $T1$ **and** $T2$

Using $T1$ and $T2$, one can obtain low degree equations:

(a) For T1, if $f(v) = u \neq 0$, then for a low degree $g = (g_1, \ldots, g_n) \in \mathrm{Ann}(f)$, one obtain a low degree equation: $0 = \langle g(v), f(v)\rangle = \langle g(v), u\rangle = \sum_{i=1}^n u_i g_i(v)$.

(b) For T2, if $f(v) = 0$, then for a low degree $h = \langle g, f\rangle \in \mathrm{Ideal}(f)$, we have a low degree equation: $h(v) = 0$.

These low degree equations can be used to mount algebraic attack to stream ciphers. Thus the following definition should be significant.

Definition 3. *Let* $f : \mathbb{F}_2^m \to \mathbb{F}_2^n$*, define* $AI_j(f) = \min\{\deg_j(g), \deg(h) | 0 \neq g \in \mathrm{Ann}(f), 0 \neq h \in \mathrm{Ideal}(f)\}$ *for* $j = 1, 2$*.*

From Definition 3, one can see that when $n = 1$, $AI_1(f) = AI_2(f) = AI(f)$, which is corresponding to the single-output case. For arbitrary n, $AI_j(f)$ should be large in certain sense.

However, when $n > 1$, only considering $AI_j(f)$ is not enough. For example, if $v \in \mathbb{F}_2^m$, $b \in \mathbb{F}_2^n$ such that $f(v) = b \neq 0$, and there is no low degree function in $Ann(f)$, but there exists a low degree function in $Ideal(f + b)$; this is possible, for $Ann(f)$ does not generally equal to $Ideal(f + b)$ when $n > 1$. Hence we also consider the following case:

Case 2: $T3$ **and** $T4$

Using $T3$ and $T4$, one can get the following attack:

(c) For T3, if $f(v) = a \neq b$, then for a low degree $g \in \mathrm{Ann}(f + b)$, we get a low degree equation $0 = \langle g(v), (f(v) + b)\rangle = \sum_{i=1}^n (a_i + b_i) g_i(v)$.

(d) For T4, if $f(v) = b$, then for a low degree $h \in \mathrm{Ideal}(f + b)$, we derive a low degree equation $h(v) = 0$.

Based on **(a)**, **(b)**, **(c)** and **(d)**, we give the following definition:

Definition 4. *Let* $f : \mathbb{F}_2^m \to \mathbb{F}_2^n$ *be a map.* $AI_j(f+b) = \min\{\deg_j(g), \deg(h) | 0 \neq g \in \mathrm{Ann}(f + b), 0 \neq h \in \mathrm{Ideal}(f + b)\}$ *for* $j = 1, 2$*.*

It is necessary that $\min_{b \in \mathbb{F}_2^n} AI_j(f + b)$ should be large in certain sense. Thus the following problem seem to be worthy of consideration:

Problem 2.1. What is the relationship of $AI(f)$ and $\min_{b \in \mathbb{F}_2^n} AI_j(f + b)$?

We shall solve Problem 2.1 in the next section.

Similar with the case $n = 1$, low degree relations among f_1+b_1, f_2+b_2,...,f_n+b_n may also produce the low degree equation. One can use $T1$ and $T3$ to mount an attack similar with **(a)** and **(b)** mentioned above.

Case 3: $T1$ **and** $T3$

Using $T1$ and $T3$, we can mount the following attack:

(e) For $T1$, if $f(v) \neq 0$, then we have similar low degree equation as in **(a)** using the low degree boolean function from $\mathrm{Ann}(f)$.

(f) For $T3$, if $f(v) = 0$, then for a low degree g as in $T3$, we obtain $0 = g(v) \circ (f(v) + b) = \sum_{i=1}^{n} b_i g_i(v)$.

Corresponding to **(e)** and **(f)**, the following definition should be considered:

Definition 5. *Let* $f : \mathbb{F}_2^m \to \mathbb{F}_2^n$, *define* $ANI_j(f) = \min\{\deg_j(g) | 0 \neq g \in \cup_{b \in \mathbb{F}_2^n} \mathrm{Ann}(f + b)\}$ *for* $j = 1, 2$.

By definition of $ANI_j(f)$, for any $b \in \mathbb{F}_2^n$, we have $ANI_j(f) = ANI_j(f + b)$. Thus a natural problem is as follows:

Problem 2.2. How to relate $AI(f)$ with $ANI_j(f)$ for $j = 1, 2$?

We shall solve Problem 2.2 in the next section.

