On the Algebraic Immunity of Symmetric Boolean Functions

An Braeken and Bart Preneel

Katholieke Universiteit Leuven
Dept. Elect. Eng.-ESAT/SCD-COSIC,
Kasteelpark Arenberg 10, 3001 Heverlee, Belgium
{an.braeken,bart.preneel}@esat.kuleuven.be

Abstract. In this paper, we analyze the algebraic immunity of symmetric Boolean functions. The algebraic immunity is a property which measures the resistance against the algebraic attacks on symmetric ciphers. We identify a set of lowest degree annihilators for symmetric functions and propose an efficient algorithm for computing the algebraic immunity of a symmetric function. The existence of several symmetric functions with maximum algebraic immunity is proven. In this way, we have found a new class of functions which have good implementation properties and maximum algebraic immunity.

1 Introduction

Symmetric functions have the property that the function value is determined by the Hamming weight of the input vector. Therefore, a symmetric function in n variables can be defined by a vector of length $n+1$ which represents the function values of the different Hamming weights of the input vectors. For this reason, symmetric functions are very interesting functions in order to obtain low memory in software. In hardware implementation, only a low number of gates is required [15]. Properties such as balancedness and resiliency, propagation characteristics and nonlinearity of symmetric functions are studied by Canteaut and Videau [3]. It is shown that these functions do not behave very well in general with respect to a combination of the properties such as nonlinearity, degree, and resiliency, which are important properties for resisting distinguishing and correlation attacks [2].

In 2002, several successful algebraic attacks on stream ciphers were proposed by Courtois [5]. The success of these attacks do not depend on the classical properties of nonlinearity or resiliency, but mainly on the weak behavior with respect to the property of algebraic immunity. In this paper we study the resistance of the symmetric functions against the algebraic attacks. We identify a set of polynomials whose linear combinations lead to lowest degree annihilators of a symmetric function. Since the size of this set is very small in comparison with the general case, the algorithm for computing the algebraic immunity (AI) of a symmetric function becomes much more efficient. We prove the existence of several symmetric functions with optimal algebraic immunity.

First, Sect. 2 deals with some background on Boolean functions and more in particular on symmetric Boolean functions. Based on the identification of a set of lowest degree annihilators of a symmetric function, we propose an algorithm for computing the algebraic immunity of symmetric functions in Sect. 3. Sect. 4 presents several classes of symmetric functions which possess maximum algebraic immunity. Finally, we conclude in Sect. 5.

2 Background

Let us first recall the basic background on Boolean functions together with some properties of symmetric Boolean functions which were proven by Canteaut and Videau [13].

Let $\mathbb{F}_2^n$ be the set of all n-tuples of elements in the field $\mathbb{F}_2$ (Galois field with two elements), endowed with the natural vector space structure over $\mathbb{F}_2$. An element $\overline{u} = (u_0, \ldots, u_{n-1})$ in $\mathbb{F}_2^n$ can be represented by an integer $\mathbb{Z}_{2^n}$ belonging to the interval $[0, 2^n - 1]$, *i.e.*, $u = \sum_{i=0}^{n-1} u_i 2^i$. We will use both notations interchangeable in the rest of the paper.

A Boolean function f on $\mathbb{F}_2^n$ is a mapping from $\mathbb{F}_2^n$ onto $\mathbb{F}_2$. It can be uniquely represented by the truth table (TT) which is the vector of length 2^n consisting of its function values. The support of the function f, $\sup(f)$ contains all vectors $\overline{x}$ for which $f(\overline{x}) = 1$. The (Hamming) weight $\mathrm{wt}(\overline{v})$ of a vector $\overline{v} \in \mathbb{F}_2^n$ is defined as the number of nonzero positions.

Another unique representation, called the ANF, is the polynomial

$$f(\overline{x}) = \bigoplus_{(a_0,\ldots,a_{n-1})\in\mathbb{F}_2^n} h(a_0, \ldots, a_{n-1}) x_0^{a_0} \ldots x_{n-1}^{a_{n-1}}, h(\overline{a}) = \sum_{\overline{x} \preceq \overline{a}} f(\overline{x}), \text{for any } \overline{a} \in \mathbb{F}_2^n \,,$$

where $\overline{x} \preceq \overline{a}$ means that $x_i \leq a_i$ for all $0 \leq i \leq n-1$. The degree of the polynomial determines the algebraic degree of this function. The ANF of a function consists of the modulo 2 sum of polynomials $(x_0 \oplus a_0 \oplus 1) \cdots (x_{n-1} \oplus a_{n-1} \oplus 1)$ for all $\overline{a} \in \mathbb{F}_2^n$ such that $f(\overline{a}) = 1$. Denote the all-zero function or vector by $\overline{0}$ and the all-one function or vector by $\overline{1}$.

The Walsh transform W_f of a function f on $\mathbb{F}_2^n$ is defined as the real valued transformation

$$W_f(\overline{w}) = \sum_{\overline{x}\in\mathbb{F}_2^n} (-1)^{f(\overline{x}) + \overline{w}\cdot\overline{x}} \,.$$

From the Walsh transform, we derive the property of nonlinearity $N_f = 2^{n-1} - \frac{1}{2} \max_{\overline{w}\in\mathbb{F}_2^n} |W_f(\overline{w})|$, which represents the smallest distance between a Boolean function and any affine function [11].

As response to the algebraic attacks, Meier et al. [10] introduced the concept of algebraic immunity (AI) for a Boolean function f on $\mathbb{F}_2^n$. This measure defines the lowest degree of a non-zero function g from $\mathbb{F}_2^n$ into $\mathbb{F}_2$ for which $f \cdot g = \overline{0}$ or $(f \oplus \overline{1}) \cdot g = \overline{0}$. The function g for which $f \cdot g = \overline{0}$ is called an *annihilator*

function of f. The set of all annihilators of f is denoted by $An(f)$. The AI is upper bounded by $\lceil \frac{n}{2} \rceil$ as proven in [4].

