Information Processing Letters 71 (1999) 149–153

Hamming weights of correlation immune Boolean functions

Subhamoy Maitra [a,*], Palash Sarkar [b,1]

[a] *Computer & Statistical Service Center, Indian Statistical Institute, 203, B.T. Road, Calcutta 700 035, India*
[b] *Applied Statistics Unit, Indian Statistical Institute, 203, B.T. Road, Calcutta 700 035, India*

Received 1 April 1999
Communicated by S.G. Akl

Abstract

The set of correlation immune (CI) Boolean functions can be partitioned into several disjoint sets depending on the Hamming weight of their output column. We show that the number of n variable CI functions of Hamming weight $2a+2$ is strictly greater than the number of such functions of weight $2a$ for $2a < 2^{n-1}$. This seemingly intuitive result turns out to be quite difficult to prove. The combinatorial structure of CI functions revealed here reduces the enumeration problem of CI functions to the enumeration problem of balanced CI functions.

Keywords: Combinatorial problems; Cryptography; Correlation immunity; Boolean functions; Bipartite graphs

1. Introduction

The concept of correlation immune Boolean functions was introduced by Siegenthaler [4]. Recently the enumeration of correlation immune Boolean functions has received a lot of attention as evident from [2,5, 3,1]. The set of n-variable Boolean functions can be partitioned into $2^n + 1$ disjoint sets depending on the Hamming weights of their output columns. In this context it is natural to consider the set of CI functions restricted to a particular weight. One particularly interesting question that immediately arises is how does the number of CI functions of a certain weight compare to the number of CI functions of a greater weight. We completely settle this question by showing that

(a) the number of CI functions of odd weight is 0,
(b) the number of CI functions of weight $2a$ is equal to the number of CI functions of weight $2^n - 2a$, and
(c) the number of CI functions of weight $2a$ is strictly less than the number of CI functions of weight $2a + 2$ for $2a < 2^{n-1}$.

Parts (a) and (b) are easy and though (c) is intuitive proving it is a nontrivial task. Our proof technique throws new light on the inherent combinatorial nature of CI functions and also reduces the enumeration problem for CI functions to the enumeration problem for balanced CI functions.

We interpret a Boolean function f as a binary string of length 2^n, given by the output column in the truth table and $wt(f)$ means the number of 1's (Hamming weight) in the string f. The string f^r is the reverse of string f and f^c is the bitwise complement of f. By $S[\tau]$ we mean the τth bit in the binary string S. Also, $\#(\phi)$ counts the number of outcomes favorable to the event ϕ. The notation $(A \mid B)$ denotes the outcomes

* Corresponding author. Email: subho@isical.ac.in.
[1] Email: palash@isical.ac.in.

favorable to A given that B has already occurred. By $D(S_1, S_2)$ we denote the Hamming distance between two strings S_1, S_2 of the same length (say λ). Also, the number of places in which S_1 and S_2 match is denoted by $M(S_1, S_2)$, i.e., $M(S_1, S_2) = \lambda - D(S_1, S_2)$. Let

$M_0(f_1, f_2) = \#\big(f_1[i] = f_2[i] = 0\big)$ and

$M_1(f_1, f_2) = \#\big(f_1[i] = f_2[i] = 1\big), \quad 0 \leqslant i \leqslant 2^n - 1.$

Thus,

$M_0(f_1, f_2) + M_1(f_1, f_2) = M(f_1, f_2).$

Next we define correlation immunity of a Boolean function [4,2].

Definition 1.1. Let f be a Boolean function of n input variables $\{X_1, X_2, \ldots, X_n\}$. Then f is correlation immune if $\text{Prob}(f = X_i) = \frac{1}{2}$, $\forall i,\ 1 \leqslant i \leqslant n$.

The set of all Boolean functions of n variables is denoted by Ω_n, and the set of all correlation immune Boolean functions of n variables is denoted by A_n. Further, $CIW_n(a) = \{f \in A_n \mid wt(f) = a\}$ denotes all n-variable CI functions of weight a and $C_n(a) = |CIW_n(a)|$ denotes the number of such functions.

We here show that,
(a) $C_n(2a + 1) = 0$,
(b) $C_n(2a) = C_n(2^n - 2a)$, and
(c) $C_n(2a) < C_n(2a + 2)$ for $2a < 2^{n-1}$.
The following simple result settles (a).