When we apply algebraic attack on the LFSR-based multi-output stream ciphers, we look for the boolean functions that satisfy Case 1, Case 2 or Case 3 by computing Gröbner basis, resulting an equation system of the lowest degree with the initial state bits as the variables. Solve the equation system and we get the initial state value. In the following, we will give an example on the LFSR-based multi-output stream cipher.

Example 1. Let the generator polynomial be $p(x) = x^7 + x + 1$, which is primitive. The filter function is a 7-variable and 2-output function $F = (f_1, f_2) : \mathbb{F}_2^7 \to \mathbb{F}_2^2$, where $f_1 : \mathbb{F}_2^7 \to \mathbb{F}_2$, $f_2 : \mathbb{F}_2^7 \to \mathbb{F}_2$.

$f_1 = x_1x_2x_3x_4x_5x_6 + x_1x_2x_3x_4x_5x_7 + x_1x_2x_3x_4x_5 + x_1x_2x_3x_4x_6x_7 + x_1x_2x_3x_4x_7 + x_1x_2x_3x_5x_6x_7 + x_1x_2x_3x_5x_6 + x_1x_2x_3x_6x_7 + x_1x_2x_3x_6 + x_1x_2x_4x_6x_7 + x_1x_2x_4x_6 + x_1x_2x_4x_7 + x_1x_2x_4 + x_1x_2x_6x_7 + x_1x_2x_7 + x_1x_3x_4x_6 + x_1x_3x_5x_6x_7 + x_1x_3x_5x_7 + x_1x_3x_5 + x_1x_3x_6x_7 + x_1x_3x_7 + x_1x_4x_5x_6x_7 + x_1x_4x_5x_7 + x_1x_4x_6x_7 + x_1x_4x_6 + x_1x_4x_7 + x_1x_4 + x_1x_5x_6 + x_1x_5x_7 + x_1x_7 + x_2x_3x_4x_5x_6x_7 + x_2x_3x_4x_5x_6 + x_2x_3x_4x_6x_7 + x_2x_3x_4x_7 + x_2x_3x_5x_7 + x_2x_3x_5 + x_2x_4x_5x_7 + x_2x_4x_6 + x_2x_5x_7 + x_2x_5 + x_2x_6x_7 + x_3x_4x_5x_6x_7 + x_3x_4x_6 + x_3x_5x_7 + x_3x_6 + x_4x_5x_7 + x_4x_7 + 1.$

$f_2 = x_1x_2x_3x_4x_5 + x_1x_2x_3x_4x_7 + x_1x_2x_3x_4 + x_1x_2x_3x_5x_6x_7 + x_1x_2x_3x_5x_7 + x_1x_2x_3x_5 + x_1x_2x_3 + x_1x_2x_4x_5x_6x_7 + x_1x_2x_4x_5 + x_1x_2x_4x_6x_7 + x_1x_2x_4x_7 + x_1x_2x_5x_6x_7 + x_1x_2x_5x_7 + x_1x_2x_5 + x_1x_2x_6x_7 + x_1x_3x_4x_5x_6x_7 + x_1x_3x_4x_6 + x_1x_3x_4x_7 + x_1x_3x_5x_6 + x_1x_3x_5x_7 + x_1x_3x_5 + x_1x_3x_6 + x_1x_3x_7 + x_1x_3 + x_1x_4x_5x_6x_7 + x_1x_4x_6x_7 + x_1x_4x_7 + x_1x_4 + x_1x_5x_7 + x_1x_5 + x_1x_6x_7 + x_1x_6 + x_1x_7 + x_1 + x_2x_3x_4x_5x_7 + x_2x_3x_4x_5 + x_2x_3x_4x_6x_7 + x_2x_3x_4x_6 + x_2x_3x_4x_7 + x_2x_3x_4 + x_2x_3x_5x_6 + x_2x_3x_5x_7 + x_2x_3x_5 + x_2x_3x_6x_7 + x_2x_3x_6 + x_2x_3 + x_2x_4x_5x_6x_7 + x_2x_4x_5x_6 + x_2x_4 + x_2x_5x_6 + x_2x_5x_7 + x_2x_6x_7 + x_2x_6 + x_2x_7 + x_2 + x_3x_4x_5x_6 + x_3x_4x_5x_7 + x_3x_4x_5 + x_3x_4x_6x_7 + x_3x_4x_6 + x_3x_4x_7 + x_3x_4 + x_3x_5x_6x_7 + x_3x_5 + x_3x_6x_7 + x_3x_7 + x_3 + x_4x_5x_6x_7 + x_4x_5x_6 + x_4x_5x_7 + x_4x_5 + x_4x_6 + x_4 + x_5x_6x_7 + x_5x_6 + x_5x_7 + x_5 + x_6x_7 + x_6 + x_7.$

Here f_1 and f_2 are both of optimum algebraic immunity, $AI(f_1) = AI(f_2) = 4$.