Symmetric functions have the property that the function value of all vectors with the same weight is equal. Consequently, the truth table of the symmetric function on $\mathbb{F}_2^n$ can be replaced by a vector v_f of length $n+1$ where the components $v_f(i)$ for $0 \leq i \leq n$ represent the function value for vectors of weight i. The vector v_f is called the value vector (VV) of the symmetric function f.

The ANF representation for a symmetric function can also be replaced by a shorter form [3, Prop. 2], called the simplified ANF (SANF). Denote the homogeneous symmetric function, which is the function that contains all terms of degree i for $0 \leq i \leq n$, by σ_i. Then, the SANF is a polynomial in $\mathbb{F}_2[x_0, \ldots, x_{n-1}]/(x_0^2 - x_0, \ldots, x_{n-1}^2 - x_{n-1})$ with basis elements the homogeneous symmetric functions σ_i for $0 \leq i \leq n$:

$$f(\overline{x}) = \bigoplus_{i=0}^{n} \lambda_f(i)\sigma_i, \quad \lambda_f(i) = \sum_{k \preceq i} v_f(k), \text{for } 0 \leq i \leq n\,.$$

The vector $\lambda_f = (\lambda_f(0), \ldots, \lambda_f(n))$ is called the simplified ANF vector (SANF vector).

3 Annihilators of Symmetric Functions

We first distinguish a set of polynomials whose linear combinations lead to lowest degree annihilators of a symmetric function. Based on this set, we propose an efficient algorithm for computing the AI of a symmetric Boolean function.

Denote the homogeneous symmetric function of degree i which depends on the j variables $\{x_{n-j}, x_{n-j+1}, \ldots, x_{n-1}\}$ with $j \geq i$ by σ_i^j. We also use the notation of P_l^k to represent the set of polynomials where each polynomial contains all k variables $\{x_0, \ldots, x_{k-1}\}$ and consists of the product of at most l factors where every factor is either the sum of two variables, one variable, or the complement of one variable. Consequently $\lceil \frac{k}{2} \rceil \leq l$. Note that the variables in the polynomials P_l^k play the same role, which means that changing the indices of the variables does not introduce new polynomials in P_l^k. Therefore, we define the role of the variables $\{x_0, \ldots, x_{k-1}\}$ in the polynomials of P_l^k as follows. Depending on l, the first factors involving the first variables (starting from $x_0, x_1, \ldots$) may consist of one variable, the complement of one variable or the sum of two variables. The following factors may consist of one variable and the sum of two variables, while the last factors consist of the sum of two variables.

Example 1. If $\lceil \frac{k}{2} \rceil = l$, only the polynomial $(x_0 \oplus x_1)(x_2 \oplus x_3) \cdots (x_{k-2} \oplus x_{k-1})$ for k even and the polynomials $x_0(x_1 \oplus x_2)(x_3 \oplus x_4) \cdots (x_{k-2} \oplus x_{k-1})$ and $(x_0 \oplus 1)(x_1 \oplus x_2)(x_3 \oplus x_4) \cdots (x_{k-2} \oplus x_{k-1})$ for k odd belong to $P^k_{\lceil \frac{k}{2} \rceil}$. If $\lceil \frac{k}{2} \rceil = l-1$, the polynomials $x_0x_1(x_2 \oplus x_3) \cdots (x_{k-2} \oplus x_{k-1})$, $(x_0 \oplus 1)x_1(x_2 \oplus x_3) \cdots (x_{k-2} \oplus x_{k-1})$, $(x_0 \oplus 1)(x_1 \oplus 1)(x_2 \oplus x_3) \cdots (x_{k-2} \oplus x_{k-1})$, $(x_0 \oplus x_1)(x_2 \oplus x_3) \cdots (x_{k-2} \oplus x_{k-1})$, belong to $P^k_{\lceil \frac{k}{2} \rceil + 1}$ for k even.

The goal of this section is to show that at least one of the lowest degree annihilators with degree strictly less than $\lceil \frac{n}{2} \rceil$ of a symmetric function on $\mathbb{F}_2^n$ is a linear combination of the polynomials of the following form:

$$
\begin{aligned}
n \text{ even:} \quad & \sigma_0^2 P_{\frac{n}{2}-1}^{n-2}, \sigma_0^3 P_{\frac{n}{2}-1}^{n-3}, \ldots, \sigma_0^{n-1} P_{\frac{n}{2}-1}^{1}, \sigma_0, \\
& \sigma_1^4 P_{\frac{n}{2}-2}^{n-4}, \ldots, \sigma_1^{n-1} P_{\frac{n}{2}-2}^{1}, \sigma_1, \ldots, \sigma_{\frac{n}{2}-2}^{n-2} P_1^2, \sigma_{\frac{n}{2}-2}^{n-1} P_1^1, \sigma_{\frac{n}{2}-2}, \sigma_{\frac{n}{2}-1} \\
n \text{ odd:} \quad & \sigma_0^1 P_{\lceil \frac{n}{2} \rceil -1}^{n-1}, \sigma_0^2 P_{\lceil \frac{n}{2} \rceil -1}^{n-2}, \ldots, \sigma_0^{n-1} P_{\lceil \frac{n}{2} \rceil -1}^{1}, \sigma_0, \\
& \sigma_1^3 P_{\lceil \frac{n}{2} \rceil -2}^{n-3}, \ldots, \sigma_1^{n-1} P_{\lceil \frac{n}{2} \rceil -2}^{1}, \sigma_1, \ldots, \sigma_{\lceil \frac{n}{2} \rceil -2}^{n-2} P_1^2, \sigma_{\lceil \frac{n}{2} \rceil -2}, \sigma_{\lceil \frac{n}{2} \rceil -1}.
\end{aligned}
$$

As $\lceil \frac{k}{2} \rceil \leq l$, the functions σ_k for $k \in \{0, \ldots, \lceil \frac{n}{2} \rceil - 1\}$ depend on $2k+2, 2k+3, \ldots, n$ variables for n even and $2k+1, 2k+2, \ldots, n$ variables for n odd in order to obtain annihilators of degree less than or equal to $\lceil \frac{n}{2} \rceil - 1$. We will call this set of polynomials AN_S. We now give some examples of annihilators which consist of the linear combination of polynomials in AN_S.

