

Enumeration of Correlation immune boolean functions

A dissertation submitted in partial fulfillment
of the requirements of M.Tech.(Computer Science)
degree of Indian Statistical Institute, Calcutta

by

Kaushik Sharma

under the supervision of

Mr. Subhomoy Maitra
Indian Statistical Institute
Calcutta-700 035.

July 2001

Indian Statistical Institute

203, Barrackpore Trunk Road,

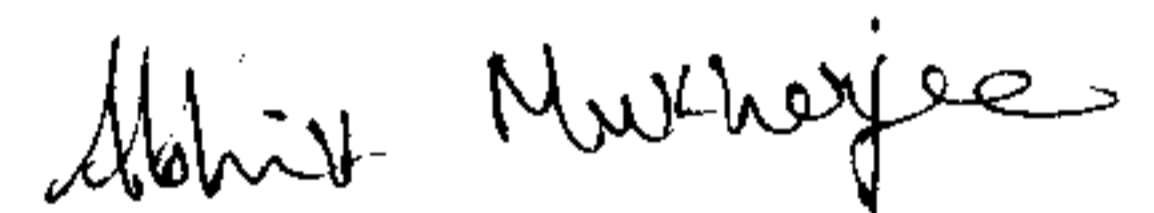
Calcutta-700 035.

Certificate of Approval

This is to certify that this thesis titled "Enumeration of Correlation immune boolean functions" submitted by Kaushik Sharma towards partial fulfillment of requirements for the degree of M.Tech in Computer Science at Indian Statistical Institute, Calcutta embodies the work done under my supervision.



Mr. Subhomoy Maitra,
Computer & Statistical Service center,
Indian Statistical Institute,
Calcutta-700 035.



ABHIK MUKHERJEE

C.S.T,

B.E College (P.U)

Acknowledgements

I take pleasure in thanking Mr Subhomoy Maitra for his friendly guidance throughout the dissertation period. His encouraging words have always kept my sagging spirits up.

I would also like to express my sincere gratitude to Mr Bimal Roy for offering the course in cryptography, which helped me a lot during my dissertation.

I would also like to thank my classmate Mr. Souradyuti paul for all the discussions we had on this subject during the last one year.

Contents

1	Introduction	1
2	Definations	1
3	Some properties of Correlation immune boolean functions	2
4	Enumeration of Correlation immune boolean functions	4
5	Some properties on hamming distance between f and f^r	6
6	Conclusion	8

1 Introduction

Correlation Immunity of boolean functions has long been recognized as one of the critical indicators. There are a lot of successful correlation attacks against a number of stream ciphers. To safeguard against such attacks it is necessary to develop boolean functions of high correlation immunity. In this paper I begin by looking at some definitions and derive some properties of correlation immune functions.

The enumeration of correlation immune functions has also received wide attention. Though till date there is no perfect enumeration there are upper bounds, lower bounds and a lot of asymptotic results on the number of Correlation immune functions. I summarize few such results.

At the end I discuss some properties of correlation immune functions f based on the hamming distance between f and f^r , though the results has been proved for 1st order correlation immune functions in [1], for higher order correlation immune functions I have observed the truth of those results for all correlation immune boolean functions with ≤ 5 input variables, the results are yet to be proved.

2 Definitions

We interpret a boolean function f as a binary string of length 2^n , given by the output column in the truth table where n is the number of input variables in the boolean function.

$\|f\|$ means the number of 1's (hamming weight) in the string f .

The string f^r is the reverse of the string f and f^c is the bitwise complement of f .

By $S[i]$, we mean the i th bit in the binary string S . $\#(\phi)$ counts the number of outcome favorable to the event ϕ . The notation (A / B) denotes the outcomes favorable to A given that the event B has already occurred.

By $D(S_1, S_2)$ we denote the Hamming distance between two strings S_1 and S_2 . Let

$$M_1(f_1, f_2) = \#(f_1[i] = f_2[i] = 1)$$

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

$$M(f_1, f_2) = M_1(f_1, f_2) + M_0(f_1, f_2) = \#(f_1[i] = f_2[i])$$

The set of all boolean functions of n variables is denoted by Ω , while the set of all correlation immune boolean functions is denoted by A_n

Though there are several definitions of correlation immunity, I prefer to work with the definition given by O.V. Denisov in [2].