Proposition 1.1. $\text{Prob}(f = X_i) = \frac{1}{2}$ *iff* $\#(f = 1 \mid X_i = 0) = \#(f = 1 \mid X_i = 1)$ $\forall i,\ 1 \leqslant i \leqslant n$. *Consequently,* $C_n(2a + 1) = 0$, *for* $a \geqslant 0$.

2. Weight distribution

It is easy to see that for $a < 2^{n-1}$, the number of n-variable functions of weight a is less than the number of n-variable functions of weight $a + 1$. This follows from simple properties of binomial coefficients. It is then intuitive to expect the same kind of results for correlation immune functions as well. In this section, we prove such a result. First we show the following which is analogous to the identity

$$\binom{m}{a} = \binom{m}{m-a}.$$

Proposition 2.1. $C_n(2a) = C_n(2^n - 2a)$.

Proof. The result follows on noting that $f \in A_n$ iff $f^c \in A_n$. □

Based on Proposition 2.1, in the rest of this section we will consider $2a < 2^{n-1}$ unless otherwise mentioned.

Proposition 2.2. *Let* $f \in CIW_n(2a)$, $n \geqslant 2$. *Then* $M(f, f^r) \equiv 0 \bmod 4$. *Consequently,* $D(f, f^r)$ *is also congruent to* $0 \bmod 4$.

Proof. Let f^u, f^l be the top and bottom halves (of equal length) of f, respectively. Since $f \in A_n$, we have $wt(f^u) = wt(f^l) = a$. Let there be k places out of the a 1's in the f^u part where the corresponding positions in $(f^l)^r$ do not match, i.e., $M_1(f^u, (f^l)^r) = (a - k)$. Thus, there are k places out of $(2^{n-1} - a)$ 0's in $(f^l)^r$ where the corresponding positions in f^u do not match, which gives $M_0(f^u, (f^l)^r) = 2^{n-1} - a - k$. Hence,

$$\begin{aligned} M(f, f^r) &= 2M\big(f^u, (f^l)^r\big) \\ &= 2\big((a - k) + (2^{n-1} - a - k)\big) \\ &= 2^n - 4k \equiv 0 \bmod 4. \qquad \square \end{aligned}$$

From the argument of the proof of Proposition 2.2, we get the following result.

Proposition 2.3. *Let* $f \in CIW_n(2a)$ *and* $M(f, f^r) = x$. *Then* $M_0(f, f^r) = 2^{n-1} - 2a + \frac{1}{2}x$ *and* $M_1(f, f^r) = -2^{n-1} + 2a + \frac{1}{2}x$. *Consequently,*

$M_0(f, f^r) - M_1(f, f^r) = 2^n - 4a.$

Now we provide a construction technique for $g \in CIW_n(2a + 2)$ from $f \in CIW_n(2a)$ and vice versa.

Definition 2.1. Let $f, g \in \Omega_n$ and there exists i_0, i_1 with $i_0 + i_1 = 2^n - 1$, such that
(1) $f[i_0] = f[i_1] = 0$,
(2) $g[i_0] = g[i_1] = 1$, and
(3) $f[j] = g[j]$ if $j \neq i_0, i_1$.
Then we say that f, g are palindromically related.

Note that values of just a specific pair of positions are toggled and the positions are at the same distances

from top and bottom of the function string. It is important to note that two functions f, g are palindromically related means that $D(f, g) = 2$, i.e., $M(f, g) = 2^n - 2$.

Proposition 2.4. *Let f, g be palindromically related. Then $M(f, f^r) = M(g, g^r)$. Equivalently, $D(f, f^r) = D(g, g^r)$.*

The following result shows the importance of Definition 2.1.