Example 2. Let $n = 16$, and suppose f is a symmetric Boolean function on $\mathbb{F}_2^n$ with value vector v_f that satisfies $v_f(i) = 0$ for $i \in \{6, 7, 10, 11\}$. Then the function $g(\overline{x}) = \sigma_2^9 x_0 (x_1 \oplus x_2)(x_3 \oplus x_4)(x_5 \oplus x_6)$ represents an annihilator of the function f. This follows from the fact that σ_2^9 is equal to 1 only for vectors in $\mathbb{F}_2^9$ with weight equal to 2,3,6,7. The function $x_0(x_1 \oplus x_2)(x_3 \oplus x_4)(x_5 \oplus x_6)$ is equal to 1 only for a subset of vectors in $\mathbb{F}_2^7$ with weight 4. Consequently the function g is equal to 1 only for a subset of vectors of weight 6,7,10,11.

If the value vector in the coordinates 2 and 6 is equal to c where $c \in \{0, 1\}$ for a symmetric function f in 10 variables, then $(x_0 \oplus 1)(\sigma_2^9 \oplus \sigma_3^9)$ represents an annihilator with degree 3 of f if $c = 0$, or $f \oplus \overline{1}$ if $c = 1$.

Theorem 1. *One of the lowest degree annihilators of a symmetric function can be constructed by means of a linear combination of the polynomials in AN_S.*

Proof. Annihilators of symmetric functions are equal to 0 for all vectors of a certain weight which belong to the support of the corresponding symmetric function. But the annihilators can be 0 or 1 for vectors which do not belong to the support of the symmetric function. Therefore, an example of an annihilator is the one which consists of the product of a symmetric function which depends on the last $n - k$ variables in order to guarantee that the function value is 1 for vectors of the same weight, together with a polynomial that depends on the other k variables and which is 1 for a subset of vectors with fixed weight. The polynomials P_l^k in the polynomials of AN_S are constructed in such way that they are equal to 1 only for a subset of vectors which have exactly one fixed and equal weight. Corollary 1, which is based on Lemma 1, proves that the annihilators constructed by means of a linear combination of the polynomials in AN_S have lowest possible degree by showing that if one of the factors of the polynomials P_l^k would consist of more than 3 variables (in order to decrease the degree), then there also exists an annihilator constructed by means of linear combinations of the polynomials of AN_S whose support is contained in the support of

this annihilator and which has smaller or equal degree. Therefore, we first prove Lemma 1. □

Remark 1. We note that the annihilators constructed by linear combinations of the polynomials in AN_S do not determine the complete basis of the ideal of annihilators with degree strictly less than $\lceil \frac{n}{2} \rceil$ of a symmetric function. For instance, the function $x_0\sigma_3$ on $\mathbb{F}_2^{10}$ is annihilator of all symmetric functions on $\mathbb{F}_2^{10}$ for which $v_f(4) = v_f(8) = 0$. But the function $x_0\sigma_3^9 \in \mathrm{AN}_S$ also satisfies this property. Both functions are linearly independent. Also note that the variables of the polynomials P_l^k play the same role in the representation, and that they only depend on the first k variables. This is possible due to the symmetry of the symmetric function. Since we are only interested in the existence of at least one annihilator in order to determine the AI of the function, we can restrict us for the search of annihilators into the set functions obtained by linear combinations of the polynomials in AN_S.

Lemma 1. *Let $r \geq 3$ and $n \geq r-1$. Define S_i^n as the symmetric function on n variables of degree i,*

$$S_i^n = \bigoplus_{0 \leq k \leq i} c_k^S \sigma_k^n \text{ where } c_k^S \in \{0,1\} \text{ for all } 0 \leq k \leq i.$$

Denote the set of weights in the support of S_i^n by V_S. Define also $S_{i-(r-1)}^{n-(r-1)} = \bigoplus_{0 \leq k \leq i} c_k^S \sigma_{k-(r-1)}^{n-(r-1)}$ where $\sigma_i = 0$ for $i < 0$ and denote its support of the value vector by $V_{S'}$. Then

$$\{a + r - 1 : a \in V_{S'}\} \subseteq \{a, a+2, \ldots, a+r-1 : a \in V_S\} \quad (1)$$

$$\{a + r : a \in V_{S'}\} \subseteq \{a+1, a+3, \ldots, a+r : a \in V_S\} \quad (2)$$

We refer to an extended version of the paper for the proof of this lemma.

Example 3. Let $n = 10, r = 3$. The support of the value vector of the function $\sigma_0^{10} \oplus \sigma_1^{10} \oplus \sigma_2^{10} \oplus \sigma_5^{10}$ belongs to $V_S = \{0,3,4,5,8\}$. The support of the value vector of $\sigma_0^8 \oplus \sigma_3^8$ belongs to $V_{S'} = \{0,1,2,4,5,6,8\}$. The theorem implies that $\{2,3,4,6,7,8,10\} \subseteq \{0,2,3,4,5,6,7,8,10\}$.

Directly from Lemma 1, we can derive

Corollary 1. *Let r be odd and $r \geq 3$, then the support of $S_i^{n-r}(x_0 \oplus \cdots \oplus x_{r-1})$ contains the support of $S_{i-(r-1)}^{n-(2r-1)} x_0(x_1 \oplus x_2) \cdots (x_{2r-3} \oplus x_{2r-2})$. The support of $S_i^{n-r}(x_0 \oplus \cdots \oplus x_{r-1} \oplus 1)$ contains the support of $S_{i-(r-1)}^{n-(2r-1)}(x_0 \oplus 1)(x_1 \oplus x_2) \cdots (x_{2r-3} \oplus x_{2r-2})$. Both pairs of functions have the same degree $i+1$.*

Let r be even and $r \geq 4$, then the support of $S_i^{n-r}(x_0 \oplus \cdots \oplus x_{r-1})$ contains the support of $S_{i-(r-2)}^{n-(2r-2)}(x_0 \oplus x_1)(x_2 \oplus x_3) \cdots (x_{2r-3} \oplus x_{2r-4})$. Both functions have the same degree $i+1$. The support of $S_i^{n-r}(x_0 \oplus \cdots \oplus x_{r-1} \oplus 1)$ contains the support of $S_{i-r}^{n-2r}(x_0 \oplus x_1)(x_2 \oplus x_3) \cdots (x_{2r-1} \oplus x_{2r-2})$. The last function has degree i in comparison with degree $i+1$ of the first function. This equation also holds for $r = 2$.

