Improved lower bound on the number of balanced symmetric functions over $GF(p)$

Pinhui Ke
Fujian Normal University
Key Laboratory of Network Security
and Cryptology
Fujian 350007, P. R. China
keph@eyou.com

Abstract

The lower bound on the number of n-variable balanced symmetric functions over finite fields GF(p) presented in [1] is improved in this paper.

1. Introduction

Symmetric Boolean functions is an interesting subclass of Boolean functions whose output depend only on the weight of the input vector. These functions can be represented in a very compact way both for their algebraic normal forms and for their value vectors. As symmetric functions are the only functions having a known implementation with a number of gates which is linear in the number of input variables, they might be good candidates in term of implementation complexity[2].

In binary case, that is $p = 2$, a lot of work have been done. Br$\ddot{u}$er [3], Mitchell [4]and later Y.X.Yang and B.Guo [5] studied the balanced symmetric functions and correlation immune symmetric functions. S.Maitra and P.Sarkar [6]studied the maximum nonlinearity of symmetric Boolean function on odd number of variables. A.Canteaut and M. Videau [2] established the link between the periodicity of simplified value vector of an symmetric Boolean functions and its degree. Especially, Algebraic immunity is a recently proposed cryptographic criteria which is used to evaluated the ability of an Boolean functions to resist algebraic attack[7]. Symmetric Boolean function had been proved to have good algebraic immunity[8, 9].

Boolean function is natural to be generalized to other finite fields of odd prime characteristic p. For example, Y.Hu and G.Xiao [10] studied the resilient functions on $GF(p)$. In [11], Li and Cusick introduced the strict avalanche criterion over $GF(p)$. In [12], they determined all the linear structures of symmetric functions over $GF(p)$. Recently, they give a lower bound for the number of balanced symmetric functions over $GF(p)$ and show the existence of nonlinear balanced symmetric functions[1].

The correspondence is organized as follows: Section 2 includes the basic background and notations. Section 3 settles some new notations and describes the result presented in [1] firstly. Based on Cusick etc's method, new classes of balanced symmetric functions over $GF(p)$ are then constructed. Also the lower bound in [1] is improved. In the last section, an equivalent problem is described.

2 Preliminaries

Let p be a prime number and $GF(p)^n$ as the set of all n-tuples of elements in the finite fields $GF(p)$. If $f(x)$: $GF(p)^n \rightarrow GF(p)$, then f can be uniquely represented as

$$f(x_1, x_2, \cdots, x_n) = \sum_{k_1,k_2,\cdots,k_n=0}^{p-1} a_{k_1,k_2,\cdots,k_n} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}$$

where each coefficient $a_{k_1,k_2,\cdots,k_n}$ is a constant in $GF(p)$. It is also called the *algebraic normal form* (ANF) of f.

Denote by F_n the set of all functions of n variables. Let S_n be the *symmetric group* on n element, that is, the collection of all bijections on$\{1, 2, \cdots, n\}$. For $f \in F_n$, f is called *symmetric* if for any permutation $\pi \in S_n$, we have $f(x_{\pi(1)}, x_{\pi(2)}, \cdots, x_{\pi(n)}) = f(x_1, x_2, \cdots, x_n)$. For any $X = (x_1, x_2, \cdots, x_n) \in GF(p)^n$, it is convenient to denote $\pi(X) = (x_{\pi(1)}, x_{\pi(2)}, \cdots, x_{\pi(n)})$ by abuse of notation.

Define an equivalent relation on $GF(p)^n$: for any $X, Y \in GF(p)^n$, write $X \sim Y$ if and only if there exists a permutation $\pi \in S_n$ such that $Y = \pi(X)$.

For $f \in F_n$, f is called *balanced* if the probability $prob(f(X) = k) = \frac{1}{p}$ for any $k = 0, 1, \cdots, p-1$. It is easy to see that f is balanced if and only if $|\{X \in GF(p)^n | f(X) = k\}| = p^{n-1}$ for any $k \in GF(p)$.

3 New classes of balanced symmetric functions

By the definition of symmetric function, we get that f is symmetric if and only if f take the same value for any n-tuple in the same equivalent class. So in order to get the number of symmetric functions, it is necessary to compute the number of different equivalent class. This number is exactly the solutions of the linear diophantine equation $i_0 + i_1 + \cdots + i_{p-1} = n$. From the viewpoint of combinatorial enumeration[13], the number of the solution of above equation can also be interpreted as the number of different ways to put n nondistinctive balls into p distinct boxes.

