Algebraic Immunity Hierarchy of Boolean Functions*

Ziran Tu, Yingpu Deng†

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China

Abstract Algebraic immunity of Boolean functions is a very important concept in recently introduced algebraic attacks of stream cipher. For a n-variable Boolean function f, the algebraic immunity $AI_n(f)$ takes values in $\{0, 1, \ldots, \lceil \frac{n}{2} \rceil\}$. For every k in this range, denote $B_{n,k}$ the set of all n-variable Boolean functions with algebraic immunity k, and we know that $B_{n,k}$ is always non-empty. According to the algebraic immunity, we can form a hierarchy of Boolean functions. Trivially, $|B_{n,0}| = 2$. In general, about this integer sequence $|B_{n,k}|, \quad k = 0, 1, \ldots, \lceil \frac{n}{2} \rceil$, very few results are known. In this paper, we show an explicit formula for $|B_{n,1}|$. That is, we obtain an exact formula for the number of Boolean functions with algebraic immunity one. This is the first exact formula for the terms in the above integer sequence. We also give a tight upper bound for nonlinearity of Boolean functions with algebraic immunity one.

Keywords: Boolean functions; algebraic attack; algebraic immunity; nonlinearity; stream cipher.

1 Introduction

Boolean functions are very important in stream ciphers, of which there are two models: combiner model and filter model. They have been proved to be theoretically equivalent, but the attacks do not works quite similarly on each model. What they have in common is that both the combining function and the filtering function should be balanced, have high algebraic degree, high nonlinearity and high correlation immunity.

Recently, a new clever attack [3] [2] [1] upon stream cipher, the so called algebraic attack, brings a completely new criterion for the design of secure stream cipher systems, known as algebraic immunity.

A Boolean function on n variables is a mapping from $\mathbb{F}_2^n$ into $\mathbb{F}_2$, which is the finite field with two

*The work was supported by NNSF of China (No. 10501049) and 973 Project (No. 2004CB318000).

†Corresponding author. E-mail address: dengyp@amss.ac.cn (Y. Deng)

elements. We denote B_n the set of all n-variable Boolean functions.

Any Boolean function f in B_n has a unique representation as multivariate polynomials over $\mathbb{F}_2$, which is called the algebraic normal form (ANF)

$$f(x_1, x_2, ..., x_n) = \sum_{I \subseteq \{1,...,n\}} a_I \prod_{i \in I} x_i$$

where the a_I's are in $\mathbb{F}_2$. The algebraic degree $\deg(f)$ of f equals the maximum degree of those monomials with nonzero coefficients in its algebraic normal form. A Boolean function f is called affine, if $\deg(f) \leq 1$. The support of f is defined as $\mathrm{Supp}(f) = \{x \in \mathbb{F}_2^n : f(x) = 1\}$, and the $\mathrm{wt}(f)$ is the number of vectors which lies in $\mathrm{Supp}(f)$.

Definition 1.1[6] The algebraic immunity $AI_n(f)$ of an n-variable Boolean function f is defined to be the lowest degree of nonzero functions g such that $fg = 0$ or $(f + 1)g = 0$.

It is known that for arbitrary n-variable Boolean function f, we have $AI_n(f) \leq \lceil \frac{n}{2} \rceil$. Let $B_{n,k} = \{f \in B_n : AI_n(f) = k\}$, where $k = 0, 1, ..., \lceil \frac{n}{2} \rceil$. From [5], we know that $B_{n,k}$ is always non-empty. Thus we have an integer sequence $|B_{n,k}|, \quad k = 0, 1, ..., \lceil \frac{n}{2} \rceil$. Trivially, $|B_{n,0}| = 2$. We are interested in what kinds of Boolean functions in $B_{n,k}$, and their cardinals. If we know this, we can successfully form a hierarchy of Boolean functions according their algebraic immunities, but unfortunately, for a general k, it seems rather difficult to determine completely the number $|B_{n,k}|$, so far as we know, there is little results about this. For example, the references [7] [4] give some lower bound for $|B_{n,\lceil \frac{n}{2} \rceil}|$.

In this paper, we have a try to understand more about this problem, we can give a definite formula to count the number of Boolean functions in $B_{n,1}$, this is the first nontrivial exact formula for the terms in the above integer sequence, and we also give a tight upper bound of nonlinearity for those functions.

2 Main Results

In this section, we give our main results and their proofs. Let us start with a simple fact.