We conclude that if one or more factors of the polynomial P_l^k would consist of the complement of two terms or more than three terms, then there always exists an annihilator constructed by means of a linear combination of polynomials in AN_S which has degree smaller or equal and whose support is contained in the support of that annihilator.

Let us now compute the number of polynomials in the set AN_S.

Theorem 2. *The number N of polynomials in AN_S is equal to*

$$N = 3 \cdot 2^{\lceil \frac{n}{2} \rceil} - 2 \cdot \left\lceil \frac{n}{2} \right\rceil - 3 \,.$$

Proof. We will compute the number for n even. In a similar way, the result is obtained for n odd. Denote R_k^n for n even and $0 \leq k \leq \frac{n}{2} - 1$ as the sum of all elements which have σ_k^i for $i = 2k+2, \ldots, n$ as factor, *i.e.*, the sum of all elements of the sets $P^i_{\frac{n}{2}-k-1}$ for $i = 0, \ldots, n-(2k+2)$:

$$R_k^n = \sum_{i=0}^{n-(2k+2)} |P^i_{\frac{n}{2}-k-1}| \,.$$

For $i = n-(2k+2)$, there is exactly one element in $P^{n-(2k+2)}_{\frac{n}{2}-k-1}$, namely the polynomial $(x_1 \oplus x_2) \cdots (x_{n-2k-2} \oplus x_{n-2k-3})$. Every decrease of i until $i = \frac{n}{2} - k - 1$ with 1 gives one more degree of freedom, which leads to a factor of two more for the possible polynomials in $P^i_{\frac{n}{2}-k-1}$. For instance, suppose the polynomial $P^i_{\frac{n}{2}-k-1}$ has the form $(x_1 \oplus x_2)(x_3 \oplus x_4) \cdots$ at step i. After removing one variable at step $i-1$, we can have two additional elements in $P^{i-1}_{\frac{n}{2}-k-1}$ namely $x_1(x_2 \oplus x_3) \cdots$ and $(x_1 \oplus 1)(x_2 \oplus x_3) \cdots$. Removing another variable leads again to two more polynomials: $(x_1 \oplus x_2) \cdots$, $x_1 x_2 \cdots$, $(x_1 \oplus 1) x_2 \cdots$, $(x_1 \oplus 1)(x_2 \oplus 1) \cdots$. For $i < \frac{n}{2} - k - 1$, due to the smaller number of variables, the total number of polynomials decreases again with a factor of 2. Therefore, we have that for $0 \leq k \leq \frac{n}{2} - 1$:

$$R_k^n = 2 \sum_{i=0}^{\frac{n}{2}-k-2} 2^i + 2^{\frac{n}{2}-k-1} \,.$$

Consequently, the total number of terms belonging to class 2 is equal to

$$N = \sum_{k=0}^{\frac{n}{2}-1} R_k^n = 2 \sum_{i=1}^{\lceil \frac{n}{2} \rceil - 1} (2^i - 1) + 2^{\lceil \frac{n}{2} \rceil} - 1 \,.$$

□

Example 4. For $n = 14$, we have that

$$\begin{aligned}
\sigma_0 &\to (|P_6^{12}|, \ldots, |P_6^0|) = (1, 2, 4, 8, 16, 32, 64, 32, 16, 8, 4, 2, 1)\\
\sigma_1 &\to (|P_5^{10}|, \ldots, |P_5^0|) = (1, 2, 4, 8, 16, 32, 16, 8, 4, 2, 1)\\
\sigma_2 &\to (|P_4^8|, \ldots, |P_4^0|) = (1, 2, 4, 8, 16, 8, 4, 2, 1)\\
\sigma_3 &\to (|P_3^6|, \ldots, |P_3^0|) = (1, 2, 4, 8, 4, 2, 1)\\
\sigma_4 &\to (|P_2^4|, \ldots, |P_2^0|) = (1, 2, 4, 2, 1)\\
\sigma_5 &\to (|P_1^2|, \ldots, |P_1^0|) = (1, 2, 1)\\
\sigma_6 &\to |P_0^0| = 1
\end{aligned}$$

3.1 An Algorithm for Computing AI

As shown in the previous section, one of the lowest degree annihilators of degree less than $\lceil \frac{n}{2} \rceil$ consists of a linear combination of the N polynomials in AN_S. As determined in Theorem 2, the size of the set AN_S is much smaller than the number of all polynomials of degree less than $\lceil \frac{n}{2} \rceil$ which is equal to $\sum_{i=0}^{\lceil \frac{n}{2} \rceil - 1} \binom{n}{i}$. Table 1 shows the comparison between both numbers for dimensions $n = 2k$ with $5 \leq k \leq 10$. We can conclude that the difference increases with the dimension.

Table 1. Comparison of the size of annihilator-set

n	10	12	14	16	18	20
$\sum_{i=0}^{\lceil \frac{n}{2} \rceil - 1} \binom{n}{i}$	386	1 586	6 476	26 333	106 762	431 910
$\lvert \mathrm{AN}_S \rvert$	83	177	367	749	1 524	3 049

The main goal of the algorithm that computes the AI of a function consists in finding suitable linear combinations within these terms. Consequently, roughly speaking the complexity for computing the AI of a symmetric function can be upper bounded by $N^{2.81} \approx 58 \cdot 2^{1.4n}$, where 2.81 corresponds to the exponent for Gaussian elimination [1].

Moreover, the additional tricks presented in [10] can be used to accelerate the algorithm even further. Due to the fact that we have much less functions to combine in the algorithm for computing the AI of a symmetric function, the AI of any arbitrary symmetric function can be computed for much larger dimensions.