A Boolean function $f(X_1, X_2, \dots, X_n)$ is called correlation immune of order k ($k= 1, \dots, n-1$), if for all $1 \leq i_1 < i_2 < \dots < i_k \leq n$ and all $\epsilon_1, \epsilon_2, \dots, \epsilon_k \in \{0, 1\}$ the equality

$$\|f_{i_1, i_2, \dots, i_k}^{\epsilon_1, \epsilon_2, \dots, \epsilon_k}\| = 2^k \|f\|$$

holds, where $\|f\|$ is the weight of the function f and $f_{i_1, i_2, \dots, i_k}^{\epsilon_1, \epsilon_2, \dots, \epsilon_k}$ is the sub function of the function f obtained if the variables X_1, X_2, \dots, X_n are replaced by $\epsilon_1, \epsilon_2, \dots, \epsilon_k$.

The set of all the k -th order correlation immune functions of n variables is denoted by $A(n, k)$

Balancedness: A function f is balanced if the number of ones in its output column is equal to the number of zeros.

3 Some properties of Correlation immune boolean functions

Here I give one more definition of correlation immune function of order one

Definition: f is correlation immune if $\text{Prob}(f = X_i) = \frac{1}{2} \forall i, 1 \leq i \leq n$.

Lemma 3.1: $\text{Prob}(f = X_i) = \frac{1}{2}$ iff $\#(f = 1/X_i = 0) = \#(f = 1/X_i = 1) \forall i, 1 \leq i \leq n$

Proof :

$$P(f = X_i) = P(f = X_i/X_i = 0)P(X_i = 0) + P(f = X_i/X_i = 1)P(X_i = 1)$$

$$= \frac{1}{2} \{P(f = 0/X_i = 0) + P(f = 1/X_i = 1)\}$$

$$= \frac{1}{2} \{1 - P(f = 1/X_i = 0) + P(f = 1/X_i = 1)\}$$

Hence $P(f = X_i) = \frac{1}{2}$ iff $P(f = 1/X_i = 0) = P(f = 1/X_i = 1)$
i.e $\#(f = 1/X_i = 0) = \#(f = 1/X_i = 1)$

Theorem 1: A boolean function f is correlation immune of order k iff the equality

$$\|f_{i_1, i_2, \dots, i_k}^{1, 1, \dots, 1}\| = 2^k \|f\|$$

holds for all $l \in \{1, 2, \dots, k\}$ and $\forall 1 \leq i_1 < i_2 < \dots < i_l \leq n$

Proof: Necessity is obvious from the fact that if a function f is correlation immune of order k , then it is also correlation immune of order $l \in 1, \dots, k$

For sufficiency I use induction on the number of variables r which are replaced by zeros.

If $r = 0$,

$$\text{then } \|f_{i_1, i_2, \dots, i_l}^{\epsilon_1, \epsilon_2, \dots, \epsilon_l}\| = \|f_{i_1, i_2, \dots, i_l}^{1, 1, \dots, 1}\| = 2^{-l} \|f\|$$

Let the equality hold for all $1 \leq i_1 < i_2 < \dots < i_l \leq n$ and $\epsilon_1, \epsilon_2, \dots, \epsilon_l$ in $0, 1$ s.t $(\epsilon_1, \epsilon_2, \dots, \epsilon_l)$ contains exactly $r = m$ zeros $m \in \{0, \dots, k-1\}$

Let us prove the equality for tuples $(\epsilon_1, \epsilon_2, \dots, \epsilon_l)$ containing $r = m+1$ 0's

W.L.O.G assume $\epsilon_l = 0$

$$\text{hence } \|f_{i_1, i_2, \dots, i_{l-1}}^{\epsilon_1, \epsilon_2, \dots, \epsilon_{l-1}}\| = 2^{-(l-1)} \|f\|$$

$$\text{hence } \|f_{i_1, i_2, \dots, i_{l-1}, i_l}^{\epsilon_1, \epsilon_2, \dots, \epsilon_{l-1}, 1}\| = 2^{-l} \|f\|$$

$$\text{Now } \|f_{i_1, i_2, \dots, i_l}^{\epsilon_1, \epsilon_2, \dots, \epsilon_l}\| = \|f_{i_1, i_2, \dots, i_{l-1}, i_l}^{\epsilon_1, \epsilon_2, \dots, \epsilon_{l-1}, 0}\|$$

$$= \|f_{i_1, i_2, \dots, i_{l-1}}^{\epsilon_1, \epsilon_2, \dots, \epsilon_{l-1}}\| - \|f_{i_1, i_2, \dots, i_{l-1}, i_l}^{\epsilon_1, \epsilon_2, \dots, \epsilon_{l-1}, 1}\|$$

$$= 2^{-(l-1)} \|f\| - 2^{-l} \|f\| = 2^{-l} \|f\|$$

Hence the theorem is proved.

