COUNTING AND CONSTRUCTING BOOLEAN FUNCTIONS WITH PARTICULAR DIFFERENCE DISTRIBUTION VECTORS

ELİF YILDIRIM

JUNE 2004

COUNTING AND CONSTRUCTING BOOLEAN FUNCTIONS WITH PARTICULAR DIFFERENCE DISTRIBUTION VECTORS

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF APPLIED MATHEMATICS

OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

ELİF YILDIRIM

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF CRYPTOGRAPHY

JUNE 2004

Approval of the Graduate School of Applied Mathematics

Prof. Dr. Aydın AYTUNA
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Ersan AKYILDIZ
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Ali DOĞANAKSOY
Supervisor

Examining Committee Members

Prof. Dr. İsmail Ş. GÜLOĞLU

Prof. Dr. Ersan AKYILDIZ

Assoc. Prof. Dr. Ali DOĞANAKSOY

Assist. Prof. Dr. Ali Aydın SELÇUK

Dr. Muhiddin UĞUZ

Abstract

COUNTING AND CONSTRUCTING BOOLEAN FUNCTIONS WITH PARTICULAR DIFFERENCE DISTRIBUTION VECTORS

Yıldırım, Elif
M.Sc., Department of Cryptography
Supervisor: Assoc. Prof. Dr. Ali DOĞANAKSOY

June 2004, 40 pages

In this thesis we deal with the Boolean functions with particular difference distribution vectors. Besides the main properties, we especially focus on strict avalanche criterion for cryptographic aspects. Not only we deal with known methods we also demonstrate some new methods for counting and constructing such functions.

Furthermore, performing some statistical tests, we observed a number of interesting properties.

Keywords: SAC, Difference Distribution Vector, Boolean Functions, Counting SAC Satisfying Functions, Cryptography.

Öz

BELİRLİ FARK DAĞILIM VEKTÖRLERİNE SAHİP BOOLE FONKSİYONLARININ SAYILMASI VE TEŞKİLİ

Yıldırım, Elif
Yüksek Lisans, Kriptografi Bölümü
Tez Yöneticisi: Doç. Dr. Ali DOĞANAKSOY

Haziran 2004, 40 sayfa

Bu tezde belirli fark dağılım vektörlerine sahip Boole fonksiyonları ile ilgilenilmiştir. Temel özelliklerinin yanında kriptografik açıdan keskin çığ etkisini sağlayan fonksiyonlar dikkate alınmıştır. Daha önceden bilinen metodların yanısıra yeni sayma ve teşkil etme yöntemlerine de yer verilmitir.

İstatistiksel metodlar kullanılarak ilginç özellikler de elde edilmiştir.

Anahtar Kelimeler: SAC, Boole fonksiyonları, Keskin Çığ Etkisi, SAC'ı Sağlayan Fonksiyonların Sayılması, Kriptografi.

To my family

Acknowledgments

I am grateful to Assoc. Prof. Dr. Ali Doğanaksoy for patiently guiding, motivating, and encouraging me throughout this study.

I want to thank my parents without whose encouragement this thesis would not be possible.

I am also grateful to Zülfükar Saygı for being with me all the way.

Table of Contents

List of Tables

Chapter 1

Introduction

Boolean functions are basic elements of most building blocks used in cryptographic applications. A Boolean function maps a number of input bits into a single bit. A cryptographic function must have high algebraic degree, as systems using Boolean functions can be attacked if the function have low degree. Also, to avoid the statistical dependence between input and output, the function should be balanced. That is, the number of 1's should be equal to the number of 0's. Moreover, the distance to affine functions, so called nonlinearity should be high, otherwise an affine aproximation of the function can be used to build attacks on the system.

For the cryptographic reasons above, it is natural to expect each input bit has some effect on the output bit. A function with this property is said to be complete. And the concept of completeness was first introduced by Kam and Davida [3] in 1979. The concept of avalanche effect which means an average of one half of the output bits should be changed whenever a singe input bit is complemented, was first introduced by Feistel [1] in 1973. Webster and Tavares [9], in order to combine the concepts of avalanche and completeness, introduced the concept of strict avalanche criterion. A cryptographic function is said to satisfy the strict avalanche criterion whenever a single input bit is complemented, each output bit changes with a probability one half. Afterwards, the notion of the strict avalanche criterion was extended in 1988 by Forre [2]. She defined the strict avalanche criterion (SAC) of order m. A function satisfies the SAC

of order m if and only if any subfunctions obtained from the original function by keeping m of its input bits constant satisfies SAC. The order of the original strict avalanche criterion is 0. In 1991, Preneel et. al. [5] introduced the concept of propagation criterion of degree k to generalize the strict avalanche criterion. A function is said to satisfy the propagation criterion of degree k if complementing at most k bits of the input, the output changes with probability exactly one half. By definition, propagation criterion of degree 1 is the strict avalanche criterion.

The outline of the thesis is as follows:

In Chapter 2, we establish some notations which are used throughout the thesis and define the desired properties of Boolean functions.

In Chapter 3, we recall the difference distribution vector. Some basic properties of functions having certain types of difference distribution vectors, and the number of such functions.

In Chapter 4, we give constructions of Boolean functions satisfying cryptographically good properties such as the strict avalanche criterion and balancedness.

In Chapter 5, we list some statistical results and observations on the distribution of difference distribution vectors.

In last chapter, we give the conclusion of our study and future work suggestions.

Chapter 2

Preliminaries

In this chapter we state the definitions and give the notations about the concepts of cryptography that we will deal with. For further definitions and notations, reader may refer to [6].