Using Case 1, Case 2 and Case 3, we find the following equation:

$f_1(x_1x_5x_7 + x_3x_5x_7 + x_1x_7 + x_3x_7 + x_5x_7 + x_7) + f_2(x_2x_3x_7 + x_1x_5x_7 + x_2x_5x_7 + x_3x_5x_7 + x_4x_5x_7 + x_2x_6x_7 + x_1x_7 + x_3x_7 + x_4x_7 + x_5x_7 + x_7) = 0.$

In the next section, we first prove some basic properties involving the invariants defined in this section.

3 Characterizations on the Algebraic Immunity of the Multi-Output Boolean Function

We begin to prove the following lemma which shows that $\min_{0\neq c\in F_2^n}(AI(\langle c,f\rangle))$ is the biggest among all these invariants:

Lemma 1. *Let $f:\mathbb{F}_2^m\to\mathbb{F}_2^n$ be a map, then $ANI_1(f)\leq \min_{0\neq c\in F_2^n}(AI(\langle c,f\rangle))$, $ANI_2(f)\leq \min_{0\neq c\in F_2^n}(AI(\langle c,f\rangle))$.*

Proof. Let $0\neq c=(c_1,\ldots,c_n)\in\mathbb{F}_2^n$ such that $AI(\langle c,f\rangle)=\min_{e\neq 0}(AI(\langle e,f\rangle))$. By definition of $AI(\langle c,f\rangle)$, there exists a boolean function g such that $AI(\langle c,f\rangle)=\deg(g)$, and $g(\langle c,f\rangle+b)=0$, where $b=0$ or 1. Let $c_{i_1}=\cdots=c_{i_s}=1$, other terms are 0. Then $\langle c,f\rangle=f_{i_1}+\cdots+f_{i_s}$. Let $\overline{g}$ be a vector of boolean functions such that i_j position is g for $j=1,\ldots,s$, other positions are 0. Hence $\langle\overline{g},f+\overline{b}\rangle=0$, where $\overline{b}$ is a vector of length n with b in the i_1 position and 0 otherwise. Thus $ANI_1(f)\leq \deg_1(\overline{g})=\deg(g)$. Similar discussion shows that $ANI_2(f)\leq\deg(g)$. □

Lemma 2. *Let $f:\mathbb{F}_2^m\to\mathbb{F}_2^n$ be a surjective map, then $AI(f)\leq ANI_2(f)$.*

Proof. Let $ANI_2(f)$ be achieved by $g=(g_1,\ldots,g_n)$. That is, $ANI_2(f)=\deg_2(g)=\deg(\langle b,g\rangle)$ for a nonzero $b\in\mathbb{F}_2^n$, and $\langle g,f+a\rangle=0$ for some $a\in\mathbb{F}_2^n$.

For $v=(v_1,\ldots,v_m)\in\mathbb{F}_2^m$ such that $f(v)=a+b$, we have $f(v)+a=b$. Hence $0=\langle g(v),f(v)+a\rangle=\langle b,g(v)\rangle=\langle b,g\rangle(v)$. Hence by definition of $AI(f)$, we have $AI(f)\leq\deg(\langle b,g\rangle)$. □

By Lemma 1 and Lemma 2, the following corollary can be directly obtained:

Corollary 1. *Let $f=(f_1,\ldots,f_n):\mathbb{F}_2^m\to\mathbb{F}_2^n$ be a surjective map, then $AI(f)\leq \min_{0\neq c\in\mathbb{F}_2^n}(AI(\langle c,f\rangle))$.*

Remark 1. Corollary 1 indicates that the algebraic immunity of a vector boolean function is less than or equal to that of its component functions. For a multi-output boolean function $f=(f_1,f_2,...,f_n)\in\mathbb{B}_m^n$, we choose $c=(1,0,0,...,0)\in\mathbb{F}_2^n$, then we can see $AI(f)\leq AI(f_1)$. Similarly, we can get that $AI(f)\leq AI(f_i)$, $i\in\{2,...,n\}$. The following example shows that the " = " can be reached in some cases.