Theorem 2.1. *Let f, g be palindromically related. Then $f \in A_n$ iff $g \in A_n$.*

Proof. Since $f \in CIW_n(2a)$, we have $\#(f = 1 \mid X_i = 0) = \#(f = 1 \mid X_i = 1) = a$ for all i. Also there exists τ such that $f[\tau] = f[2^n - 1 - \tau] = 0$. Consider the column of X_i in the truth table as a binary string. Note that,

$$X_i[\tau] = \big(X_i[2^n - 1 - \tau]\big)^c.$$

Thus, if we consider the function g, then we have,

$$\#(g = 1 \mid X_i = 0) = \#(g = 1 \mid X_i = 1) = a + 1.$$

Thus $g \in A_n$. The other direction can be proved similarly. □

Corollary 2.1. *All palindromic functions are CI.*

Proof. The identity function 0 is trivially correlation immune. The result then follows from Theorem 2.1 by induction on the weight of a palindrome. □

This result has also been proved differently [2]. As an immediate consequence of Corollary 2.1, we have

$$C_n(2a) \geqslant \binom{2^{n-1}}{a}.$$

Interestingly, the exact proportion of $C_n(2a)$ among all functions of weight $2a$ is an open question. However, we have an exact result for $2a = 2$. Since the set $CIW_n(2)$ contains only the palindromes, we get $C_n(2) = 2^{n-1}$.

If we take $f \in CIW_n(2a)$, $2a < 2^{n-1}$, then from Proposition 2.3, we have $M_0(f, f^r) > 0$. Thus there exists at least one position τ such that $f[\tau] = f[2^n - 1 - \tau] = 0$. Then using Definition 2.1, we can get some $g \in CIW_n(2a + 2)$, by replacing the pair of 0's by a pair of 1's. Moreover, if there exists more than one τ, such that $f[\tau] = f[2^n - 1 - \tau] = 0$, then different functions of $CIW_n(2a + 2)$ can be constructed from f. Let us now consider the other way around. Let $g \in CIW_n(2a + 2)$. If $M_1(g, g^r) > 0$, then using Definition 2.1, some $f \in CIW_n(2a)$ can be found. However, it is important to note that there may exist some $g \in CIW_n(2a + 2)$ with $M_1(g, g^r) = 0$. In that case it is not possible to get a function $f \in CIW_n(2a)$ by changing one pair of positions. Motivated by this discussion we make the following definition.

Definition 2.2. Let G_n be an undirected graph where the vertices are the elements of A_n and two vertices are connected if they are palindromically related. We call such a graph an n-variable correlation immune graph.

The following theorem describes the components of G_n.

Theorem 2.2. *If two vertices f, g of the CI graph G_n belong to the same component of G_n then $M(f, f^r) = M(g, g^r)$.*

Proof. The path $u = u_0, u_1, \ldots, u_{k-1}, u_k = v$ exists in G_n iff u_i, u_{i+1}, $0 \leqslant i \leqslant k - 1$, are palindromically related. The result then follows from Proposition 2.4. □

Next we define another graph which is basically a subgraph of the CI graph.

Definition 2.3. By $G_n(2a, 2a + 2)$, we define an undirected bipartite graph such that $G_n(2a, 2a + 2) = (CIW_n(2a) \cup CIW_n(2a + 2), E)$, where there is an edge between $f \in CIW_n(2a)$ and $g \in CIW_n(2a + 2)$ if they are palindromically related.

The following defines a special type of bipartite graph.

Definition 2.4. Let $G = (V_1 \cup V_2, E)$ be a connected bipartite graph. Then G is called homogeneous if all the vertices of V_1 are of the same degree d_1 and all the vertices of V_2 are of the same degree d_2.

Homogeneous bipartite graphs have the following simple property which will prove to be useful later.

Proposition 2.5. *Let $G = (V_1 \cup V_2, E)$ be a homogeneous graph. Let the degree of each vertex of V_1 be d_1 and the degree of each vertex of V_2 be d_2. Then $|V_1| \times d_1 = |V_2| \times d_2 = |E|$. Consequently, if $d_1 > d_2$ then $|V_1| < |V_2|$.*

Our next task is to show that the components of $G_n(2a, 2a+2)$ are homogeneous. We need the following additional notation. Let

$CIW_{n,x}(2a) = \{f \in CIW_n(2a) \mid M(f, f^r) = x\}$ and $C_{n,x}(2a) = |CIW_{n,x}(2a)|$.

By $G_{n,x}(2a, 2a+2)$ we mean the subgraph of $G_n(2a, 2a+2)$ induced by the vertices of $CIW_{n,x}(2a) \cup CIW_{n,x}(2a+2)$. The following relates $G_{n,x}(2a, 2a+2)$ to $G_n(2a, 2a+2)$.