Let $\mathcal{V}_n$ be the vector space of all n-tuples of elements from $GF(2)$. By setting $k = \sum_{i=1}^{n} a_i 2^{n-i}$, an element $\alpha_k = (a_1, a_2, \ldots, a_n) \in \mathcal{V}_n$ corresponds the integer k. So there is a one-to-one correspondence between vectors of $\mathcal{V}_n$ and integers $\{0, 1, \ldots, 2^n - 1\}$. Via this representation, $\mathcal{V}_n$ assumes a natural ordering called the *lexicographic ordering*. We denote the element of $\mathcal{V}_n$ corresponding to the integer k by α_k so that, $\mathcal{V}_n = \{\alpha_0, \ldots, \alpha_{2^n-1}\}$ and $\alpha_0 \leq \alpha_1 \leq \cdots \leq \alpha_{2^n-1}$.

For $\alpha, \beta \in \mathcal{V}_n$, the *sum* $(\alpha \oplus \beta) \in \mathcal{V}_n$ is obtained by adding corresponding components of α and β modulo 2.

In $\mathcal{V}_n$, we denote the vectors having only one nonzero entry in the i-th position by e_i, and these vectors constitute a basis for V_n, called *standard basis*.

The standard inner product on $\mathcal{V}_n$ is defined by $< \alpha, \beta >= \sum_{i=1}^{n} a_i b_i \in GF(2)$ for $\alpha = (a_1, a_2, \ldots, a_n)$ and $\beta = (b_1, b_2, \ldots, b_n)$.

2.1 Boolean Functions

A *Boolean function* is a $GF(2)$ valued map defined on $\mathcal{V}_n$ and the set of all Boolean functions on $\mathcal{V}_n$ is denoted by $\mathcal{F}_n$. In the rest of this thesis, unless otherwise stated, "function" will stand for "Boolean function".

Any Boolean function $f \in \mathcal{F}_n$ has a unique representation in each of the following forms:

- The ordered tuple

$$T_f = (f(\alpha_0), f(\alpha_1), \ldots, f(\alpha_{2^n-1}))$$

  is called the *truth table* of f.

- The unique polynomial representation

$$f(x) = a_0 \oplus a_1 x_1 \oplus \cdots \oplus a_n x_n \oplus a_{12} x_1 x_2 \oplus \cdots \oplus a_{12\ldots n} x_1 x_2 \cdots x_n$$

  is called the *algebraic normal form* . In the algebraic normal form of a function, each product of variables appearing as a part of the sum is called a *term*. Number of variables in each term is called the degree of that term. The *degree of a function* is the degree of the term with largest degree in its algebraic normal form, and is denoted by $deg(f)$. The *degree of a variable* x_i is the degree of highest degree term in which the variable x_i appears, denoted by $deg(f, x_i)$.

The *Hamming weight* of a function is defined as the number of nonzero entries in the truth table of f and is denoted by $w(f)$.

$1_n, 0_n \in \mathcal{F}_n$ are constant functions which map all inputs to $1, 0 \in GF(2)$, respectively. Truth table of the constant function 1_n is $(1, 1, \ldots, 1)$ and that of 0_n is $(0, 0, \ldots, 0)$. Note that $w(1_n) = 2^n$ and $w(0_n) = 0$.

For a function $f \in \mathcal{F}_n$, the *complement function* $\bar{f} \in \mathcal{F}_n$ is defined as $\bar{f}(x) = f(x) \oplus 1$ for all $x \in \mathcal{V}_n$. From this definition, it follows that $w(\bar{f}) = 2^n - w(f)$.

Support of a function $f \in \mathcal{F}_n$ is defined to be the set $\{\alpha \in \mathcal{V}_n | f(\alpha) = 1\}$ and is denoted by $Supp(f)$. It is clear that $|Supp(f)| = w(f)$ and using the above definition, $Supp(f) \cap Supp(\bar{f}) = \emptyset$.

A function is called *balanced* if the number of 1's is equal to the number of 0's in its truth table. It follows that, $f \in \mathcal{F}_n$ is balanced if and only if $w(f) = 2^{n-1}$. Clearly, the number of balanced functions is $\binom{2^n}{2^{n-1}}$.

A function $f \in \mathcal{F}_n$ is called *linear* if $f(x \oplus y) = f(x) \oplus f(y)$ for all $x, y \in \mathcal{V}_n$. Any linear function is of the form $f(x) = a_1x_1 \oplus a_2x_2 \oplus \cdots \oplus a_nx_n$. The set of linear functions is denoted by $\mathcal{L}_n$. A function $f \in \mathcal{F}_n$ is called *affine* if $f(x \oplus y) = f(x) \oplus f(y) \oplus a$ for all $x, y \in \mathcal{V}_n$ where $a \in \{0, 1\}$. Any affine function is of the form $f(x) = a_0 \oplus a_1x_1 \oplus a_2x_2 \oplus \cdots \oplus a_nx_n$. The set of affine functions is denoted by $\mathcal{A}_n$. It is clear from the above definitions that, $\mathcal{L}_n \subset \mathcal{A}_n$ and that $2|\mathcal{L}_n| = |\mathcal{A}_n| = 2^{n+1}$.

Theorem 2.1.1. *[6] Any non-constant affine function is balanced .*

2.2 Binary operations

For $f,\ g \in \mathcal{F}_n$, the *sum* $f \oplus g$ and the *product* fg are defined by

$$(f \oplus g)(x) = f(x) \oplus g(x)$$

and

$$(fg)(x) = f(x)g(x)$$

for all $x \in \mathcal{V}_n$, respectively.