Example 2. Let $f=(f_1,f_2):\mathbb{F}_2^6\to\mathbb{F}_2^2$ be a surjective map.
$f_1=x_1x_2x_3x_4x_5+x_1x_2x_3x_4x_6+x_1x_2x_3x_4+x_1x_2x_4x_5x_6+x_1x_2x_4x_5+x_1x_2x_4x_6+x_1x_2x_5x_6+x_1x_2x_5+x_1x_2x_6+x_1x_2+x_1x_3x_4+x_1x_3x_5+x_1x_3x_6+x_1x_4x_5+x_1x_4+x_1+x_2x_3x_4x_5x_6+x_2x_3x_4x_6+x_2x_3x_6+x_2x_4x_5+x_2x_4x_6+x_2x_4+x_2x_5x_6+x_2x_5+x_3x_4x_6+x_3x_4+x_3x_6+x_5+x_6$.

$f_2 = x_1x_2x_3x_4x_5+x_1x_2x_3x_5x_6+x_1x_2x_3x_5+x_1x_3x_4x_5x_6+x_1x_3x_4x_5+x_2x_3x_4x_5x_6+x_1x_2x_3x_6+x_1x_2x_3+x_1x_2x_4+x_1x_3x_4+x_2x_3x_4x_6+x_1x_3x_5+x_1x_4x_5+x_2x_4x_5+x_1x_3x_6+x_2x_3x_6+x_1x_4+x_2x_4x_6+x_2x_4+x_3x_4x_6+x_3x_4+x_2x_5x_6+x_2x_5+x_1+x_3x_6+x_5+x_6$.
We can compute that $AI(f) = 1$, $AI(f_1) = AI(f_2) = 3$, and $AI(f_1 + f_2) = 1$.

Example 2 verifies that $AI(f) \leq \min_{0\neq c\in\mathbb{F}_2^n}(AI(\langle c, f\rangle))$, and the " = " can be reached.

In order to investigate the problems in Section 2, we introduce a new invariant:

Definition 6. *Let $f = (f_1, \ldots, f_n) : \mathbb{F}_2^m \to \mathbb{F}_2^n$ be a map. Define* $\mathrm{DI}(f) = \min\{\deg(g)|0 \neq g \in \cup_{a\in Im(f)}\mathrm{Ideal}(f + a)\}$*, where $Im(f)$ is the image set of f.*

With the above notation, the following lemma provides an algebraic description of $AI(f)$, which seems to be obvious from the view of algebraic geometry and the definition of $AI(f)$. Since it is a basis of our subsequent discussions, we give here a rigorous proof.

Lemma 3. *Let $f = (f_1, \ldots, f_n) : \mathbb{F}_2^m \to \mathbb{F}_2^n$ be a map, then $AI(f) = \mathrm{DI}(f)$.*

Proof. Let $\overline{\mathbb{F}}_2$ be the algebraically closure of $\mathbb{F}_2$. Let $f_i(X)$ be the polynomial obtained from the boolean function $f_i(x)$ by replacing the boolean variables x_i by the polynomial variables X_i for $i = 1, \ldots, m$, where $x = (x_1, \ldots, x_m)$, and $X = (X_1, \ldots, X_m)$.

Let J be the ideal generated by $f_1(X)+a_1, \ldots, f_n(X)+a_n, X_1^2-X_1, \ldots, X_m^2-X_m$ in $\mathbb{F}_2[X_1, \ldots, X_m]$, and $J_1 = J\overline{\mathbb{F}}_2[X_1, \ldots, X_m]$ is the extension of J in $\overline{\mathbb{F}}_2[X_1, \ldots, X_m]$. For $a \in \mathbb{F}_2^n$, it is easy to know $f^{-1}(a) = \{v \in \overline{\mathbb{F}}_2^m|\forall h \in J_1, h(v) = 0\}$.

Let $0 \neq g \in \mathbb{B}_m$ such that g vanishes on $f^{-1}(a)$ with $\deg(g) = AI(f)$, where $f^{-1}(a) \neq \emptyset$. Let $g_1 \in \mathbb{F}_2[X_1, \ldots, X_m]$ be any polynomial corresponding to g, i.e., $g_1 \bmod I = g$, where $I = \mathrm{Ideal}(X_1^2 - X_1, \ldots, X_m^2 - X_m)$. Then g_1 vanishes on $f^{-1}(a)$. Hence by Hilbert's Nullstellensatz, $g_1^r \in J_1$ for some positive integer r. Therefore, $g_1^r \in J_1 \cap \mathbb{F}_2[X_1, \ldots, X_m] = J$ ([3], page 212).

Note that $g_1^r \bmod I = (g_1 \bmod I)^r = g^r = g = g_1 \bmod I$, we have $g_1^r - g_1 \in I$. By $I \subseteq J$, we have $g_1 \in J$. Hence there exists the following equation in $\mathbb{F}_2[X_1, \ldots, X_m]$:

$g_1 = \sum_{i=1}^n h_i(X_1, \ldots, X_m)(f_i(X_1, \ldots, X_m) + a_i) + h$, where $h \in I$.