Lemma 2.1 *Let $f \in B_n$ be a non-constant Boolean function, then $AI_n(f) = 1$ if and only if there exists a hyperplane (i.e. $(n-1)$-dimensional subspace of $\mathbb{F}_2^n$) H in $\mathbb{F}_2^n$ such that Supp(f)$\subseteq H$ or Supp(f)$\supseteq H$ or Supp(f)$\subseteq \overline{H}$ or Supp(f)$\supseteq \overline{H}$, where $\overline{H} = \mathbb{F}_2^n \backslash H$.*

Proof. $AI_n(f) = 1$ means there exists a degree-1 function g such that $f \cdot g = 0$ or $(f + 1) \cdot g = 0$, the support of g is a hyperplane or its complement, then it's easy to derive the lemma. □

Lemma 2.2 *We choose m distinct vectors from $\mathbb{F}_2^n \backslash \{(0,0,..,0)\}$ to form a matrix over $\mathbb{F}_2$ with rank r, denote the total number of this kind of matrices by $f_n(m,r)$, then*

$$f_n(m,r) = \begin{cases} 0 & \text{if } r > m \\ f_n(m-1,r)\cdot(2^r - m) + f_n(m-1,r-1)\cdot(2^n - 2^{r-1}) & \text{otherwise} \end{cases}$$

Proof. Suppose we've already a matrix composed by $m-1$ distinct non-zero vectors $\alpha_1, \alpha_2, ...,$ α_{m-1} in $\mathbb{F}_2^n$, we need to choose α_m such that $\text{rank}\{\alpha_1, \alpha_2, ..., \alpha_m\} = r$, there are two cases to be considered: first, if $\text{rank}\{\alpha_1, \alpha_2, ..., \alpha_{m-1}\} = r$, then we should choose α_m in the subspace spanned by $\alpha_1, \alpha_2, ..., \alpha_{m-1}$, there are $2^r - m$ choices for α_m; second, if $\text{rank}\{\alpha_1, \alpha_2, ..., \alpha_{m-1}\} = r-1$, we should choose α_m not in the subspace spanned by $\alpha_1, \alpha_2, ..., \alpha_{m-1}$, there are $2^n - 2^{r-1}$ possibilities, then we obtain our recursive relation. □

When $m = r$, from [8] we know

$$f_n(r,r) = (2^n - 1)\cdot(2^n - 2)\cdot ... \cdot (2^n - 2^{r-1})$$

and by Lemma 2.2 we can obtain iteratively all $f_n(m,r)$.

Lemma 2.3 *We denote $F_n(m,r)$ the number of possibilities to choose m distinct non-zero vectors from F_2^n whose rank is r, then $F_n(m,r) = f_n(m,r)/m!$.*

Proof. It is obvious. □

Now, we can deduce our definite formula to count the number of n-variable Boolean functions with algebraic immunity one, this is the following theorem.

Theorem 2.4 *We have $|B_{n,1}| = 2 - 2^{n+1} + \sum_{m=1}^{2^n-1}\sum_{r=1}^{n} F_n(m,r)\cdot 2^{r+1}\cdot(2^{2^{n-r}} - 1)\cdot(-1)^{m+1}$.*

Proof. By Lemma 2.1, we only need to consider the following set

$A = \{X \subseteq \mathbb{F}_2^n : X \neq \emptyset$ and $X \neq \mathbb{F}_2^n$, there exists a hyperplane H such that $X \subseteq H$ or $X \subseteq \overline{H}$ or $X \supseteq H$ or $X \supseteq \overline{H}\}$, and $|A|$ is what we want, that is, $|A| = |B_{n,1}|$.

Let us give an order on all $2^n - 1$ nonzero vectors in $\mathbb{F}_2^n$, let α_i be the ith vector and H_i be the hyperplane which is $\{x \in \mathbb{F}_2^n :< x, \alpha_i >= 0\}$, where $< x, \alpha_i >$ denotes the inner-product of x and α_i, $i = 1, ..., 2^n - 1$.

We denote $A_i = \{X \subseteq \mathbb{F}_2^n : X \neq \emptyset$, $X \neq \mathbb{F}_2^n$, $X \neq H_i$, $X \neq \overline{H_i}$ and $X \supseteq H_i$ or $X \subseteq H_i$ or $X \supseteq \overline{H_i}$ or $X \subseteq \overline{H_i}\}$, we have $|A| = |\bigcup_{i=1}^{2^n-1} A_i| + 2^{n+1} - 2$, in which $2^{n+1} - 2$ is the number of non-constant

affine functions. By the Inclusion and Exclusion-Principle, then

$$|\bigcup_{i=1}^{2^n-1} A_i| = \sum_i |A_i| - \sum_{i,j} |A_i \bigcap A_j| + ... + (-1)^{m+1} \sum_{i_1,i_2,...,i_m} |\bigcap_{j=1}^{m} A_{i_j}| + ... + |\bigcap_{i=1}^{2^n-1} A_i|.$$