Lemma 2.1. *The subgraphs $G_{n,x}(2a, 2a+2)$ of the graph $G_n(2a, 2a+2)$ are homogeneous for all possible values of x.*

Proof. To prove the statement consider $f, f_1 \in CIW_{n,x}(2a)$. Then, $M(f, f^r) = M(f_1, f_1^r) = x$. Hence, degree of f,

$$d(f) = \frac{M_0(f, f^r)}{2} = \frac{M_0(f_1, f_1^r)}{2} = d(f_1).$$

Similarly, for $g, g_1 \in CIW_{n,x}(2a+2)$, we have, $d(g) = d(g_1)$. Thus $G_{n,x}(2a, 2a+2)$ is homogeneous. □

Lemma 2.2. *Let f, g be vertices of $G_{n,x}(2a, 2a+2)$ where, $f \in CIW_{n,x}(2a)$, $g \in CIW_{n,x}(2a+2)$ and $2a < 2^{n-1}$. Then, the degree of f is $d(f) = \frac{1}{2}M_0(f, f^r)$ and the degree of g is $d(g) = \frac{1}{2}M_1(g, g^r)$ with $d(f) > d(g) > 0$. Consequently, $C_{n,x}(2a) < C_{n,x}(2a+2)$.*

Proof. It is easy to check that, $d(f) = \frac{1}{2}M_0(f, f^r)$ and $d(g) = \frac{1}{2}M_1(g, g^r)$. Using Proposition 2.3 we have

$$d(f) - d(g) = \frac{2^n - 4a - 2}{2}$$

which gives $d(f) > d(g)$ since $2a \leqslant 2^{n-1} - 2$. Also, $M_1(g, g^r) = M_1(f, f^r) + 2$ and hence $d(g) > 0$. The last statement follows from Proposition 2.5. □

Theorem 2.3. *$C_n(2a) < C_n(2a+2)$ for $2a < 2^{n-1}$ and $C_n(2a) > C_n(2a+2)$ for $2a \geqslant 2^{n-1}$.*

Proof. Using Proposition 2.1 it is sufficient to show $C_n(2a) < C_n(2a+2)$ for $2a < 2^{n-1}$. Let there be t distinct values $x_1, x_2, \ldots, x_t$ of $M(f, f^r)$ for $f \in CIW_n(2a)$. From Lemma 2.2, we have, $C_{n,x_i}(2a) < C_{n,x_i}(2a+2)$. Also, it is easy to see that $CIW_{n,x_i}(2a) \cap CIW_{n,x_j}(2a) = \emptyset$ for $x_i \neq x_j$. Hence,

$$\begin{aligned} C_n(2a) &= \sum_{i=1}^{t} C_{n,x_i}(2a) \\ &< \sum_{i=1}^{t} C_{n,x_i}(2a+2) \leqslant C_n(2a+2). \end{aligned}$$

Thus,

$$C_n(2a) < C_n(2a+2) \quad \text{for } 2a < 2^{n-1}. \qquad \square$$

3. Balanced functions

In this section we show that if the set of balanced correlation immune functions $CIW_n(2^{n-1})$ can be enumerated with the cardinality of each partition $CIW_{n,x}(2^{n-1})$ separately, then the exact enumeration of A_n is possible. First we express the proportional cardinality of two sets which follows from Proposition 2.3.

Lemma 3.1.

$$\frac{C_{n,x}(2^{n-1} - 2(i+1))}{C_{n,x}(2^{n-1} - 2i)} = \frac{\frac{1}{2}x - 2i}{\frac{1}{2}x + 2i + 2},$$

for $\frac{1}{2}x - 2i > 0,\ i \geqslant 0$.

Let

$$MATCH_n = \{x \mid M(f, f^r) = x, \text{ for some } f \in CIW_n(2a),\ 2a \leqslant 2^{n-1}\}$$

and

$$BMATCH_n = \{x \mid M(f, f^r) = x, \text{ for some } f \in CIW_n(2^{n-1})\}.$$

The next result shows $MATCH_n = BMATCH_n$.

Lemma 3.2. *If* $f \in CIW_{n,x}(2a)$, $2a < 2^{n-1}$, *then there exists a function* g *such that* $g \in CIW_{n,x}(2^{n-1})$.