Instead of checking the whole set of 2^{n+1} symmetric functions for functions on $\mathbb{F}_2^n$ with maximum AI, we first present some properties on the value vector of a symmetric function with maximum AI. These properties can be immediately derived from the existence of the annihilators constructed by means of linear combinations of polynomials in AN_S.

3.2 Properties

Theorem 3. *Let f be a symmetric Boolean function on $\mathbb{F}_2^n$ with value vector v_f. If $v_f(\lceil \frac{n}{2} \rceil - 1) = v_f(\lceil \frac{n}{2} \rceil + 1)$ for all n, or in addition for n odd $v_f(\lceil \frac{n}{2} \rceil - 2) = v_f(\lceil \frac{n}{2} \rceil)$, then f can not have maximum AI.*

Theorem 4. *Let $2^j \leq n < 2^{j+1} - 1$ where $j \geq 1$ and f be a symmetric Boolean function on $\mathbb{F}_2^n$ with value vector v_f. Define for all $0 \leq i < 2^{j-1}$ the set $V_i = \{l : l \equiv i \mod 2^{j-1} \text{ for } 0 \leq l < n\}$. If there exists an $i \in \{0, \ldots, 2^{j-1} - 1\}$ such that $v_f(k) = 0$ (resp. 1) for all $k \in V_i$, then the AI of f is less than or equal to $2^{j-1} - 1$. For $n = 2^{j+1} - 1$ where $j \geq 1$, the value vector of f should be of the form $(\overline{a}|\overline{a}^c)$ where $\overline{a} \in \mathbb{F}_2^j$ in order to reach the maximum AI.*

Finally, we want to mention that also the condition on the weight of a Boolean function, as derived in [6], is very strong for symmetric functions with an odd number of variables. It implies that maximum AI can only be obtained for balanced functions if n is odd. A large set of balanced functions in n odd are the trivially balanced functions, *i.e.*, the functions with value vector $v_f(i) = v_f(n - i) \oplus 1$ for all $0 \leq i \leq \lfloor \frac{n}{2} \rfloor$. In fact, the trivially balanced functions form the whole set of balanced functions for n odd and $n \leq 128$, except in dimensions $n \in \{13, 29, 31, 33, 35, 41, 47, 61, 63, 73, 97, 103\}$ as shown in [14].

3.3 Experiments

For the computation of the AI, we can use a more efficient algorithm than the algorithm of [10] as explained above and thus reach higher dimensions.

If n is odd, the condition of trivially balancedness is very powerful. We checked until $n \leq 17$ and can conclude that the only trivially balanced functions with maximum AI have value vector v_f such that

$$v_f(i) = \begin{cases} 0 \text{ for } i < \lceil \frac{n}{2} \rceil \\ \\ 1 \text{ for } i \geq \lceil \frac{n}{2} \rceil . \end{cases} \tag{3}$$

In [12], the complete set of non-trivially balanced functions for $n = 13$ is described. From this description, we derive that the AI of the non-trivial balanced functions in 13 variables is less than or equal to 3 due to Theorem 4. Therefore, we conclude that all symmetric functions in n odd and $n \leq 17$ with maximum AI have value vector defined by (3). We will show in the next section that a symmetric function with such value vector always has maximal AI for every n odd. Moreover, it can be easily proven that for $n = 2^i - 1, 2^i + 1$, with $i \geq 2$, only the trivially balanced functions with value vector determined by (3) have maximum AI. In these dimensions, the property of Theorem 4 is very powerful.

For n even, we found more symmetric functions with maximum AI. In the next section, we will theoretically prove the maximum AI for some of these functions. The theorems will cover all symmetric functions with maximum AI

in dimensions less than or equal to 12 and all but one in dimensions 14 and 16. We refer to the extended version of the paper for the complete set of symmetric Boolean functions with maximum AI in dimensions $n = 6, 8, 10, 12, 14, 16$.

4 Symmetric Functions with Maximum AI

In this section, we show the existence of several symmetric functions with maximum AI for all dimensions n. Let us first recall that the property of AI is invariant under affine transformation in the input variables, *i.e.*, $f(\overline{x})$ and $f(\overline{x}A \oplus \overline{b})$, where A is an $n \times n$ nonsingular matrix and $\overline{b} \in \mathbb{F}_2^n$ will have the same AI. This follows from the fact that if g is annihilator of f, then $g(\overline{x}A \oplus \overline{b})$ is annihilator of $f(\overline{x}A \oplus \overline{b})$.

However, the AI of two functions $f(\overline{x})$ and $f(\overline{x}) \oplus \overline{c} \cdot \overline{x}$ with $\overline{c} \in \mathbb{F}_2^n$ can differ at most with 1. This can be easily seen as follows. Let g be annihilator of f such that $f(\overline{x}) \cdot g(\overline{x}) = 0$, then $g(\overline{x})(\overline{c} \cdot \overline{x} \oplus \overline{1})$ is annihilator of $(f(\overline{x}) \oplus \overline{c} \cdot \overline{x})$ because $(f(\overline{x}) \oplus \overline{c} \cdot \overline{x})g(\overline{x})(\overline{c} \cdot \overline{x} \oplus \overline{1}) = f(\overline{x})g(\overline{x})(\overline{c} \cdot \overline{x} \oplus 1) \oplus (\overline{c} \cdot \overline{x})g(\overline{x})(\overline{c} \cdot \overline{x} \oplus \overline{1}) = 0$. The last equality follows from the fact that $\overline{c} \cdot \overline{x} \oplus \overline{1}$ is annihilator of $\overline{c} \cdot \overline{x}$.

We now investigate the affine transformations on the input variables which will transform a symmetric function into a new symmetric function. Due to the following lemma proven by Dawson and Wu, we only need to check the transformations $\overline{x} \mapsto \overline{x}A \oplus c\overline{1}$, where A is a nonsingular $n \times n$ binary matrix and $c \in \mathbb{F}_2$.