Lemma 3.2: A boolean function f is correlation immune of order k then any subfunction obtained from it by replacing any one variable by a constant is correlation immune of order $k-1$.

Proof: By definition f is correlation immune of order k iff $\|f_{i_1, i_2, \dots, i_l}^{\epsilon_1, \epsilon_2, \dots, \epsilon_l}\| = 2^{-l} \|f\|$

now if we obtain a sub-function f_i^c it has to be correlation immune of order $k-1$, as if not so we can suitably fix $k-1$ variables and the equality A will be violated.

Lemma 3.3: For f a k -th order correlation immune boolean function $\|f\|$ is divisible by 2^k

Proof: As the hamming weight of all the $(n-k)$ variable sub-function is $2^{-k} \|f\|$. It is obvious that $\|f\|$ is divisible by 2^k

lemma 3.4: If we divide the string of a kth order correlation immune function f into 2^k equal parts then each chunk of the truth table has same number of 1's

Proof: Each chunk is a subfunction of f where the 1st k variables X_1, X_2, \dots, X_k are fixed. Hence each chunk has same weight.

4 Enumeration of Correlation immune boolean functions

Recently enumeration of correlation immune boolean function has received wide attention, though no exact formula is obtained for the number of k-th order correlation boolean function on n input variables, a lot of asymptotic results are available. The enumeration of 1st order correlation immune functions is also an open problem. In their paper [1] the authors has proved that the enumeration problem of 1st order correlation immune functions can be reduced to the problem of enumerating the set of balanced correlation immune functions. Unfortunately enumerating all balanced correlation immune functions is also an open problem.

The paper also pointed out exactly how the number of correlation immune functions of a certain weight is related to the number of correlation immune functions of greater weight. It was shown that

- a) The number of correlation immune functions of odd weight is zero.
- b) The number of correlation immune functions of weight $2a$ is equal to the number of correlation immune functions of weight $2^n - 2a$.
- c) The number of correlation immune functions of weight $2a$ is strictly less than the number of correlation immune functions of weight $2a+2$ for $2a < 2^{n-1}$.

This paper also showed that all palindromic functions are correlation immune. Here I provide a much easier proof for the same problem.

Theorem 2: All boolean functions f whose truth-table representation is a palindromic string is 1st order correlation immune

Proof: Though this result is already proved in [1], I provide a easier prove.

If f is a palindromic string.

$$f[\tau] = 1 \Leftrightarrow f[2^n - 1 - \tau] = 1$$

Moreover $X_i[\tau] = X_i[2^n - 1 - \tau]^c$

Hence for each τ s.t $f[\tau] = 1$ and $X_i[\tau] = 1$ (0)

we get $f[2^n - 1 - \tau] = 1$ and $X_i[2^n - 1 - \tau] = 0$ (1)

hence $\#(f = 1/X_i = 0) = \#(f = 1/X_i = 1)$

and hence f is correlation immune.

This result gives a very weak lower bound on the number of correlation immune boolean functions

$$|A_n| \geq 2^{2^{n-1}}$$

Weak since there are many functions which are not palindrome but are correlation immune.

For enumerating the correlation immune the authors partitioned the set Ω_n in two different ways

1) In terms of hamming weights.

2) In terms of the hamming distance between f and f^r

$$CIW_{n,x}(2a) = \{f : f \in A_n \text{ with } \|f\| = 2a \wedge M(f, f^r) = x\}$$

$$C_{n,x} = |CIW_{n,x}(2a)|$$

With this kind of partitioning they used some properties of homogeneous bipartite graphs to show that

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

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

It was also shown that if $f \in CIW_{n,x}(2a)$ s.t $2a \nmid 2^{n-1}$ then \exists a boolean function g s.t $g \in CIW_{n,x}(2^{n-1})$ and finally it was shown

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

This paper also had a result on the distance between f and f^r for a correlation immune function f . It says that this distance will be divisible by 4. I will discuss this later.