Lemma 3.1 *[1] The number of n-variable symmetric functions over $GF(p)$ is $p^{C(n+p-1,n)}$.*

Here $C(n,k) = \frac{n!}{k!(n-k)!}$ is the *binomial coefficient*.

For each equivalent class $\widetilde{X}$, we may choose those elements $X = (x_1, x_2, \cdots, x_n), x_1 \le x_2 \le \cdots \le x_n$ as representative elements, denote it as $\overline{X}$. Let $\overline{X} = (\underbrace{0,...,0}_{x_{i_0}}\underbrace{1,...,1}_{x_{i_1}},\cdots,\underbrace{p-1,...,p-1}_{x_{i_{p-1}}})$, where $i_0+i_1+\cdots+i_{p-1} = n, 0 \le i_j \le n, j = 0,1,\cdots,p-1$. Then the cardinality of the set $\widetilde{X}$ equals the multinomial coefficient $C(n, i_0, i_1, \cdots, i_{p-2}) = \frac{n!}{i_0!i_1!\cdots i_{p-1}!}$.

If $X = (..., c, \underbrace{d,...,d}_{k}, e, ...), c \ne d$ and $d \ne e$, we call that $\underbrace{d,...,d}_{k}$ a *run* of X and the length of the run is k. For a fixed n-tuple X, we can write out all the run length, called the *run distribution* . For example, let $X = (0,0,0,1,2)$ has runs "000","1" and "2" and the lengths are 3,1 and 1 respectively. Then the run distribution is 113.

For a chosen n and p, one can list all the possible run distribution of different representative elements. Using integral partition (more concretely, dividing n into at most p parts), we can easily get the number of different run distribution. For example, for $n = 4, p = 3$, there are exactly 4 different run distribution $13, 112, 22$ and 4. For each run distribution, it always contains several representative elements (or,equivalently, several equivalent classes). For example, representative elements $0111, 0222, 1222$ have the same run distribution 13. Furthermore, let m_i denotes the number of the runs of length i, then the number of different equivalent class having the same run distribution is $\frac{p!}{m_0!m_1!\cdots m_n!}$. Cusick etc.[1] observed that under mild condition $\frac{p!}{m_0!m_1!\cdots m_n!}$ is a multiple of p.

Lemma 3.2 *[1] Let n, p be positive integers, with p a prime number. If $m_i \le p$ for some i (and so for all i), or if $gcd(n,p) = 1$, then p divides $\frac{p!}{m_0!m_1!\cdots m_n!}$.*

To get balanced symmetric polynomials, we need to participate the $C(n+p-1, n)$ equivalent classes into p groups such that each group consisting of p^{n-1} elements. By Lemma 3.2, Cusick etc.[1] divide each class having the same run distribution into p groups. So they constructed a class of balanced symmetric functions. By enumerating the functions, they presented the following lower bound on the number of balanced symmetric functions over $GF(p)$.

Theorem 3.1 *[1] Let N be the number of n-variable balanced symmetric functions over GF(p). If $m_i \le p$ for some i (or $gcd(n,p) = 1$), then*

$$N \ge \prod_{\substack{\sum_{j=0}^{n} m_j = p, \\ \sum_{j=0}^{n} j m_j = n}} \frac{(\frac{p!}{m_0!\cdots m_n!})!}{((\frac{(p-1)!}{m_0!\cdots m_n!})!)^p} \tag{1}$$

For fixed n and p such that $gcd(n,p) = 1$, assume the number of different run distributions is t (just as we have pointed out, this number can be easily obtained by integral partition, which can be calculated by generating function[13]). let $\triangle_i$ be the collection of equivalent classes with the same run distribution, $1 \le i \le t$. Denotes the cardinality of each collection $\triangle_i$ as k_i. That is, $k_i = |\triangle_i| = \frac{p!}{m_0!m_1!\cdots m_n!}$, here m_i is the number of run of length i. If $i_0 i_1 \cdots i_{p-1}$ be a run distribution, then define $h_i = \frac{n!}{i_0!i_1!\cdots i_{p-1}!}$. By the definition, each equivalent class belong the collection $\triangle_i$ contains the same number of elements h_i. It is easy to verified that

$$\sum_{i=1}^{t} k_i h_i = p^n. \tag{2}$$

Take $n = 5, p = 3$ as example, we concluded these numbers in the following table:

$\triangle_i$'s run distribution	k_i	h_i
5	3	1
1 4	6	5
2 3	6	10
1 1 3	3	20
1 2 2	3	30

Table 1: When $n = 5, p = 3$,
$\triangle_i$: collection of equivalent classes with certain run distribution
$k_i = |\triangle_i|$
h_i : the number of elements in each equivalent class

Using the above notation, the lower bound in Theorem 3.1 can be written as

$$\begin{aligned}&\prod \sum_{\substack{\sum_{j=0}^{n} m_j=p,\\ \sum_{j=0}^{n} jm_j=n}} \frac{(\frac{p!}{m_0!\cdots mn!})!}{((\frac{(p-1)!}{m_0!\cdots mn!})!)^p}\\ &= \prod_{i=1}^{t} C(k_i,\tfrac{k_i}{p})C(k_i-\tfrac{k_i}{p},\tfrac{k_i}{p})\cdots C(\tfrac{k_i}{p},\tfrac{k_i}{p})\\ &= \prod_{i=1}^{t} \frac{(k_i)!}{(\frac{k_i}{p})!^p}\end{aligned} \tag{3}$$

In Cusick's enumeration, each collection of equivalent classes with the same run distribution is divided evenly into p group. When $n=5, p=3$, we demonstrate a partition as follows:

	0	1	2
$\triangle_1$	(1)	(1)	(1)
$\triangle_2$	(5)(5)	(5)(5)	(5)(5)
$\triangle_3$	(10)(10)	(10)(10)	(10)(10)
$\triangle_4$	(20)	(20)	(20)
$\triangle_5$	(30)	(30)	(30)

Here each $(\cdot)$ denotes an equivalent class of $\triangle_i$ and the number in $(\cdot)$ is h_i, the cardinality of each equivalent class. However many potential balanced symmetric functions may be left out in this way . For example, by $30 = 20 + 10$, we can divide $\triangle_i$ as follows:

	0	1	2
$\triangle_1$	(1)	(1)	(1)
$\triangle_2$	(5)(5)	(5)(5)	(5)(5)
$\triangle_3$	(10)(10)(10)	(10)	(10)(10)
$\triangle_4$	(20)(20)		(20)
$\triangle_5$		(30)(30)	(30)

That is, the equivalent classes in each collection may be divided unevenly. Noted that the modified function is also a balanced symmetric function. We will use this idea to look for more symmetric balanced functions and thus improve the lower bound greatly.

For fixed n, p, let h_i be defined as above. Without loss of generality, we may assume that $h_1 \le h_2 \le \cdots \le h_t$. consider the following muli-variable equation with restricted conditions:

$$\sum_{i=1}^{t} x_i h_i = 0, x_i \in Z, |x_i| \le \frac{k_i}{p} \tag{4}$$

Obviously, $X = (0, 0, \cdots, 0)$ is a trivial solution. Two solutions $X = (x_1, x_2, \cdots, x_t)$ and $Y = (y_1, y_2, \cdots, y_t)$ are said to be equivalent if $X = \pm Y$. In this case, the solutions whose most right nonzero component is positive are called *uniformed*. For equivalent solutions, we choose the uniformed solution and discard the other one. Denote the set of nontrivial solutions of equation (4) as $S_{n,p}$ (for equivalent solutions, only uniformed solutions are chosen).

Example 1: Let $n = 5, p = 3$. By table 1, we have equation:

$x_1 \cdot 1 + x_2 \cdot 5 + x_3 \cdot 10 + x_4 \cdot 20 + x_5 \cdot 30 = 0$, such that $x_1 \in \{-1, 0, 1\}, x_2 \in \{-2, -1, 0, 1, 2\}, x_3 \in \{-2, -1, 0, 1, 2\}, x_4 \in \{-1, 0, 1\}, x_5 \in \{-1, 0, 1\}$.

Then $S_{5,3} = \{(0, -2, 1, 0, 0), (0, -2, -1, 1, 0), (0, 0, -2, 1, 0), (0, -2, 0, -1, 1), (0, 0, -1, -1, 1), (0, -2, -2, 0, 1)\}$.