Lemma 2. *[8] Let $a \in \mathbb{F}_2^n \setminus \{\overline{0}, \overline{1}\}$. If f is a symmetric Boolean function, then $f(\overline{x} \oplus \overline{a})$ is symmetric if and only if f is affine.*

Theorem 5. *In n even, the only binary linear transformation on the input variables of a symmetric function that will compute a new symmetric function on $\mathbb{F}_2^n$ is the transformation $T = \overline{x} \mapsto \overline{x}A$, where A is a nonsingular $n \times n$ matrix over $\mathbb{F}_2$ with the property that the sum of the elements in each row and column of A is equal to $n - 1$. For n odd, no such transformations exist.*

The transformation $(x_0, \ldots, x_{n-1}) \mapsto (x_0 \oplus 1, \ldots, x_{n-1} \oplus 1)$ for all n will map a symmetric function with value vector v_f to a symmetric function with value vector equal to the reverse of this value vector, i.e., v_f^r.

Proof. A minimal requirement for a binary linear transformation $x \mapsto \overline{x}A$ which maps a symmetric function onto a symmetric function is that the weight W of the columns and rows of A is equal, since all variables play the same role in a symmetric function. If W is greater than 1 and smaller than $n - 1$, the transformation is not bijective or does not lead to a symmetric function.

Consider n even and $W = n-1$. If $\mathrm{wt}(\overline{x})$ is odd and equal to i, then we show that $\mathrm{wt}(\overline{x}A)$ is equal to $n - i$. Denote by $V = \{i : x_i \neq 0\}$. The coordinates j with $j \in \{0, \ldots, n-1\}$ in the vector $\overline{x}A$ are 1 if and only if the elements on the corresponding column j of A are 1 exactly on the i positions of the set V. (Note that it is not possible that there are $i - 2k$ with $k \geq 1$ elements in the columns of A which are 1 and $2k$ elements which are 0 due to the fact that $W = n - 1$.)

The number of such columns in A is equal to $\binom{n-i}{n-i-1} = n-i$ for i odd and $1 \leq i \leq n-1$.

Now we show that if $\mathrm{wt}(\overline{x})$ is even and equal to i, then $\mathrm{wt}(\overline{x}A) = i$. Denote by $V = \{i : x_i \neq 0\}$. The coordinates j with $j \in \{0, \ldots, n-1\}$ in the vector $\overline{x}A$ are 1 if and only if the elements on the corresponding column j of A are 1 on exactly $i-1$ positions of the set V. There are $\binom{i}{i-1} = i$ possibilities for this to occur.

For n odd, the transformation T is not bijective which follows immediately from the fact that vectors of weight 0 and n are both mapped onto vectors of weight 0.

Finally, since the transformation $(x_0, \ldots, x_{n-1}) \mapsto (x_0 \oplus 1, \ldots, x_{n-1} \oplus 1)$ maps a vector of weight i onto a vector of weight $n-i$, this transformation corresponds to the mapping of $v_f(i)$ onto $v_f(n-i)$ for every i with $0 \leq i \leq n$. □

We now present three basic classes of symmetric functions with maximum AI. We refer to the extended version of the paper for the proofs of the theorems in this section.

Class 1

Theorem 6. *The symmetric function f in $\mathbb{F}_2^n$ with value vector*

$$v_f(i) = \begin{cases} 0 \text{ for } i < \left\lceil \frac{n}{2} \right\rceil \\ 1 \text{ else} \end{cases} \tag{4}$$

has maximum AI. Let us denote this function f by F_k where k is equal to the threshold $\left\lceil \frac{n}{2} \right\rceil$.

Remark 2. The maximum AI of this class of symmetric functions was independently proven in [7] using a different proof method. This result was also presented at [2].

For n even, we prove that also the function which only differs from the threshold function $F_{\left\lceil \frac{n}{2} \right\rceil}$ in the function value of the vector $(1, \ldots, 1)$ has maximum AI. Denote the zero vector on $\mathbb{F}_2^{n+1}$ with 1 on position i by $\overline{e}_i$ for $0 \leq i \leq n$.

Theorem 7. *The symmetric function f with value vector $v_{F_{\left\lceil \frac{n}{2} \right\rceil}} \oplus \overline{e}_n$ in $\mathbb{F}_2^n$ for n even has maximum AI. The degree of f is equal to n if $n \neq 2^i$ for $i \geq 1$ and equal to 2^{i-1} else.*

Class 2

For $n \geq 8$ and even, we can distinguish another class of symmetric functions with maximum AI. These symmetric functions differ from $F_{\frac{n}{2}}$ in two symmetric positions such that they possess the same weight as $F_{\frac{n}{2}}$. Denote by $\overline{s}_i$ the all zero vector on $\mathbb{F}_2^{n+1}$ with 1 on positions $i, n-i$ for $0 \leq i < \frac{n}{2}$.

Theorem 8. *Let $n = 2k$ and $k \geq 4$. The symmetric function f with value vector $v_{F_{\frac{n}{2}}} \oplus \overline{s}_{k-4}$ on $\mathbb{F}_2^n$ has maximum AI.*

Again, the symmetric functions f which differ from the functions presented in Theorem 8 only in the all-one vector have maximum AI for $n \geq 10$. This can be obtained by using the proof technique of Theorem 7 for showing the non-existence of annihilators with degree less than $\frac{n}{2}$ for f and the proof technique of Theorem 8 for $f \oplus \overline{1}$.

Theorem 9. *Let $n = 2k$ and $k \geq 5$. The symmetric function f with value vector $v_{F_{\frac{n}{2}}} \oplus \overline{s}_{k-4} \oplus \overline{e}_n$ on $\mathbb{F}_2^n$ has maximum AI. The degree of f is equal to n if $n \neq 2^i$ with $i \geq 1$ and equal to 2^{i-1} else.*

We also present another class of functions which differs from $F_{\frac{n}{2}}$ in two symmetric positions. These functions coincide with the function defined in Theorem 7 for $n = 8$.

Theorem 10. *Let f be a symmetric function on $\mathbb{F}_2^n$ with n even. If $\binom{n}{\frac{n}{2}} \equiv 1 \mod 4$, then the function with value vector $v_{F_{\frac{n}{2}}} \oplus \overline{s}_0$ has maximum AI.*

Example 5. The numbers $n = 2^i$ for $i \geq 3$ satisfy the property that $\binom{n}{\frac{n}{2}} \equiv 1 \mod 4$.