Let $g_1, g_2, \ldots, g_{2^k} \in \mathcal{F}_m$, then by $f = (g_1 \ || \ g_2 \ || \ \ldots \ || \ g_{2^k}) \in \mathcal{F}_{m+k}$ we denote the function whose truth table is the *concatenation* of the truth tables of $g_1, g_2, \ldots, g_{2^k}$ in the given order. In other words, f is a function whose truth table is $T_f = (T_{g_1}||T_{g_2}|| \ldots ||T_{g_{2^k}})$, where $||$ stands for the concatenation of 2^m-tuples. We write $f(x) = f(y, z) = (g_1(y) \ || \ g_2(y) \ || \ \ldots \ || \ g_{2^k}(y)) \in \mathcal{F}_{m+k}$, where $x = (y, z)$, $y = (x_1, \ldots, x_m)$, and $z = (x_{m+1}, \ldots, x_{m+k})$.

Given $f \in \mathcal{F}_k$ and $g \in \mathcal{F}_m$ the *Kronecker product* $f \otimes g$ of f and g is defined by setting

$$f \otimes g = (f(\alpha_0)g \mid\mid f(\alpha_1)g \mid\mid \cdots \mid\mid f(\alpha_{2^k-1})g).$$

From the definition, it follows that $f \otimes g$ and $g \otimes f$ are in $\mathcal{F}_{k+m}$ but, in general, $f \otimes g \neq g \otimes f$. By $(f \otimes g)(x) \in \mathcal{F}_{k+m}$, we mean $(f \otimes g)(y, z) = f(y)g(z)$ where $x = (x_1, \ldots, x_{k+m})$, $y = (x_1, \ldots, x_k)$ and $z = (x_{k+1}, \ldots, x_{k+m})$

Theorem 2.2.1. *For $f \in \mathcal{F}_k$ and $g \in \mathcal{F}_m$, the algebraic normal form of $(f \otimes g)(x) = (f \otimes g)(y, z)$ is the product of the algebraic normal forms of $f(y)$ and $g(z)$ where $x = (x_1, \ldots, x_{k+m})$, $y = (x_1, \ldots, x_k)$ and $z = (x_{k+1}, \ldots, x_{k+m})$.*

Proof: From the definition,
$f \otimes g \in \mathcal{F}_{k+m}$ is $(f \otimes g)(x) = (f \otimes g)(y, z) = f(y)g(z)$. □

Example 2.2.2. Let $f \in \mathcal{F}_2$ and $g \in \mathcal{F}_3$ be given by the truth tables $T_f = (1, 1, 0, 1)$ and $T_g = (1, 0, 0, 1, 1, 1, 0, 0)$. Then the truth table of $f \otimes g$ is

$$T_{f\otimes g} = (\mathbf{1,0,0,1,1,1,0,0}, 1, 0, 0, 1, 1, 1, 0, 0, \mathbf{0,0,0,0,0,0,0,0}, 1, 0, 0, 1, 1, 1, 0, 0)$$

On the other hand the truth table of $g \otimes f$ is

$$T_{g\otimes f} = (\mathbf{1,1,0,1}, 0, 0, 0, 0, \mathbf{0,0,0,0}, 1, 1, 0, 1, \mathbf{1,1,0,1}, 1, 1, 0, 1, \mathbf{0,0,0,0}, 0, 0, 0, 0).$$

The algebraic normal form of f is

$$f(x_1, x_2) = 1 \oplus x_1 \oplus x_1x_2$$

and the algebraic normal form of g is

$$g(x_1, x_2, x_3) = 1 \oplus x_2 \oplus x_3 \oplus x_1x_3.$$

From the above theorem it follows that, the algebraic normal form of $f \otimes g$ is

$$(f\otimes g)(x_1, x_2, x_3, x_4, x_5) = f(x_1, x_2)g(x_3, x_4, x_5) = (1\oplus x_1\oplus x_1x_2)(1\oplus x_4\oplus x_5\oplus x_3x_5)$$

$= 1 \oplus x_1 \oplus x_4 \oplus x_5 \oplus x_1x_2 \oplus x_1x_4 \oplus x_1x_5 \oplus x_3x_5 \oplus x_1x_2x_4 \oplus x_1x_2x_5 \oplus x_1x_3x_5 \oplus x_1x_2x_3x_5.$
The algebraic normal form of $g \otimes f$ is,

$(g \otimes f)(x_1, x_2, x_3, x_4, x_5) = g(x_1, x_2, x_3)f(x_4, x_5) = (1 \oplus x_2 \oplus x_3 \oplus x_1x_3)(1 \oplus x_4 \oplus x_4x_5)$

$= 1 \oplus x_2 \oplus x_3 \oplus x_4 \oplus x_1x_3 \oplus x_2x_4 \oplus x_3x_4 \oplus x_4x_5 \oplus x_1x_3x_4 \oplus x_2x_4x_5 \oplus x_3x_4x_5 \oplus x_1x_3x_4x_5.$

The *Hamming distance* of two functions $f, g \in \mathcal{F}_n$ is the cardinality of the set $\{x \in \mathcal{V}_n | f(x) \neq g(x)\}$ and is denoted by $d(f, g)$. It is clear from the definitions that $d(f, g) = w(f \oplus g) = w(f) + w(g) - 2w(fg)$ for any $f, g \in \mathcal{F}_n$.

2.3 Desired Cryprographic Properties of Boolean Functions

For any $f \in \mathcal{F}_n$, it is important to compare the output of $f(x)$ and $f(x \oplus \alpha)$ for $\alpha \in \mathcal{V}_n$. For a "random" function, one expects to have $f(x) = f(x \oplus \alpha)$ with probability one half or equivalently, to have $\sum_x f(x) \oplus f(x \oplus \alpha) = 2^{n-1}$ or $\sum_x (-1)^{f(x) \oplus f(x \oplus \alpha)} = 0$. The quantity $\sum_x (-1)^{f(x) \oplus f(x \oplus \alpha)}$ is called the *auto correlation coefficient* and is denoted by $\Delta_f(\alpha)$.