Proof. Let $f \in CIW_{n,x}(2a)$. Then it is easy to check that there is a path $f = f_0, f_1, \ldots, f_k = g$ of length $k = 2^{n-2} - a$ in the CI graph G_n, where
(a) $wt(f_i) = 2 + wt(f_{i-1})$ and
(b) $M_0(f_i, f_i^r) = 2^{n-1} - (2a + 2i) + \frac{1}{2}x$ for $i \geqslant 1$.
Thus $g \in CIW_n(2^{n-1})$. Using Proposition 2.4, we have $M(f, f^r) = M(g, g^r)$ and so $g \in CIW_{n,x}(2^{n-1})$. □

Let $BMATCH_n = \{x_1, \ldots, x_t\}$. For $1 \leqslant j \leqslant t$, $0 < 2i \leqslant 2^{n-1}$, $\sigma_{j,2i} = 1$ if $\frac{1}{2}x_j - 2i > 0$ and $\sigma_{j,2i} = 0$, otherwise.

Theorem 3.1.

$$C_n(2^{n-1} - 2i) = \sum_{j=1}^{t} \sigma_{j,2i} C_{n,x_j}(2^{n-1}) \prod_{k=0}^{i-1} \frac{\frac{1}{2}x_j - 2k}{\frac{1}{2}x_j + 2k + 2}.$$

Proof. Using Lemma 3.1 we get, for $i > 0$,

$$C_{n,x_j}(2^{n-1} - 2i) = C_{n,x_j}(2^{n-1}) \prod_{k=0}^{i-1} \frac{\frac{1}{2}x_j - 2k}{\frac{1}{2}x_j + 2k + 2}.$$

Since
(a) $CIW_n(2^{n-1} - 2i)$ is a disjoint union of the sets $CIW_{n,x_j}(2^{n-1} - 2i)$ for $x_j \in BMATCH_n$ and
(b) $CIW_{n,x_j}(2^{n-1} - 2i) = \emptyset$ iff $\sigma_{j,2i} = 0$,
the result holds. □

Theorem 3.2.

$$|A_n| = C_n(2^{n-1}) + 2\sum_{j=1}^{t} C_{n,x_j}(2^{n-1}) \sum_{i=1}^{x_j/4} \prod_{k=0}^{i-1} \frac{\frac{1}{2}x_j - 2k}{\frac{1}{2}x_j + 2k + 2}.$$

Proof. This follows from Theorem 3.1 and Proposition 2.1. The expression $\sigma_{j,2i}$ is removed by summing i from 1 to $\frac{1}{4}x_j$. □

Theorems 3.1 and 3.2 provide formulae for $C_n(2a)$ and $|A_n|$ respectively. To use them one has to determine $BMATCH_n$ and the $C_{n,x_j}(2^{n-1})$'s. These are open problems and could prove to be nontrivial tasks. These also show that the enumeration problem of CI functions reduces to the enumeration problem of balanced CI functions.

The correlation immune functions with the same values of $M(f, f^r)$ form an equivalence class and in each equivalence class there is a hierarchy depending on $wt(f)$. Also, each set $CIW_n(2a)$ can be partitioned depending on $M(f, f^r)$. This gives a new direction to characterize and enumerate the correlation immune Boolean functions.

Acknowledgement

We are grateful to our students Ms. Aditi M. Kolhatkar and Mr. Kaushik Nath for carefully reading this paper and pointing out that the converse of Theorem 2.2 does not necessarily hold.

References

[1] S. Maitra, P. Sarkar, Enumeration of correlation immune Boolean functions, in: Proc. 4th Australasian Conference on Information, Security and Privacy, Lecture Notes in Computer Sci., Vol. 1587, Springer, Berlin, 1999, pp. 12–25.
[2] C.J. Mitchell, Enumerating Boolean functions of cryptographic significance, J. Cryptology 2 (3) (1990) 155–170.
[3] S.M. Park, S. Lee, S.H. Sung, K. Kim, Improving bounds for the number of correlation immune Boolean functions, Inform. Process. Lett. 61 (4) (1997) 209–212.
[4] T. Siegenthaler, Correlation-immunity of nonlinear combining functions for cryptographic applications, IEEE Trans. Inform. Theory IT-30 (5) (1984) 776–780.
[5] Y.X. Yang, B. Guo, Further enumerating Boolean functions of cryptographic significance, J. Cryptology 8 (3) (1995) 115–122.