We need to compute $|\bigcap_{j=1}^{m} A_{i_j}|$. If $m = 1$, it is easy to compute that $|A_i| = 2^2 \cdot (2^{2^{n-1}} - 2)$. Suppose $m > 1$, we can divide $\bigcap_{j=1}^{m} A_{i_j}$ into two parts

$$\bigcap_{j=1}^{m} A_{i_j} = \bigcup_{S_{i_j}=H_{i_j} or \overline{H_{i_j}}} \{X \subseteq \mathbb{F}_2^n : X \neq \emptyset, X \subseteq \bigcap_{j=1}^{m} S_{i_j}\} \bigcup \bigcup_{S_{i_j}=H_{i_j} or \overline{H_{i_j}}} \{X \subseteq \mathbb{F}_2^n : X \neq \mathbb{F}_2^n, X \supseteq \bigcup_{j=1}^{m} S_{i_j}\}.$$

Since $\{X \subseteq \mathbb{F}_2^n : X \neq \emptyset, X \subseteq \bigcap_{j=1}^{m} S_{i_j}\}$ and $\{X \subseteq \mathbb{F}_2^n : X \neq \mathbb{F}_2^n, X \supseteq \bigcup_{j=1}^{m} S_{i_j}\}$ are symmetric, these two parts have the same cardinal, we can only consider the first part. If $rank\{\alpha_{i_1}, \alpha_{i_2}, ..., \alpha_{i_m}\} = r$, then $\bigcap_{j=1}^{m} H_{i_j}$ is a $(n-r)$-dimensional subspace, and then $\bigcap_{j=1}^{m} S_{i_j}$ is either $\emptyset$ or a $(n-r)$-dimensional flat, note that the components of the first part are disjoint, in other words, there are 2^r disjoint flats with dimension $(n-r)$, we get

$$|\bigcup_{S_{i_j}=H_{i_j} or \overline{H_{i_j}}} \{X \subseteq \mathbb{F}_2^n : X \neq \emptyset, X \subseteq \bigcap_{j=1}^{m} S_{i_j}\}| = 2^r \cdot (2^{2^{n-r}} - 1).$$

then

$$|\bigcap_{j=1}^{m} A_{i_j}| = 2^{r+1} \cdot (2^{2^{n-r}} - 1).$$

When we choose randomly m non-zero vectors from $\mathbb{F}_2^n$, its rank may distribute from 1 to $\text{Min}\{n, m\}$, by lemma 2.3, there are $F_n(m, r)$ possibilities that the rank of this group of vectors is r, we have

$$\sum_{i_1,i_2,...,i_m} |\bigcap_{j=1}^{m} A_{i_j}| = \sum_{r=1}^{n} F_n(m, r) \cdot 2^{r+1} \cdot (2^{2^{n-r}} - 1).$$

Finally, don't forget that we should take into account the $2^{n+1} - 2$ non-constant affine functions, we can get

$$\begin{aligned} |A| &= 2^{n+1} - 2 + \sum_{m=2}^{2^n-1} \sum_{r=1}^{n} F_n(m, r) \cdot 2^{r+1} \cdot (2^{2^{n-r}} - 1) \cdot (-1)^{m+1} + F_n(1, 1) \cdot 2^2 \cdot (2^{2^{n-1}} - 2) \\ &= 2 - 2^{n+1} + \sum_{m=1}^{2^n-1} \sum_{r=1}^{n} F_n(m, r) \cdot 2^{r+1} \cdot (2^{2^{n-r}} - 1) \cdot (-1)^{m+1}. \end{aligned}$$

This proves our theorem. □

Remark From our formula, we have the following table

n	$\lvert B_{n,1}\rvert$	$\lvert B_{n,1}\rvert/\lvert B_n\rvert$
1	2	0.5
2	14	0.875
3	198	0.7734375
4	10582	0.161468505859
5	7666550	0.00178500777110457420349121093750
6	1081682871734	0.00000005863814597371810183323859178O
7	9370945806264076577334	2.75387346428130707474160629154766355497062e-17

We can see from the above table, that $B_{n,1}$ constitutes only a very small part of B_n, and as n grows up, the proportion of $B_{n,1}$ in B_n approaches 0.

It is well known that for any $\alpha \in \mathbb{F}_2^n$, the value

$$\mathrm{W}_f(\alpha) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+<x,\alpha>}$$

is called the Walsh coefficient of f at α. The nonlinearity of Boolean function f can be expressed via its Walsh coefficients by the next formula

$$\mathrm{nl}(f) = 2^{n-1} - \frac{1}{2}\mathrm{Max}_{u \in \mathbb{F}_2^n} |\mathrm{W}_f(u)|.$$

We also derive a tight upper bound on the nonlinearity of Boolean functions with algebraic immunity one.