An element $\alpha \in \mathcal{V}_n$ is called a *linear structure* of f if $\Delta_f(\alpha) = 2^n$. On the other hand, if $\Delta_f(\alpha) = 0$, f is said to satisfy *propagation criterion* with respect to α. If f satisfies the propagation criterion with respect to all α with $w(\alpha) \leq k$ then f is said to satisfy *propagation criterion of order k*, and denoted by $PC(k)$.

It is observed that, a function $f \in \mathcal{F}_n$ satisfies $PC(1)$, whenever an input bit is complemented, the output bit changes with probability one half. Naturally, this is a desired cryptographic property. This important property is called the *strict avalanche criterion*.

Chapter 3

Strict Avalanche Criterion

In this chapter, we deal with strict avalanche criterion for Boolean functions. Besides basic properties already known, we also introduce some new results.

For a given $f \in \mathcal{F}_n$, $S_i(f)$ denotes $1/2(2^n - \Delta_f(e_i))$, that is

$$S_i(f) = 1/2[2^n - \sum_x (-1)^{f(x)\oplus f(x\oplus e_i)}]$$

or equivalently,

$$S_i(f) = \sum_x f(x) \oplus f(x \oplus e_i).$$

By $S(f)$ we denote the vector $(S_1(f), S_2(f), \ldots, S_n(f))$. When the function is clear for the given context, we just write $S = (S_1, S_2, \ldots, S_n)$. $S(f)$ is called the *difference distribution vector* of f. It follows that $f \in \mathcal{F}_n$ satisfies the strict avalanche criterion if and only if $S_i(f) = 2^{n-1}$ for $i \in \{1, 2, \ldots, n\}$. Also by $S_{max}(f)$, we denote the maximum value of $S_i(f)$ for $i \in \{1, 2, \ldots, n\}$.

Alternatively, the following characterization of $S_i(f)$ for $i \in \{1, 2, \ldots, n\}$ can be given by using the partial derivative of a function. Namely, since for any x_i, the function $f \in \mathcal{F}_n$ can be written as $f = x_i u \oplus v$ where $u, v \in \mathcal{F}_{n-1}$ and are independent of x_i , $S_i(f) = 2w(u)$. Note that, since the characteristic of the field we deal with is 2, u is in fact the formal partial derivative of f with

respect to x_i. It follows that,

$$S_i(f) = 2w(\frac{\partial f}{\partial x_i}).$$

3.1 Properties of Difference Distribution Vector

For any fixed $a \in \{0, \ldots, 2^n\}$ the number of functions satisfying $S_i(f) = a$ does not depend on the choice of $i \in \{1, \ldots, n\}$. That is,

$$|\{f \in F_n | S_i(f) = a\}| = |\{f \in F_n | S_j(f) = a\}|$$

for any pair $i, j \in \{1, \ldots, n\}$. For $a = 2^{n-1}$, this number is denoted by $S(n, 1)$. We generalize this idea in an obvious manner to define $S(n, k)$ as follows. The number of functions satisfying the condition $S_{i_1}(f) = a_1, S_{i_2}(f) = a_2, \cdots, S_{i_k}(f) = a_k$, does not depend on the choice of the subset $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$. Thus we define $S(n, k)$ to be the number of functions satisfying $S_{i_1}(f) = S_{i_2}(f) = \cdots = S_{i_k}(f) = 2^{n-1}$, where $\{i_1, \ldots, i_k\}$ is any subset of $\{1, \ldots, n\}$ with cardinality k.

Lemma 3.1.1. *Let $f_1, f_2, \ldots, f_{2^k} \in \mathcal{F}_k$. If $T_{f_1}, T_{f_2}, \ldots, T_{f_{2^k}}$ are linearly independent in $\mathcal{V}_{2^k}$, then any $f \in \mathcal{F}_n$ $(k \leq n)$ can be written uniquely as*

$$f = (f_1 \otimes g_1) \oplus (f_2 \otimes g_2) \oplus \cdots \oplus (f_{2^k} \otimes g_{2^k})$$

where $g_1, g_2, \ldots, g_{2^k} \in \mathcal{F}_{n-k}$.

Proof: From the definition of Kronecker product, it follows that the truth

table of f is

$$\begin{aligned} T_f &= ([f_1(\alpha_0)T_{g_1} \oplus f_2(\alpha_0)T_{g_2} \oplus \cdots \oplus f_{2^k}(\alpha_0)T_{g_{2^k}}], \\ &\quad [f_1(\alpha_1)T_{g_1} \oplus f_2(\alpha_1)T_{g_2} \oplus \cdots \oplus f_{2^k}(\alpha_1)T_{g_{2^k}}], \cdots, \\ &\quad [f_1(\alpha_{l-1})T_{g_1} \oplus f_2(\alpha_{l-1})T_{g_2} \oplus \cdots \oplus f_{2^k}(\alpha_{l-1})T_{g_{2^k}}], \cdots, \\ &\quad [f_1(\alpha_{2^k-1})T_{g_1} \oplus f_2(\alpha_{2^k-1})T_{g_2} \oplus \cdots \oplus f_{2^k}(\alpha_{2^k-1})T_{g_{2^k}}]). \end{aligned}$$

Hence, the components of T_f between positions $(l-1)\cdot 2^m + 1$ and $l \cdot 2^m$ are

$$\begin{aligned} [f(\alpha_{(l-1)2^m}, \ldots, f(\alpha_{(l2^m)-1})] &= \bigoplus_{i=1}^{2^k} f_i(\alpha_{l-1})Tg_i \\ &= \bigoplus_{i=1}^{2^k} f_i(\alpha_{l-1})[g_i(\alpha_0), g_i(\alpha_1), \ldots, g_i(\alpha_{2^m-1})] \end{aligned}$$

where $m = n - k$.