Principle of inclusion and exclusion can be applied to enumerate A_n in terms of other sets

Defination: A function f is called correlation immune with respect to a variable X_i if $\#(f = X_i) = \#(f \neq X_i)$

$$A_{X_i} = \{ \#(f = X_i) \neq \#(f \neq X_i) \}$$

then $A_n = \overline{A_{X_1}} \cap \overline{A_{X_2}} \cap \dots \cap \overline{A_{X_n}}$

Hence using principle of inclusion exclusion we get $|A_n| = 2^{2^n} - \binom{n}{1} |\overline{A_{X_1}}| - \binom{n}{2} |\overline{A_{X_1}} \cap \overline{A_{X_2}}| \dots + (-1)^n \binom{n}{n} |\overline{A_{X_1}} \cap \overline{A_{X_2}} \cap \dots \cap \overline{A_{X_n}}|$

$|A_{X_i}| = \binom{2^n}{2^{n-1}}$ since we have to choose 2^{n-1} position out of 2^n and then ensure that those positions are filled up s.t $f=X_i$. For the remaining positions f

But it is not easy to know $|\overline{A_{X_1}} \cap \overline{A_{X_2}} \cap \dots \cap \overline{A_{X_i}}|$ hence attacking the enumeration problem in this way is not likely to succeed.

5 Some properties on hamming distance between f and f^r

Theorem 3: If f is a 1st order correlation immune function then $M(f, f^r)$ is divisible by 4.

Proof: Let $f = [f_u, f_l]$, where f_u is the 1st half of the truth table string of f while f_l is the 2nd half of the truth table string of f .

$$D(f, f^r) = D(f_u, f_l^r) + D(f_l, f_u^r)$$

But $D(f_u, f_l^r) = D(f_l, f_u^r)$

hence $D(f, f^r) = 2D(f_u, f_l^r)$

$\|f_u\| = \|f_l\| = a$ as proved in lemma 3.4

Let there be k positions out of the a 1's in f_u , corresponding to which we have 0's in f_l

My claim is that there are exactly k positions in f_u which has 0's and corresponding to those positions f_l has 1's

As $M_1(f_u, f_l^r) = a - k = M_1(f_l, f_u^r)$

Thus $D(f_u, f_l) = 2k$

Similarly the hamming distance $D(f_l, f_u)$ is $2k$.

Hence the hamming distance $D(f, f^r)$ is $4k$.

Hence if $n \geq 2$ $M(f, f^r) = 2^n - 4k$ is divisible by 4.

This result was proved in [1] in slightly different way.

A very intuitive result will be that for all 2nd order correlation immune functions $M(f, f^r)$ will be divisible by 8 and a generalization that for all k th order correlation immune functions $M(f, f^r)$ will be divisible by 2^{k+1} .

To gain more confidence I run a program and calculated $M(f, f^r)$ for all 2nd order correlation immune functions on 4 variables and found out that the result is true.

Then I also run the program for 2nd order and 3rd order correlation immune functions of 5 variables I found the result to be true.

Though it took a lot of time to run this program I also found that if for a 2nd order correlation immune function the output string and its reverse are matched by sectioning them into four parts then the number of mismatches in all the four parts are same, i.e if $f = [f_1 f_2 f_3 f_4]$ then the number of mismatches between $f_1 \& f_4^r$, $f_2 \& f_3^r$, $f_3 \& f_2^r$, $f_4 \& f_1^r$ are same moreover in each of these subsections the number of positions s.t the L.H.S column has 0 while R.H.S column has 1 is same as the number of positions s.t the L.H.S column has 1 while R.H.S column has 0. A similar observation was made when I run the program for 2nd and 3rd order correlation immune

functions on five input variables.

So I tried to prove that for a k th order correlation immune function if the output column of f and f^r are divided into 2^k sections then for each section the number of mismatch is same and is of form $2t$. If somehow such a result can be proved then my conjecture that $M(f, f^r)$ is divisible by 2^{k+1} will come true. Unfortunately this does not seem to be an easy problem either.

6 Conclusion

So enumeration problem for the set of correlation immune functions remains unsolved. But a reasonably good characterization of correlation immune functions has been made in terms of the Hamming distance between f and f^r . I conclude my work with the conjecture that for any k th order correlation immune function f , $D(f, f^r)$ is divisible by 2^{k+1} .

References

- [1] Subhomoy Maitra, Palash Sarkar *Hamming weights of correlation immune Boolean functions* Information Processing Letters 71(1999) pg 149-153.
- [2] O.V Denisov, *An asymptotic formula for the number of correlation immune of order k boolean functions*, Discrete Math Application, Vol 2, No.4, pg 407-426.
- [3] Yuriy Tarannikov, *Ramsey-like Theorems on the structure and Numbers of Higher Order Correlation immune functions.*