Class 3

For n even, the third class of functions with maximum AI differs from $F_{\frac{n}{2}}$ in only one position. Therefore these functions have weight different from the weight of the functions of class 1 or 2.

Theorem 11. *Let f be a symmetric function on $\mathbb{F}_2^n$ with n even. For $1 \leq i < \lfloor \frac{n}{4} \rfloor$, if $\binom{\frac{n}{2}+t-i}{t} \equiv 1 \mod 2$ for all $t \in \{1, \ldots, i\}$, then the function f with value vector $v_{F_{\frac{n}{2}}} \oplus \overline{e}_{n-i}$ has maximum AI.*

Example 6. For $n = 14$, since $\binom{7}{1} \equiv 1 \mod 2$, the function value vector $v_{F_7} \oplus \overline{e}_{13}$ has maximum AI. Also $\binom{7}{3} \equiv 1 \mod 2, \binom{6}{2} \equiv 1 \mod 2, \binom{5}{1} \equiv 1 \mod 2$, and thus the function with value vector $v_{F_7} \oplus \overline{e}_{11}$ represents a function with maximum AI.

Functions Derived From Classes 1, 2, and 3

For n even, the symmetric functions from classes 1, 2, and 3 can be used to derive other symmetric functions by means of the affine transformation $(x_0, \ldots, x_{n-1}) \mapsto (x_0 \oplus x_1 \oplus \cdots \oplus x_{n-2}, x_1 \oplus x_2 \oplus \cdots \oplus x_{n-1}, \ldots, x_{n-1} \oplus x_0 \oplus \cdots \oplus x_{n-3})$. As already explained in the proof of Theorem 4, this transformation maps vectors of odd weight i to vectors with weight $n - i$. If the weight is even, then nothing is changed.

Corollary 2. *Let f be a symmetric functions on $\mathbb{F}_2^n$ which belongs to class 1 or 2. If $n = 4k$, then $f \oplus \sigma_1$ has maximum AI. If $n = 4k + 2$, then the symmetric function with value vector $v_{f\oplus\sigma_1} \oplus \overline{e}_{\frac{n}{2}}$ has maximum AI.*

Let f be a symmetric functions on $\mathbb{F}_2^n$ which belongs to class 3. If $n = 4k$, then the function with value vector $v_{f\oplus\sigma_1} \oplus c\overline{e}_{n-i}$, where $c = 1$ if i is odd and $c = 0$ otherwise, has maximum AI. If $n = 4k + 2$, then the function with value vector $v_{f\oplus\sigma_1} \oplus \overline{e}_{\frac{n}{2}} \oplus c\overline{e}_{n-i}$, where $c = 1$ if i is odd and $c = 0$ otherwise has maximum AI.

Remark 3. We want to note that the symmetric Boolean functions f derived from the function $F_{\lceil \frac{n}{2} \rceil}$ and also $F_{\lceil \frac{n}{2} \rceil} \oplus \sigma_n$ if n is even have very simple annihilators. For instance, it can be easily seen that the functions $x_{i_1} \cdots x_{i_{\lceil \frac{n}{2} \rceil}}$ with $0 \le i_1 < i_2 < \cdots < i_{\lceil \frac{n}{2} \rceil} \le n-1$ are annihilators of $F_{\lceil \frac{n}{2} \rceil} \oplus \overline{1}$. Moreover, they form exactly the basis of the set of annihilators for $F_{\lceil \frac{n}{2} \rceil} \oplus \overline{1}$. The basis of the annihilators of $F_{\lceil \frac{n}{2} \rceil} \oplus \sigma_n \oplus \overline{1}$ consists of the elements $\{x_0 \cdots x_{\lceil \frac{n}{2} \rceil - 1} \oplus x_{i_1} \cdots x_{i_{\lceil \frac{n}{2} \rceil}} : 0 \le i_1 < i_2 < \cdots < i_{\lceil \frac{n}{2} \rceil} \le n-1, (i_1, \ldots, i_{\lceil \frac{n}{2} \rceil}) \ne (0, \ldots, \lceil \frac{n}{2} \rceil - 1)\}$.

A high number of terms in the equations is another important criteria for the algebraic attacks. Therefore, one should be very careful in choosing the taps of the filter function and the taps of the LFSR when using these symmetric functions in a filter generator. The annihilators of the affine equivalent functions are more complicated. However, this does not change the situation, since one can always replace the filter generator by an equivalent generator with different initial state and connection polynomial of the LFSR and with filter function equal to the affine equivalent one (see [9]).

Annihilators of degree $\frac{n}{2}$ of symmetric functions which belong to classes 2 or 3 are more complicated and consist of more terms.

Properties

Properties such as degree, weight and maximum value in the Walsh spectrum of the functions from classes 1, 2, and 3 for n even are summarized in Table 2. The property of degree can be easily derived by using Proposition 2 and Proposition 4 of [3]. The nonlinearity of the functions is immediately derived from the weight since one can show that $\max_{\overline{w} \in \mathbb{F}_2^n} |W_f(\overline{w})| = |W_f(\overline{0})|$. This is proven in detail by Dalai *et. al* in [7].

Table 2. Properties of Symmetric function on $\mathbb{F}_2^n$ with Maximum AI for n even

Function	Degree	weight	$\max \lvert W_f \rvert$
$F_{\frac{n}{2}}$	$2^{\lfloor \log_2 n \rfloor}$	$2^{n-1} + \frac{1}{2}\binom{n}{\frac{n}{2}}$	$\binom{n}{\frac{n}{2}}$
$F_{\frac{n}{2}} \oplus \overline{s}_{\frac{n}{2}-4}$	$2^{\lfloor \log_2 n \rfloor}$	$2^{n-1} + \frac{1}{2}\binom{n}{\frac{n}{2}}$	$\binom{n}{\frac{n}{2}}$
$F_{\frac{n}{2}} \oplus \overline{e}_{n-i}$	$\ge n - i$	$2^{n-1} + \frac{1}{2}\binom{n}{\frac{n}{2}} - \binom{n}{n-i}$	$\binom{n}{\frac{n}{2}} - 2\binom{n}{n-i}$

The functions from class 1 for n odd are trivially balanced. The nonlinearity of these functions is equal to $2^{n-1} - \binom{n-1}{\frac{n-1}{2}}$. This follows from the fact that the restriction to the subspace $x_n = 0$ (resp. $x_n = 1$) is equal to the symmetric function (resp. complement of symmetric function) of class 1 in $\mathbb{F}_2^{n-1}$. As mentioned in [3], trivially balanced functions satisfy the property that the derivative with respect to the all one vector is constant, *i.e.*, $D_{\overline{1}} f = \overline{1}$. Also $W_f(\overline{v}) = 0$ for all vectors $\overline{v}$ of even weight.