Fixing an arbitrary $t \in \{1, \ldots, 2^m\}$ we obtain the following 2^k equations

$$\begin{aligned} f(\alpha_t) &= f_1(\alpha_0)g_1(\alpha_t) \oplus f_2(\alpha_0)g_2(\alpha_t) \oplus \cdots \oplus f_{2^k}(\alpha_0)g_{2^k}(\alpha_t) \\ f(\alpha_{2^m+t}) &= f_1(\alpha_1)g_1(\alpha_t) \oplus f_2(\alpha_1)g_2(\alpha_t) \oplus \cdots \oplus f_{2^k}(\alpha_1)g_{2^k}(\alpha_t) \\ &\vdots \\ f(\alpha_{((2^k-1)2^m)+t}) &= f_1(\alpha_{2^k-1})g_1(\alpha_t) \oplus f_2(\alpha_{2^k-1})g_2(\alpha_t) \oplus \cdots \oplus f_{2^k}(\alpha_{2^k-1})g_{2^k}(\alpha_t). \end{aligned}$$

The columns of the coefficient matrix are truth tables of f_i's, whose linear independence is guaranteed by the hypothesis, so the system has a unique solution for $g_1(\alpha_t), g_2(\alpha_t), \ldots, g_{2^k}(\alpha_t)$. Since t is arbitrary, lemma follows. □

Now we introduced two distinguished examples that will be used throughout the chapter.

Example 3.1.2. Let $f_1, f_2 \in \mathcal{F}_1$. If T_{f_1} and T_{f_2} are linearly independent in $\mathcal{V}_2$ then any f in $\mathcal{F}_n$ can be written uniquely as $f = (f_1 \otimes g_1) \oplus (f_2 \otimes g_2) \in \mathcal{F}_n$ where $g_1, g_2 \in \mathcal{F}_{n-1}$. For some particular choices of f_1 and f_2 we compute weight of f and $S_i(f)$, $i = 1, \ldots, n$:

a) Let $T_{f_1} = (1,0)$, $T_{f_2} = (0,1)$.

$$\begin{aligned} w(f) &= w(g_1) + w(g_2), \\ S_1(f) &= 2w(g_1 \oplus g_2), \\ S_i(f) &= S_{i-1}(g_1) + S_{i-1}(g_2), \quad i = 2, \ldots, n. \end{aligned}$$

b) Let $T_{f_1} = (1,1)$, $T_{f_2} = (0,1)$.

$$\begin{aligned} w(f) &= w(g_1) + w(g_1 \oplus g_2), \\ S_1(f) &= 2w(g_2), \\ S_i(f) &= S_{i-1}(g_1) + S_{i-1}(g_1 \oplus g_2), \quad i = 2, \ldots, n. \end{aligned}$$

Example 3.1.3. For $f_1, f_2, f_3, f_4 \in \mathcal{F}_2$ let the truth tables be linearly independent in $\mathcal{V}_4$. Then any f in $\mathcal{F}_n$ can be written uniquely as $f = (f_1 \otimes g_1) \oplus (f_2 \otimes g_2) \oplus (f_3 \otimes g_3) \oplus (f_4 \otimes g_4) \in \mathcal{F}_n$ where $g_1, g_2, g_3, g_4 \in \mathcal{F}_{n-2}$. Now we compute the weight of f, and $S_i(f)$ $i = 1, \ldots, n$, for some particular choices of f_1, f_2, f_3 and f_4:

a) Let $T_{f_1} = (1,0,0,0)$, $T_{f_2} = (0,1,0,0)$, $T_{f_3} = (0,0,1,0)$, $T_{f_4} = (0,0,0,1)$

$$\begin{aligned} w(f) &= w(g_1) + w(g_2) + w(g_3) + w(g_4), \\ S_1(f) &= 2w(g_1 \oplus g_3) + 2w(g_2 \oplus g_4), \\ S_2(f) &= 2w(g_1 \oplus g_2) + 2w(g_3 \oplus g_4), \\ S_i(f) &= S_{i-2}(g_1) + S_{i-2}(g_2) + S_{i-2}(g_3) + S_{i-2}(g_4), \quad i = 3, \ldots, n. \end{aligned}$$

b) In case $T_{f_1} = (0,0,1,1)$, $T_{f_2} = (0,1,0,1)$, $T_{f_3} = (0,0,0,1)$, $T_{f_4} = (1,1,1,1)$

$$\begin{aligned} w(f) &= w(g_4) + w(g_2 \oplus g_4) + w(g_1 \oplus g_4) + w(g_1 \oplus g_2 \oplus g_3 \oplus g_4), \\ S_1(f) &= 2w(g_1) + 2w(g_1 \oplus g_3), \\ S_2(f) &= 2w(g_2) + 2w(g_2 \oplus g_3), \end{aligned}$$

and for $i = 3, \ldots, n$

$$S_i(f) = S_{i-2}(g_4) + S_{i-2}(g_2 \oplus g_4) + S_{i-2}(g_1 \oplus g_4) + S_{i-2}(g_1 \oplus g_2 \oplus g_3 \oplus g_4).$$

From now on we deal with the properties of the difference distribution vectors of Boolean functions which will be used in the following sections.

Lemma 3.1.4. *For any $f \in \mathcal{F}_n$ we have $S(f) = S(\bar{f})$.*

Proof: It is enough to show that $S_i(f) = S_i(\bar{f})$ for a fixed $i \in \{1, 2, \ldots, n\}$.
$S_i(f) = \sum_x f(x) \oplus f(x \oplus e_i) = \sum_x (f(x) \oplus 1) \oplus (f(x \oplus e_i) \oplus 1)$
$= \sum_x \bar{f}(x) \oplus \bar{f}(x \oplus e_i) = S_i(\bar{f})$. □

In the following proposition we consider the weight of the functions whose difference distribution vectors are known.