Lemma 3.3 *Let n, m and t be positive integers, $0 \le t \le n$, then*

$$C(mn, n-t)C(mn-n+t, n+t)C((m-2)n, n)\cdots C(n, n) = \frac{(mn)!}{(n-t)!(n+t)!(n!)^{m-2}}$$

Proof: It is easily verified that

$$\begin{aligned}&C(mn, n-t)C(mn-n+t, n+t)C((m-2)n, n)\cdots C(n, n)\\ &= \frac{(mn)!}{((m-1)n+t)!(n-t)!}\frac{((m-1)n+t)!}{((m-2)n)!(n+t)!}\\ &\cdot\frac{((m-2)n)!}{((m-3)n)!n!}\cdots\frac{n!}{n!}\\ &= \frac{(mn)!}{(n-t)!(n+t)!(n!)^{m-2}}.\end{aligned}$$

Let $X = (x_1, x_2, \cdots, x_t) \in S_{n,p}$. We now construct some new classes of balanced symmetric functions. For those zero components x_i in X, the corresponding equivalent classes $\triangle_i$ must be divided evenly. The number of partitions is

$$C(k_i, \frac{k_i}{p})C(k_i - \frac{k_i}{p}, \frac{k_i}{p})\cdots C(\frac{k_i}{p}, \frac{k_i}{p}) = \frac{(k_i)!}{(\frac{k_i}{p})!^p}.$$

For those nonzero components in X, choose $\frac{k_i}{p} - x_i$ equivalent class from $\triangle_i$ firstly, $\frac{k_i}{p} + x_i$ secondly, and the rest are divided evenly. The number of partitions is

$$p\cdot(p-1)\cdot C(k_i, \frac{k_i}{p} - x_i)C(k_i - \frac{k_i}{p} + x_i, \frac{k_i}{p} + x_i)$$

$$C(k_i - 2\frac{k_i}{p}, \frac{k_i}{p})\cdots C(\frac{k_i}{p}, \frac{k_i}{p})$$

where the first term $p\cdot(p-1) = \frac{p!}{1!1!(p-2)!}$ is to take into account the different orderings of the p groups. By Lemma 3.3, this product can be written as

$$p\cdot(p-1)\cdot\frac{k_i!}{(\frac{k_i}{p} - x_i)!(\frac{k_i}{p} + x_i)!(\frac{k_i}{p}!)^{p-2}}$$

In order to get balanced symmetric functions, we require that for those nonzero components in X, once an order is specified for a group which is a partition of an equivalent class collection $\triangle_i$, the other groups corresponding to nonzero components in X also take the same order.

Now we give our main result.

Theorem 3.2 *Let n, p be two co-prime integers, t be the number of different run distribution, k_i be the cardinality of equivalent classes with the same run distribution, h_i be number of the elements contained in equivalent class of each collection, $1 \leq i \leq t$, $S_{n,p}$ be the set of the uniformed nontrivial solution of equation*

$$\sum_{i=1}^{t} x_i h_i = 0, x_i \in \mathrm{Z}, |x_i| \leq \frac{k_i}{p}, \tag{5}$$

Then the number of n-variable balanced symmetric functions over GF(p) has the lower bound

$$\prod_{i=1}^{t} \frac{(k_i)!}{(\frac{k_i}{p})!^p} + \sum_{X=(x_1,x_2,\cdots,x_t)\in S_{n,p}} p(p-1)$$

$$(\prod_{\substack{x_i=0,\\ 1\leq i\leq t}} \frac{(k_i)!}{(\frac{k_i}{p})!^p} \prod_{\substack{x_i\neq 0,\\ 1\leq i\leq t}} \frac{k_i!}{(\frac{k_i}{p}-x_i)!(\frac{k_i}{p}+x_i)!(\frac{k_i}{p}!)^{p-2}}). \tag{6}$$

Proof. It is obvious that all the functions constructed are symmetric. Just as proved in [1], the functions constructed by Cusick etc is balanced, which also corresponds to the case of trivial solution of equation (5). Now we prove the functions constructed from the nontrivial solutions of (5) are also balanced and they are different from Cusick etc's construction.