Passing to $\mathbb{B}_m$ by modulo I, we get $g = \sum_{i=1}^n h_i(x_1, \ldots, x_m)(f_i(x_1, \ldots, x_m)+a_i)$, thus $g \in \mathrm{Ideal}(f + a)$ in $\mathbb{B}_m$. Hence $\deg(g) \geq \mathrm{DI}(f)$, i.e., $AI(f) \geq \mathrm{DI}(f)$.

On the other hand, let $0 \neq h \in \cup_{a\in Im(f)}\mathrm{Ideal}(f + a)$ such that $\deg(h) = \mathrm{DI}(f)$. Thus for some $a \in Im(f)$ and boolean functions g_i, $h = \sum_{i=1}^n g_i(f_i + a_i)$. Thus for any $v \in f^{-1}(a)$, we have $h(v) = 0$. Hence $\mathrm{DI}(f) = \deg(h) \geq AI(f)$. Thus we have proved $AI(f) = \mathrm{DI}(f)$. □

The following theorem gives the relationship of $AI(f)$ and $ANI_2(f)$.

Theorem 1. *Let $f = (f_1, \ldots, f_n) : \mathbb{F}_2^m \to \mathbb{F}_2^n$ be a map. Then $AI(f) \geq ANI_2(f)$. In particular, when f is surjective, $AI(f) = ANI_2(f)$.*

Proof. By Lemma 3, $AI(f) = \mathrm{DI}(f)$. Hence there exists a nonzero $g \in \mathrm{Ideal}(f + a)$ for some $a = (a_1, \ldots, a_n) \in \mathbb{F}_2^n$ such that $AI(f) = \deg(g)$.

Therefore there exists $h_i \in \mathbb{B}_m$ for $i = 1, \ldots, n$ such that

$$g = h_1(f_1 + a_1) + h_2(f_2 + a_2) + \cdots + h_n(f_n + a_n). \tag{1}$$

Since $g \neq 0$, there exists some i with $h_i \neq 0$. Without loss of generality, we may assume $i = 1$. Multiplying two sides of (1) by $f_1 + a_1$, we get:

$$g(f_1 + a_1) = h_1(f_1 + a_1) + h_2'(f + a_2) + \cdots + h_n'(f_n + a_n), \tag{2}$$

where $h_i' = h_i(f_i + a_i)$ for $i = 2, \ldots, n$.

Combining (1) and (2), we obtain:

$$g(f_1 + a_1 + 1) + w_2(f_2 + a_2) + \cdots + w_n(f_n + a_n) = 0, \tag{3}$$

where $w_i = h_i + h_i'$ for $i = 2, \ldots, n$.

From (3), we get $g' = (g, w_2, \ldots, w_n) \in \mathrm{Ann}(f + a')$, where $a' = (a_1 + 1, a_2, \ldots, a_n)$. Hence $ANI_2(f) \leq \deg_2(g') \leq \deg(g) = AI(f)$.

In particular, when f is surjective, by Lemma 2, $AI(f) \leq ANI_2(f)$. Thus we have $AI(f) = ANI_2(f)$. □

Until now, we have solved Problem 2.2.

We would like to give an example to verify that when f is surjective, $AI(f) = ANI_2(f)$.

Example 3. Let $f = (f_1, f_2) : \mathbb{F}_2^5 \to \mathbb{F}_2^2$ be a surjective map.
$f_1 = x_1x_2x_3x_5 + x_1x_2x_5 + x_1x_2 + x_1x_3x_4x_5 + x_1x_3x_4 + x_1x_3x_5 + x_1x_4x_5 + x_1x_4 + x_2x_3 + x_2x_4x_5 + x_2x_5 + x_3x_4 + x_4x_5 + 1$.
$f_2 = x_1x_2x_3x_4 + x_1x_2x_3 + x_1x_2x_4x_5 + x_1x_2x_4 + x_1x_2x_5 + x_1x_3x_5 + x_1x_3 + x_1x_5 + x_2x_3x_4x_5 + x_2x_3x_5 + x_2x_4x_5 + x_2x_5 + x_3x_4 + x_3 + x_4x_5 + 1$.
We can compute that $AI(f) = ANI_2(f) = 2$, which completes our verification.