Proposition 3.1.5. *For any $f \in \mathcal{F}_n$, $(S_{max}(f)/2) \leq w(f) \leq 2^n - (S_{max}(f)/2)$.*

Proof: Without loss of generality, we may assume that $S_{max}(f) = S_1$. Write $f = (f_1 \otimes g_1) \oplus (f_2 \otimes g_2)$, where $f_1 = (1, 1)$, $f_2 = (0, 1)$, for some $g_1, g_2 \in \mathcal{F}_{n-1}$. By Example 3.1.2 b), $S_1 = 2w(g_2)$, and $w(f) = w(g_1) + w(g_1 \oplus g_2)$. Then $w(f) = w(g_2) + 2(w(g_1) - w(g_1 g_2)) = S_{max}(f)/2 + 2k$ where $k \in \{0, 1, \ldots, 2^{n-1}\}$. We know that $w(f) + w(\bar{f}) = 2^n$, and from the above lemma, it follows that $S_{max}(f) = S_{max}(\bar{f})$. Thus, $S_{max}/2 \leq w(\bar{f})$, which implies that $w(f) \leq 2^n - S_{max}(f)/2$. □

Corollary 3.1.6. *Given $f \in \mathcal{F}_n$. If f satisfies the strict avalanche criterion, then $2^{n-2} \leq w(f) \leq 3 \cdot 2^{n-2}$.*

Proof: Since f satisfies the strict avalanche criterion, $S_i = S_{max}(f) = 2^{n-1}$ for all $i \in \{1, \ldots, n\}$. By the previous proposition $(2^{n-1}/2) \leq w(f) \leq 2^n - (2^{n-1}/2)$, that is, $2^{n-2} \leq w(f) \leq 3 \cdot 2^{n-2}$. □

Proposition 3.1.7. *Let $f \in \mathcal{F}_n$ and $i \in \{1, 2, \ldots n\}$. Then $S_i(f) = 0$ if and only if $deg(f, x_i) = 0$.*

Proof: Let $S_i(f) = 0$. Then $\sum_x f(x) \oplus f(x \oplus e_i) = S_i(f) = 0$. Consequentely $f(x) = f(x \oplus e_i)$ for all $x \in \mathcal{V}_n$, hence $deg(f, x_i) = 0$
If $deg(f, x_i) = 0$ for some $i \in \{1, \ldots, n\}$ then $f(x) = f(x \oplus e_i)$ for all $x \in \mathcal{V}_n$. Therefore, $S_i(f) = \sum_x f(x) \oplus f(x \oplus e_i) = 0$.

□

Proposition 3.1.8. *Let $f \in \mathcal{F}_n$ and $i \in \{1, 2, \ldots n\}$. Then, $S_i(f) = 2^n$ if and only if $deg(f, x_i) = 1$.*

Proof: $S_i(f) = \sum_x f(x) \oplus f(x \oplus e_i) = 2^n$ implies that $f(x \oplus e_i) = f(x) \oplus 1$ for all $x \in \mathcal{V}_n$. Equivalently, $deg(f, x_i) = 1$.
If $deg(f, x_i) = 1$, then $f(x \oplus e_i) = f(x) \oplus 1$ for all $x \in \mathcal{V}_n$. Therefore, $S_i(f) = \sum_x f(x) \oplus f(x \oplus e_i) = 2^n$. □

Corollary 3.1.9. *Let $f \in \mathcal{F}_n$. If $S_i(f) = 2^n$ for some $i \in \{1, 2, \ldots n\}$, then f is balanced.*

Proof: By proposition 3.1.5 we have $(S_{max}(f)/2) \leq w(f) \leq 2^n - (S_{max}(f)/2)$. $S_{max}(f) = 2^n$ implies $(2^{n-1} \leq w(f) \leq 2^{n-1})$, hence f is balanced. □

Theorem 3.1.10. *Let $f \in \mathcal{F}_n$. Then,*
a) $S_i \equiv 0$ mod 2 for all $i \in \{1, \ldots, n\}$.
b) $S_i \equiv 2$ mod 4 for all $i \in \{1, 2, \ldots, n\}$ if and only if $w(f)$ is odd.
c) $S_i \equiv 0$ mod 4 for all $i \in \{1, 2, \ldots, n\}$ if and only if $w(f)$ is even .

Proof:
a) Since $S_i(f) = 2w(\frac{\delta f}{\delta x_i})$ for all $i \in \{1, \ldots, n\}$, we have $S_i(f) \equiv 0$ *mod* 2.
b) It is sufficient to prove theorem for $i = 1$. Write f as $f = (f_1 \otimes g_1) \oplus (f_2 \otimes g_2)$ where $g_1, g_2, \in \mathcal{F}_{n-1}$ and $T_{f_1} = (1, 1)$, $T_{f_2} = (0, 1)$. Then, by Example 3.1.2 b), $S_1(f) = 2w(g_2)$ and $w(f) = w(g_1) + w(g_1 \oplus g_2) = 2[w(g_1) - w(g_1 g_2)] + w(g_2)$.
Now, if $S_1(f) \equiv 2$ *mod* 4, then $w(g_2)$ is odd, which implies that $w(f)$ is odd.
On the other hand, if $w(f)$ is odd then $w(g_2)$ is odd, that is $2w(g_2) \equiv 2$ *mod* 4.
c) This part is equivalent to part b). □

Theorem 3.1.11. *Let $f \in \mathcal{F}_n$. If $S_i = 2^n$ for some $i \in \{1, 2, \ldots, n\}$, then $S_j \equiv S_k \ mod\ 8$ where j, k are different from i.*