In order to prove that a function is balanced, we only need to prove that the cardinality of the pre-image of each function value is p^{n-1}. Let $X \in S_{n,p}$. Then each $\triangle_i, 1 \leq i \leq t$, are divided into p groups. According to the value of x_i, the corresponding $\triangle_i$ are divided in different ways. So the pre-image of the each value $\{0, 1, \cdots, p-1\}$ must belong to one of the following cases:

1. It is consisted of $\frac{k_i}{p}$ equivalent classes of $\triangle_i$, for all $1 \leq i \leq t$.
2. It is consisted of $\frac{k_i}{p} + x_i$ equivalent classes of $\triangle_i$ for $x_i \neq 0$, and $\frac{k_i}{p}$ equivalent classes of $\triangle_i$ for $x_i = 0$.
3. It is consisted of $\frac{k_i}{p} - x_i$ equivalent classes of $\triangle_i$ for $x_i \neq 0$, and $\frac{k_i}{p}$ equivalent classes of $\triangle_i$ for $x_i = 0$.

It is straightforward to calculate the number of elements in each cases. In case 1, by (2) the number of elements is

$$\sum_{i=1}^{t} \frac{k_i}{p} h_i = \frac{1}{p} \sum_{i=1}^{t} k_i h_i = p^{n-1}.$$

In case 2, the number of elements is

$$\sum_{\substack{x_i=0,\\ 1\leq i\leq t}} \frac{k_i}{p} h_i + \sum_{\substack{x_i\neq 0,\\ 1\leq i\leq t}} (\frac{k_i}{p} + x_i) h_i$$

$$= \frac{1}{p} \sum_{i=1}^{t} k_i h_i + \sum_{\substack{x_i\neq 0,\\ 1\leq i\leq t}} x_i h_i$$

For $X \in S_{n,p}$, we have

$$\sum_{\substack{x_i\neq 0,\\ 1\leq i\leq t}} x_i h_i = 0.$$

So the number of elements in case 2 is also p^{n-1}. Case 3 can be proved similarly. In conclusion, each function we constructed is balanced.

Because X is a nontrivial solution of equation (4), at lease two components of X are nonzero and then two corresponding collections $\triangle_i$ will be partitioned unevenly. So the construction we presented is different from Cusick etc's construction. On the other hand, by the definition of $S_{n,p}$, each $X \in S_{n,p}$ is uniformed solution of equation (4), so the functions we constructed are different from each other. Furthermore, as we have described in our construction, for those $x_i \neq 0$, given an order of the p groups of an $\triangle_i$, the other groups which are consisted of $\frac{k_i}{p} + x_i$ (and $\frac{k_i}{p} - x_i$ respectively) equivalent classes must lie in the same position. So these functions are enumerated as follows:

$$\sum_{X=(x_1,x_2,\cdots,x_t)\in S_{n,p}} p(p-1)$$

$$(\prod_{\substack{x_i=0,\\ 1\leq i\leq t}} \frac{(k_i)!}{(\frac{k_i}{p})!^p} \prod_{\substack{x_i\neq 0,\\ 1\leq i\leq t}} \frac{k_i!}{(\frac{k_i}{p}-x_i)!(\frac{k_i}{p}+x_i)!(\frac{k_i}{p}!)^{p-2}}),$$

Plusing the number of functions constructed by Cusick etc, new lower bound (6) is then obtained. Thus the proof is completed.

Except special cases, equation (5) always has nontrivial solutions. So Theorem 3.2 improves the lower bound in Theorem 3.1. To illust our result, take $n = 5, p = 3$ as example. By Theorem 3.1, Cusick etc's lower bound is 1749600. And by Example 1, equation (5) has 6 nontrivial solutions. Omitting the detail of the calculations, the number of the functions constructed by our method is 32659200. So our result improves Cusick etc's lower bound greatly.

4 An equivalent characterization

Let p be a prime number and n be an arbitrary positive integer. In this section, we prove that the enumeration of the number of balanced symmetric n-variable functions over $GF(p)$ is equivalent to solve an equation system.