In order to give a characterization of $AI(f)$ in view of the minimum of $AI_2(f + a)$ and $AI_1(f + a)$, we note the following lemma:

Lemma 4. *Let $f : \mathbb{F}_2^m \to \mathbb{F}_2^n$ be a surjective map. Then $AI(f) \leq AI_2(f)$.*

Proof. Let $g = (g_1, \ldots, g_n) \in \mathrm{Ann}(f)$ such that $\deg_2(g) = \min\{\deg_2(h) | 0 \neq h \in \mathrm{Ann}(f)\}$.

Assume that $\deg_2(g) = \deg(\langle b, g\rangle)$ for some nonzero $b \in \mathbb{F}_2^n$. Then $\forall v \in f^{-1}(b)$, we have $f(v) = b$. Hence $0 = \langle g(v), f(v)\rangle = (\langle b, g\rangle)(v)$, which shows that $AI(f) \leq \deg(\langle b, g\rangle)$.

On the other hand, let $0 \neq h \in \mathrm{Ideal}(f)$ such that $\deg(h)$ reaches the minimum. Thus $h = \sum_{i=1}^n g_i f_i$ for some boolean functions $g_1, \ldots, g_n$. $\forall v \in f^{-1}(0)$, we have $f(v) = (f_1(v), \ldots, f_n(v)) = 0$. Hence $h(v) = 0$, which shows that $AI(f) \leq \deg(h)$. Thus we have proved $AI(f) \leq AI_2(f)$. □

The following theorem reveals the relations among $AI(f)$ and $AI_j(f + b)$ for $b \in \mathbb{F}_2^n$ and $j = 1, 2$, which can be thought as a solution to Problem 2.1.

Theorem 2. *Let $f : \mathbb{F}_2^m \to \mathbb{F}_2^n$ be a surjective map. Then $AI(f) = \min_{b\in\mathbb{F}_2^n} AI_2(f+b) = \min_{b\in\mathbb{F}_2^n} AI_1(f+b)$.*

Proof. By Lemma 4, we have $AI(f) \leq AI_2(f)$. Hence for any $b \in \mathbb{F}_2^n$, we have $AI(f) = AI(f+b) \leq AI_2(f+b)$. Thus $AI(f) \leq \min_{b\in\mathbb{F}_2^n} AI_2(f+b) \leq \min_{b\in\mathbb{F}_2^n} AI_1(f+b)$.

On the other hand, $\min_{b\in\mathbb{F}_2^n} AI_1(f+b) \leq \min_{b\in\mathbb{F}_2^n}\{\min\deg(g)|0 \neq g \in \text{Ideal}(f+b)\}$. The latter is $AI(f)$ by Lemma 3. Hence $\min_{b\in\mathbb{F}_2^n} AI_1(f+b) \leq AI(f)$. Thus we have proved the conclusion. □

Inspired by Corollary 1, we conjecture that for the LFSR-based single-output stream ciphers, maybe we can use our method in Section 2 to derive equation system of degree lower than the optimum algebraic immunity of the filter function. In the next section, we apply our attack on single-output stream ciphers by using augmented function to transform the model into multi-output.

4 Examples on LFSR-Based Single-Output Stream Ciphers Using Augmented Function

For a LFSR-based single-output stream cipher, Courtois and Meier [5] indicated that it can be seen as using several functions defined as: $f(s), f(L(s)), f(L^2(s))...$, which resulted to the attack scenarios **S5** mentioned in Section 3. Qichun Wang and Thomas Johansson presented a method called higher order algebraic attack on stream ciphers [14]. They also apply their attack on augmented functions. However, they did not give the detailed and systematic attack method that can give a way on how to look for the low degree equations.

When the boolean function is single-output, we can construct augmented function to transform it into multi-output case.

First, let us describe the model of single-output stream ciphers. The LFSR is the same with the one described in Section 3. While the filter function $f : \mathbb{F}_2^m \to \mathbb{F}_2$ is single-output. The algebraic immunity of f is optimum, which means that the lowest degree of the equation system we can find equals to the algebraic immunity of f when we apply the conventional algebraic attack given in [4] on this model.

Let the initial state of the LFSR be $s^0 = (s_0, s_1, ..., s_{m-1})$, then the state of the LFSR at time t is

$$s^t = (s_t, s_{t+1}, ..., s_{t+m-1}) = L^t(s_0, s_1, ..., s_{m-1}).$$

Denote the output of the filter generator by $c_0, c_1, c_2, ...$, where $c_i \in \mathbb{F}_2$, then we can get the following equation system:

$$\begin{cases} c_0 = f \quad (s_0, s_1, ..., s_{m-1}) \\ c_1 = f(L\,(s_0, s_1, ..., s_{m-1})) \\ c_2 = f(L^2(s_0, s_1, ..., s_{m-1})) \\ \vdots \end{cases}$$