Proof: It is sufficient to prove for $i = 1$, since $[S(f(x))] = [S(f(\pi(x)))]$ for any permutation π. Write $f = ((1,0) \otimes g) \oplus ((0,1) \otimes \bar{g})$ for some $g \in \mathcal{F}_{n-1}$. By Example 3.1.2 a), we have $S_1(f) = 2^n$ and $S_j(f) = S_{j-1}(g) + S_{j-1}(\bar{g}) = 2S_{j-1}(g)$ for $j \in \{2, \ldots, n\}$. Since $S_{j-1}(g) \equiv S_{k-1}(g) \ mod\ 4$ for all $j, k \in \{2, \ldots, n\}$, $S_j(f) \equiv S_k(f) \ mod\ 8$. □

Theorem 3.1.12. *Let $f \in \mathcal{F}_n$ and $w(f) \equiv\ 0 \ mod\ 4$. If $S_i = 0$ for some $i \in \{1, 2, \ldots, n\}$, then $S_j \equiv\ 0 \ mod\ 8$ for any $j \in \{1, 2, \ldots, n\}$.*

Proof: It is sufficient to prove for $i = 1$, since $[S(f(x))] = [S(f(\pi(x)))]$ for any permutation π. Write $f(x) = ((1,0) \otimes g) \oplus ((0,1) \otimes g)$ where $g \in \mathcal{F}_{n-1}$. As in Example 3.1.2 a), we have $w(f) = 2w(g)$, $S_1(f) = 0$ and $S_k(f) = S_{k-1}(g) + S_{k-1}(g) = 2S_{k-1}(g)$ for $k \in \{2, \ldots, n\}$. Since $w(f) \equiv\ 0 \ mod\ 4$, $w(g)$ is even. Therefore by Theorem 3.1.10 c), $S_{k-1}(g) \equiv\ 0 \ mod\ 4$ for all $k \in \{2, \ldots, n\}$. Hence $S_k(f) \equiv\ 0 \ mod\ 8$ implies $S_j(f) \equiv\ 0 \ mod\ 8$ for any $j \in \{1, 2, \ldots, n\}$. □

3.2 Number of functions with a particular difference distribution vector

In this chapter, we especially deal with the numbers of functions having some particular types of difference distribution vectors. We introduce the notation $[a] = [a_1, a_2, \ldots, a_n]$ to denote the number of functions in $\mathcal{F}_n$ having difference distribution vector $a = (a_1, a_2, \ldots, a_n)$.

First we note that, for any permutation π defined on n objects, $S(f(\pi(x_1, \ldots, x_n))) = \pi(S_1, \ldots, S_n)$. Then, it follows that $[\pi a] = [a]$.

Property 3.2.1. $[a_1, a_2, \ldots, a_n] = [2^n - a_1, a_2, \ldots, a_n]$.

Proof: Let $f \in \mathcal{F}_n$ and $g, h \in \mathcal{F}_{n-1}$ such that

$f = ((1,0) \otimes g) \oplus ((0,1) \otimes h)$. Then according to the Example 3.1.2 a)

$$S_i(f) = S_{i-1}(g) + S_{i-1}(h) \text{ where } i \in \{2, \ldots, n\}$$

$$S_1(f) = 2w(g \oplus h).$$

Now define $f' \in \mathcal{F}_n$ as $f' = ((1,0) \otimes g) \oplus ((0,1) \otimes \bar{h})$. Then, $S_1(f') = 2w(g \oplus \bar{h}) = 2(2^{n-1} - w(g \oplus h)) = 2^n - 2w(g \oplus h) = 2^n - S_1(f)$ and, $S_i(f') = S_{i-1}(g) + S_{i-1}(\bar{h}) = S_i(f)$ for $i \in \{2, \ldots, n\}$. This construction associates each $f \in \mathcal{F}_n$ with a unique $f' \in \mathcal{F}_n$ such that $S_1(f') = 2^n - S_1(f)$ and $S_i(f') = S_i(f)$ for $i = 2, \ldots, n$. $\square$

The following is an immediate result of above property.

Corollary 3.2.2. *For some $f, g \in \mathcal{F}_n$, if $S_i(f) = S_i(g)$ or $S_i(f) + S_i(g) = 2^n$ for all $i \in \{1, 2, \ldots, n\}$, then $[S(f)] = [S(g)]$.*

Theorem 3.2.3. *The number of functions with $S_1 = \lambda$ is*

$$\binom{2^{n-1}}{\frac{\lambda}{2}} 2^{2^{n-1}}$$

where λ is a nonnegative even integer.

Proof: Using Example 3.1.2 a), any $f \in \mathcal{F}_n$ can be written as

$$((1,1) \otimes g_1) \oplus ((0,1) \otimes g_2)$$

for some $g_1, g_2 \in \mathcal{F}_{n-1}$. Then we have, $\lambda = S_1(f) = 2w(g_2)$. Then, it is obvious that, there are $\binom{2^{n-1}}{\frac{\lambda}{2}}$ possible choices for g_2 and $2^{2^{n-1}}$ possible choices for g_1. $\square$

By substituting $\lambda = 2^{n-1}$ in the above theorem, one obtains

Corollary 3.2.4. *$S(n,1)$ [4] is*

$$\binom{2^{n-1}}{2^{n-2}} 2^{2^{n-1}}.$$

It is also possible to find $S(n,1)$ as follows :
Let $f \in \mathcal{F}_n$ be such that $((1,1) \otimes g_1) \oplus ((0,1) \otimes g_2)$ where $g_1, g_2 \in \mathcal{F}_{n-1}$. Then,

$2^{n-1} = S_1(f) = 2w(g_2)$. After fixing $g_1 \cdot g_2$, the number of ways of choosing the functions g_1 and g_2 satisfying the above condition is

$$\sum_{w(g_1 \cdot g_2)=0}^{2^{n-2}} \binom{2^{n-1}}{w(g_1 \cdot g_2)} \binom{2^{n-1} - w(g_1 \cdot g_2)}{2^{n-2} - w(g_1 \cdot g_2)} 2^{2^{n-2}}$$

or that is,

$$S(n,1) = \sum_{k}^{2^{n-2}} \binom{2^{n-1}}{k} \binom{2^{n-1} - k}{2^{n-2}} 2^{2^{n-2}}.$$

The number of balanced Boolean functions in $\mathcal{F}_n$ is $\binom{2^n}{2^{n-1}}$. Now we obtain this result following an indirect way which will enable us to count the balanced Boolean functions satisfying certain conditions.