Let $t, \triangle_i, k_i$ and $h_i, 1 \leq i \leq t$, be defined as in section 3. Because n and p are not required to be co-prime, k_i is not necessarily a multiple of p. In order to get balanced symmetric functions, $\triangle_i, 1 \leq i \leq t$, must be participated into p parts properly. In detail, let $x_{ij}, 1 \leq i \leq t, 1 \leq j \leq p$, be a partition of $\triangle_i$. Then $x_{ij}, 1 \leq i \leq t, 1 \leq j \leq p$, must satisfy that

$$\begin{cases} \sum_{i=1}^{t} x_{ij} h_i = p^{n-1}, 1 \leq j \leq p \\ \sum_{j=1}^{p} x_{ij} = k_i, 1 \leq i \leq t. \end{cases} \tag{7}$$

And there are always several functions correspond to each solution of (7). Because the structure of the solutions are not clear, it is different for us to enumerate exactly.

Contrarily, if a balanced symmetric function exits, there is a set of positive integers $x_{ij}, 1 \leq i \leq t, 1 \leq j \leq p$, satisfying equation system (7). Thus, we get the following result.

Theorem 4.1 *Let notations be defined as before. Then the number of n-variable balanced symmetric functions over GF(p) is not less than the number of solutions of equation systems over $\mathcal{Z}^+$:*

$$\begin{cases} \sum_{i=1}^{t} x_{ij} h_i = p^{n-1}, 1 \leq j \leq p \\ \sum_{j=1}^{p} x_{ij} = k_i, 1 \leq i \leq t. \end{cases}$$

here $\mathcal{Z}^+$ denotes the set of positive integers.

The equation system (7) can also be regarded as an strengthened version of *Knapsack problem*, which is a so-called NP-complete problems. Hence, it seems hard to give an exact number of balanced symmetric functions over $GF(p)$.

5 Conclusion

Based on the Cusick etc's construction, new classes of balanced symmetric functions over $GF(p)$ are constructed and the lower bound in [1] is improved in this paper. For general case, an equivalent characterization is also presented.

References

[1] T.W.Cusick, Y. Li and P.Stănică. Balanced symmetric functions over GF(p). *IEEE Trans. on Infor. theory*,54(3),pp.1304-1307, 2008.

[2] A.Canteaut, M. Videau. Symmetric Boolean functions. *IEEE Trans. on Infor. theory*,51(8),pp.2791-2811, 2005.

[3] J.O.Brüer. On pseudorandom sequences as crypto generators. In *International Zurich Seminar on Digital Communications*, pp.157-161, IEEE, New York, 1984.

[4] C.J.Mitchell. Emunerating Boolean functions of cryptographic significance. *J.Cryptology*, 2(3), pp.155-170, 1990.

[5] Y.X.Yang, B.Guo. Further enumerating Boolean fucntions of cryptographic signifiance. *J.Cryptology*, 8(3), pp.115-122, 1995.

[6] S.Maitra, P.Sarkar. Maximum nonlinearity of symmetric Boolean functions on odd number of variables, *IEEE Trans. on Infor. theory*, 48(9), pp: 2626-2630 , 2002.

[7] S.Meier, E.Pasalic and C.Carlet. Alegbraic attacks and decomposition of Boolean functions, C.Cachin and J Camenisch editors, in Advances in *Cryptology- Eurocrypt'04*, LNCS 3027, Berlin, Germay: Springer-Verlag , pp.474-491, 2004.

[8] A.Braeken, B.Preneel. On the algebraic immunity of symmetric Boolean functions. in *INDOCRYPT 2005*,LNCS 3797, Berlin, Germay: Springer-Verlag , pp.35-48, 2005.

[9] D.K.Dalai, S.Maitra and S. Sarkar. Basic theory in construction of Boolean functions with maximum possible annihilatro immunity. *Des. Codes, Cryptogr.*, 40(1), pp.41-58,2006.

[10] Y.Hu, G.Xiao. Resilient functions over finite fields, *DIEEE Trans. on Infor. theorys*, 49,pp.2040-2046, 2003.

[11] T.W.Cusick, Y.Li. k-th order symmetric SAC Boolean functions and bisecting binomial coefficients. *Discr.Appl.Math* Vol.149,pp.73-86,2005.

[12] Y.Li, T.W.Cusick. Linear structures of symmetric functions over finite fields. *Inf. Process. Lett.*, Vol.97, pp.124-127,2006.

[13] R.A.Brualdi. Introductory Combinatorics , Prentice Hall/Pearson, 2005.