When we apply the conventional algebraic attack on this model, we can get equations of degree no less than $\lceil \frac{m}{2} \rceil$. While according to Corollary 1, we may get equations of degree less than $\lceil \frac{m}{2} \rceil$ if we use the augmented function. First, let us recall the definition of augmented function:

Definition 7. *For the nonlinear filter function $f : \mathbb{F}_2^m \to \mathbb{F}_2$, the n-th augmented function of f is defined as $F^n : \mathbb{F}_2^{n+m-1} \to \mathbb{F}_2^n$:*
$F^n(s_0, s_1, ..., s_{n+m-1}) = (f(s_0, s_1, ..., s_{m-1}), f(s_1, s_2, ..., s_m), ..., f(s_n, s_{n+1}, ..., s_{n+m-1}))$

The generator polynomial is $p(x) = p_0 + p_1 x + ... + p_{m-1}x^{m-1} + x^m$, which is primitive. Then we can get the generator matrix of the sequence generated by LSFR:

$$M = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ 0 & 0 & 0 & ... & 1 \\ p_0 & p_1 & p_2 & ... & p_{m-1} \end{pmatrix} \tag{4}$$

Then Definition 7 can also be written as:

Definition 8. *For the nonlinear filter function $f : \mathbb{F}_2^m \to \mathbb{F}_2$, the n-th augmented function of f is defined as $F^n : \mathbb{F}_2^m \to \mathbb{F}_2^n$:*

$$F^n(s_0, s_1, ..., s_{m-1}) = (f(s_0, s_1, ..., s_{m-1}), f(M(s_0, s_1, ..., s_{m-1})^T), ..., f(M^n(s_0, s_1, ..., s_{m-1})^T)).$$

According to Corollary 1, we derive that the algebraic immunity of the augmented function $AI(F^n) \leq AI(f)$. It means that although $AI(f)$ is optimum, we may find equation system of degree less than $AI(f)$ by targeting the augmented function F^n.

Here we would like to give examples to show that when the single-output filter function f has optimum algebraic immunity, attack on augmented function case can find equation system of degree less than $AI(f)$.

Example 4. Let the generator polynomial is $p(x) = x^5+x^2+1$. The filter function is a 5-variable Carlet-Feng function $f(x_1, x_2, ..., x_5) = x_1x_2x_3x_5 + x_1x_2x_4x_5 + x_1x_2x_4 + x_1x_3x_4x_5 + x_1x_3 + x_1x_4x_5 + x_2x_3x_4 + x_2x_3x_5 + x_2x_3 + x_2 + x_3x_5 + x_4x_5 + x_4 + 1$.

Then we know that $AI(f) = 3$, $deg(f) = 4$. So if we apply the conventional algebraic attack on this model, we can get equation system with degree no less than $AI(f) = 3$.

In the following, we will show the results when we apply the method in Section 2 on the augmented function, which is constructed by the expressions of the keystream bits of different clocks. In this case, we construct the augmented function $F^2 : \mathbb{F}_2^5 \to \mathbb{F}_2^2$ as follows:

$$F^2 = (f_1, f_2) = (f(x_1, x_2, ..., x_5), f(M(x_1, x_2, ..., x_5)^T).$$

Where M is the companion matrix of $p(x)$.

$$M = \begin{pmatrix} 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \\ 1\,0\,1\,0\,0 \end{pmatrix} \tag{5}$$

Similar with the attack on multi-output boolean function, we mount algebraic attack on this model and get the following equation:

$$f_1(x_1 + x_3 + x_4 + x_5 + 1) + f_2(x_1x_4 + x_2 + x_3 + 1) = 0.$$

$AI(F^2) = 1 < AI(f) = 3$. Then we can get equation system of degree less than 3.

Remark 2. This example indicates that the algebraic immunity of the multi-output boolean function is less than or equal to that of the component functions, which is consistent with Corollary 1.

According to Definition 7, we can learn that if the LFSR-based stream cipher is equidistant, there is no need to involve all the initial state bits of the LFSR in the augmented function. For instance, assume the length of the LFSR is m and the filter function is of k-variable ($m > 10k$), if we adapt 2-th augmented function, then we only need to consider $k + 1$ variables instead of the whole m initial state bits. Hence the number of variables involved in the augmented function can be reduced in a large extent, which cuts down the cost of deriving the low degree boolean functions that satisfy Case 1, Case 2 or Case 3. To illustrate the method intuitively, we would like to give the following example.

Example 5. Let the length of the LFSR be $m = 128$. The generator polynomial is $p(x) = x^{11} + x^2 + 1$. The filter function is a 11-variable Carlet-Feng function. The support of f is $\{0, 1, \alpha, \alpha^2, ..., \alpha^{2^{11-1}-2}\}$, where α is a primitive element of the finite field F_2^{12}. The stream cipher is equidistant.