Theorem 2.5 *Let f be in B_n with $AI_n(f) = 1$, then $nl(f) \leq 2^{n-2}$, and this bound is tight.*

Proof. Suppose f and g in B_n, it's easy to verify that

$$2 \cdot (-1)^{f \cdot g} = 1 + (-1)^f + (-1)^g - (-1)^{f+g}.$$

By the definition of Walsh coefficient, we have

$$2 \cdot \mathrm{W}_{f \cdot g}(\alpha) = \mathrm{W}_0(\alpha) + \mathrm{W}_f(\alpha) + \mathrm{W}_g(\alpha) - \mathrm{W}_{f+g}(\alpha),$$

if $f \cdot g = 0$, then

$$2^n \delta_{\alpha,0} + \mathrm{W}_{f+g}(\alpha) = \mathrm{W}_f(\alpha) + \mathrm{W}_g(\alpha).$$

Since $AI_n(f) = 1$, we assume $g(x) =< \beta, x > +a_0$, in which β is nonzero in $\mathbb{F}_2^n$ and a_0 in $\mathbb{F}_2$. Let $\alpha = 0$, we get

$$2^n + (-1)^{a_0} \mathrm{W}_f(\beta) = \mathrm{W}_f(0).$$

Then

$$2^n \leq |\mathrm{W}_f(0)| + |\mathrm{W}_f(\beta)| \leq 2 \cdot \mathrm{Max}_{u \in \mathbb{F}_2^n} |\mathrm{W}_f(u)|.$$

Finally

$$\mathrm{nl}(f) = 2^{n-1} - \frac{1}{2}\mathrm{Max}_{u \in \mathbb{F}_2^n} |\mathrm{W}_f(u)| \leq 2^{n-1} - 2^{n-2} = 2^{n-2}.$$

Note that the upper bound we obtained above is also tight. For $n = 1$, the above bound gives $\mathrm{nl}(f) \leq \frac{1}{2}$, that is, $\mathrm{nl}(f) = 0$. Suppose $n \geq 2$. Consider $f(x_1, x_2, \ldots, x_n) = x_1 x_2$ in B_n, clearly $AI_n(f) = 1$, because $x_1 x_2 (x_1 + x_2) = 0$. The Walsh coefficient of f at $(a_1, a_2, ..., a_n) \in \mathbb{F}_2^n$ is

$$\mathrm{W}_f(a_1, a_2, ..., a_n) = \sum_{x \in \mathbb{F}_2^n} (-1)^{x_1 x_2 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n} = \sum_{x_1, x_2 \in \mathbb{F}_2} (-1)^{x_1 x_2 + a_1 x_1 + a_2 x_2} \prod_{i=3}^{n} \sum_{x_i \in \mathbb{F}_2} (-1)^{a_i x_i}.$$

If $(a_3, a_4, ..., a_n) \neq (0, 0, ..., 0)$, then $W_f(a_1, a_2, ..., a_n) = 0$. So $W_f(0, 0, 0, ..., 0) = 2^{n-1}$, $W_f(0, 1, 0, ..., 0) = W_f(1, 0, 0, ..., 0) = 2^{n-1}$ and $W_f(1, 1, 0, ..., 0) = -2^{n-1}$, we get $\mathrm{nl}(f) = 2^{n-2}$. □

3 Conclusion

According to the algebraic immunity, we can form a hierarchy of Boolean functions. It is very difficult to determine the number of Boolean functions with a specified algebraic immunity. In this paper, we obtain the first complete answer to this problem, that is, we give the exact formula for the number of any n-variable Boolean functions with algebraic immunity one, and we also give a tight upper bound of nonlinearity for those functions.

References

[1] F. Armknecht, Improving fast algebraic attacks, FSE 2004, Springer LNCS vol. 3017, pp. 65–82.

[2] N. T. Courtois, Fast algebraic attacks on stream ciphers with linear feedback, Crypto 2003, Springer LNCS vol. 2729, pp. 176–194.

[3] N. T. Courtois, W. Meier, Algebraic attacks on stream ciphers with linear feedback, Eurocrypt 2003, Springer LNCS vol. 2656, pp. 345–359.

[4] N. Li, W. F. Qi, Construction and count of Boolean functions of an odd number of variables with maximum algebraic immunity. preprint.

[5] M. Lobanov. Tight bound between nonlinearity and algebraic immunity. Cryptology ePrint Archive 2005/441.

[6] W. Meier, E. Psalic, C. Carlet, Algebraic attacks and decomposition of Boolean functions, Eurocrypt 2004, Springer LNCS vol. 3027 pp. 474–491.

[7] L. Qu, G. Feng, C. Li, On the Boolean functions with maximum possible algebraic immunity: construction and a lower bound of the count. preprint.

[8] Z. Wan, Geometry of Classical Groups over Finite Fields, Second Edition, Science Press, Beijing, 2002.