5 Conclusions

We have presented in this paper an efficient algorithm for computing the AI of a symmetric Boolean function. We have identified several classes of symmetric functions with maximum AI.

Since the nonlinearity of functions with maximum AI is not sufficiently high for resisting distinguishing attacks and correlation attacks as explained in [2], we also investigated the existence of symmetric functions with suboptimal AI and better nonlinearity. As shown in the extended version of the paper, it seems that it is not possible to obtain a sufficient order of AI (in the order of 7) together with a reasonable nonlinearity (in the order of $\epsilon = 2^{-9}$) for symmetric functions which depend on less than 32 variables. Therefore in order to use symmetric functions in practise, one should use them as a building block for instance by means of the direct sum with a highly nonlinear Boolean function. Examples of functions with high nonlinearity and which have still reasonable hardware complexity are the Boolean functions which are affine equivalent with the trace function of the power functions.

On the other hand, it is clear that a symmetric function has lots of structure. Therefore, it is an interesting research question whether this structure can be exploited in an attack. Also, the use of the direct sum of two functions has been pointed out as a possible weakness in the design. But again, no attack is known for this. There are two straightforward ways to destroy the symmetry and to still maintain a large set of the properties such as nonlinearity, AI and degree. The first way is by affine transformation on the input variables which keeps the AI, nonlinearity and degree invariant. However, this method is not a good solution, since one can construct an equivalent cipher, with different initial state and different connection polynomial for the LFSR(s) where the function is again symmetric (see [9]). The second way is to add an affine function, which keeps the nonlinearity and degree invariant, but will decrease the AI with 1 in general. For this transformation, it is not immediately clear how to rewrite it to an equivalent scheme where the symmetric function is again obtained.

Acknowledgement

We thank Anne Canteaut and Marion Videau for their useful comments and in particular for providing an algorithm that computes the AI of symmetric

functions. We would also like to thank the anonymous referees for their helpful comments. This work was supported in part by the Concerted Research Action (GOA) Ambiorics 2005/11 of the Flemish Government and by the European Commission through the IST Programme under Contract IST2002507932 ECRYPT. An Braeken is an F.W.O. Research Assistant, sponsored by the Fund for Scientific Research - Flanders (Belgium).

References

1. D.H. Bailey, K. Lee, and H.D. Simon. Using Strassen's algorithm to accelerate the solution of linear systems. *Journal of Supercomputing*, 4:357–371, 1990.
2. A. Braeken and J. Lano. Design principles for LFSR-based stream ciphers. In B. Preneel and S. Tavares, editors, *Selected Areas in Cryptography SAC 2005*, Lecture Notes in Computer Science. Springer-Verlag, 2005.
3. A. Canteaut and M. Videau. Symmetric Boolean functions. *IEEE Transactions on Information Theory*, IT-51(8):2791–2811, 2005.
4. N. Courtois and W. Meier. Algebraic attacks on stream ciphers with linear feedback. In E. Biham, editor, *Eurocrypt 2003*, volume 2656 of *Lecture Notes in Computer Science*, pages 345–359. Springer-Verlag, 2003.
5. N.T. Courtois. Higher order correlation attacks, XL algorithm, and cryptanalysis of Toyocrypt. In P.J. Lee and C.H. Lim, editors, *Information Security and Cryptology ICISC, 2002*, volume 2587 of *Lecture Notes in Computer Science*, pages 182–199. Springer-Verlag, 2002.
6. D.K. Dalai, K.C. Gupta, and S. Maitra. Results on algebraic immunity for cryptographically significant Boolean functions. In A. Canteaut and K. Viswanathan, editors, *Indocrypt 2004*, volume 3348 of *Lecture Notes in Computer Science*, pages 92–106. Springer-Verlag, 2004.
7. D.K. Dalai, S. Maitra, and S. Sarkar. Basic theory in construction of Boolean functions with maximum possible algebraic immunity. Cryptology ePrint Archive, Report 2005/229.
8. E. Dawson and C.-H. Wu. On the linear structures of symmetric Boolean functions. 16:87–102, 1996.
9. C. Ding, G. Xiao, and W. Shan. *Stability Theory of Stream Ciphers*. Springer-Verlag, 1991. ISBN 3-540-54973-0, 0-387-54973-0.
10. W. Meier, E. Pasalic, and C. Carlet. Algebraic attacks and decomposition of Boolean functions. In C. Cachin and J. Camenisch, editors, *Eurocrypt 2004*, volume 3027 of *Lecture Notes in Computer Science*, pages 474–491. Springer-Verlag, 2004.
11. W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. In J.-J. Quisquater and J. Vandewalle, editors, *Eurocrypt 1989*, volume 434 of *Lecture Notes in Computer Science*, pages 549–562. Springer-Verlag, 1989.
12. P. Sarkar and S. Maitra. Balancedness and correlation immunity of symmetric Boolean functions. *Electronic Notes in Discrete Mathematics*, 15:178–183, 2002.
13. M. Videau. On some properties of symmetric Boolean functions. In D.J. Costello and J.B. Hajek, editors, *IEEE International Symposium on Information Theory, 2004, Proceedings*, page 500. IEEE Press, 2004.
14. J. von zur Gathen and J.R. Roche. Polynomials with two values. *Combinatorica*, 17(3):345–362, 1997.
15. I. Wegener. *The Complexity of Boolean Functions*. Wiley, 1987.