If we use the whole initial state bits as the variables of the augmented function, then the number of variables is 128, which is almost impossible for us to derive the low degree functions by using computer.

We take advantage of the equidistant feature of the cipher and adapt the augmented function $F^2 := (f_1, f_2)$, where $f_1 = f, f_2 = f(x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12})$. We can check that $AI(f_1) = AI(f_2) = 6$.

Then the number of variables for F^2 is 12, less than $m = 128$. We search the low degree boolean functions quickly and get the following equation of degree 2:

$$(f_1 + f_2)(x_1x_{12} + x_1 + x_{12} + 1).$$

In the same way, we can find low degree equations with the other initial state bits as the variables.

Remark 3. The examples in this section are very meaningful. They suggests for the LFSR-based single-output stream ciphers, when the filter function is of optimum algebraic immunity, it is possible to get equation systems of degree less than the algebraic immunity.

5 Conclusion

This paper provides several new characterizations for algebraic immunity of multi-output Boolean functions. These descriptions might be helpful for a clear understanding of algebraic immunity of multi-output Boolean functions given by Armknecht and Krause. Some examples are given to illustrate these results.

References

1. Armknecht, F., and Krause, M: Algebraic attacks on combiners with memory. In D. Boneh, editor, CRYPTO, volume 2729 of Lecture Notes in Computer Science, pages 162-175. Springer, 2003.
2. Armknecht, F., Krause, M.: Constructing single-and multi-output boolean functions with maximal algebraic immunity. LNCS. 4052, pp. 180-191, 2006.
3. Becker, E., Weispfenning, V.: Gröbner bases, A computational approach to commutative algebra. Graduate texts in Mathematics 141, Springer, 1993
4. Courtois, N., Meier, W.: Algebraic attacks on stream ciphers with linear feedback, Eurocrypt 2003, Lecture notes in computer science 2656, pp. 345-359, 2003.
5. Courtois, N., Meier, W.: Algebraic attacks on stream ciphers with linear feedback, extended version of the Eurocrypt 2003.
6. Courtois, N.: Algebraic attacks on combiners with memory and several outputs. ICISC 2004, LNCS 3506, pp.3-20, 2005.
7. Courtois, N.: Higher order correlation attacks, XL algorithm and cryptanalysis of Toyocrypt, ICISC 2002, November 2002, Seoul, Korea, LNCS 2587, pp. 182-199, Springer.
8. Extended slides by Courtois. *www.nicolascourtois.com/papers/toyolili_slides.pdf*.
9. Carlet, C., Feng, K.: An Infinite Class of Balanced Functions with Optimal Algebraic Immunity, Good Immunity to Fast Algebraic Attacks and Good Nonliearity. ASIACRYPT, LNCS 5350, pp. 425-440, 2008.
10. Carlet, C., Dalai, D.K., Gupta, K. C., Maitra, S.: Algebraic immunity for cryptographically significant Boolean functions: analysis and construction. IEEE Trans. Inform. Theory 52(7), 3105-3121, 2006.
11. Claude C., Xiangyong Z., Chunlei L. and Lei H.: Further Properties of several classes of Boolean functions with optimum algebraic immunity. Designs, Codes and Cryptography, volume 52, vumber 3, pp. 303-338, 2009.
12. Deepak K. D., Subhamoy M. and Sumanta S.: Basic Theory in Construction of Boolean Functions with Maximum possible Annihilator Immunity. Designs, Codes and Cryptography, pp. 41-58, 2006.
13. Feng, K., Liao, Q., Yang, J.: Maximum Values of Gneralized Algebraic Immunity. Designs, Codes and Cryptography,volume 50, number 2, pp. 243-252, 2009.
14. Qichun Wang, Thomas Johansson: Higher Order Algebraic Attacks on Stream Ciphers. http://eprint.iacr.org/2012/013.pdf, 2012.

15. Wang, Q., Peng,J., Kan, H and Xue, X: Constructions of cryptographically significant Boolean functions using primitive polynomials. IEEE Trans. Inf. Theory, vol. 56, no. 6, pp. 3048C3053, 2010.
16. Rueppel, R: Analysis and Design of Stream Ciphers. Berlin, Germany, Springer-Verlag, 1986.
17. Meier, W., Pasalic, E., Carlet, C.: Algebraic attacks and decomposition of Boolean functions. Eurocrypt 2004, Lecture notes in computer sciences vol. 3027, pp. 474-491, 2004.