Let $f \in \mathcal{F}_n$ be such that $((1,1) \otimes g_1) \oplus ((0,1) \otimes g_2)$ where $g_1, g_2 \in \mathcal{F}_{n-1}$. Then, it follows that $w(f) = w(g_1) + w(g_1 \oplus g_2)$. After fixing g_1, the number of ways of choosing g_2 satisfying the above condition is

$$\sum_{w(g_1 g_2)=0}^{w(g_1)} \binom{w(g_1)}{w(g_1 g_2)} \binom{2^{n-1} - w(g_1)}{w(g_1) - w(g_1 g_2)}$$

$w(g_1)$ can take any value between 0 and 2^{n-1}, and for each $k \in \{0, \ldots, 2^{n-1}\}$ there are $\binom{2^{n-1}}{k}$ functions with $w(g_1) = k$. Hence the number of balanced functions is

$$\binom{2^n}{2^{n-1}} = \sum_{k=0}^{2^{n-1}} \binom{2^{n-1}}{k} \sum_{l=0}^{k} \binom{k}{l} \binom{2^{n-1} - k}{k - l}.$$

Theorem 3.2.5. *The number of balanced functions in $\mathcal{F}_n$ with the property $S_1 = \lambda$ is*

$$\sum_{k=0}^{2^{n-1}} \binom{2^{n-1}}{k} \binom{k}{2^{n-2} - \frac{\lambda}{4}} \binom{2^{n-1} - k}{2^{n-2} - \frac{\lambda}{4}}$$

where λ is a nonnegative integer, divisible by 4.

Proof: Given $f \in \mathcal{F}_n$, the number of balanced Boolean functions is

$$\sum_{w(g_1)=0}^{2^{n-1}} \binom{2^{n-1}}{w(g_1)} \sum_{w(g_1g_2)=0}^{w(g_1)} \binom{w(g_1)}{w(g_1g_2)} \binom{2^{n-1}-w(g_1)}{w(g_1)-w(g_1g_2)}$$

Now $S_1 = 2w(g_2)$ implies that $w(g_2) = \frac{\lambda}{2}$. Since f is balanced, we have $w(f) = 2^{n-1} = w(g_1)+w(g_1 \oplus g_2) = w(g_2)+2w(g_1)-2w(g_1g_2) = \frac{\lambda}{2}+2w(g_1)-2w(g_1g_2)$ So we again have the same formula as above, but now $w(g_1g_2) = w(g_1) + \frac{\lambda}{4} - 2^{n-2}$. Then the sought number is

$$\sum_{w(g_1)=0}^{2^{n-1}} \binom{2^{n-1}}{w(g_1)} \binom{w(g_1)}{w(g_1g_2)} \binom{2^{n-1}-w(g_1)}{w(g_1)-w(g_1g_2)}$$

$$= \sum_{w(g_1)=0}^{2^{n-1}} \binom{2^{n-1}}{w(g_1)} \binom{w(g_1)}{w(g_1)+\frac{\lambda}{4}-2^{n-2}} \binom{2^{n-1}-w(g_1)}{w(g_1)-(w(g_1)+\frac{\lambda}{4}-2^{n-2})}$$

$$= \sum_{w(g_1)=0}^{2^{n-1}} \binom{2^{n-1}}{w(g_1)} \binom{w(g_1)}{2^{n-2}-\frac{\lambda}{4}} \binom{2^{n-1}-w(g_1)}{2^{n-2}-\frac{\lambda}{4}}.$$

□

We denote by $SB(n,k)$ the number of balanced functions counted in $S(n,k)$. And letting $\lambda = 2^{n-1}$ in the above theorem, the following is obtained.

Theorem 3.2.6. *For any integer $n > 1$,*

$$SB(n,1) = \sum_{k=0}^{2^{n-1}} \binom{2^{n-1}}{k} \binom{k}{2^{n-3}} \binom{2^{n-1}-k}{2^{n-3}}$$

In [4], a computation of $S(n,2)$ is given by

$$S(n,2) = \sum_{i=0}^{2^{n-3}} \left(\begin{array}{c} 2^{n-2} \\ 2i \end{array} \right) 8^{2^{n-2}-2i} 2^{2i} \sum_{j=0}^{i} \left(\begin{array}{c} 2i \\ 2j \end{array} \right) \left(\begin{array}{c} 2j \\ j \end{array} \right) \left(\begin{array}{c} 2i-2j \\ i-j \end{array} \right).$$

We give a more general, but yet more simple formula which computes not only $S(n,2)$ but also the number of functions $f \in \mathcal{F}_n$ with $S_1(f) = \lambda_1$, $S_2(f) =$

λ_2 for arbitrary choosen nonnegative even integers λ_1 , λ_2.

Theorem 3.2.7. *Given nonnegative even integers $\lambda_1, \lambda_2 \leq 2^n$ with $\lambda_1 \equiv \lambda_2 \ mod\ 4$. Then, the number of functions in $\mathcal{F}_n$ such that $S_1 = \lambda_1$, $S_2 = \lambda_2$ is given by*

$$2^{2^{n-2}} \sum_{t=0}^{2^{n-2}} \binom{2^{n-2}}{t} \binom{2^{n-2}-t}{\frac{\lambda_1-2t}{4}} \binom{2^{n-2}-t}{\frac{\lambda_2-2t}{4}} 2^{2t}$$

where t ranges over only odd or even integers depending on whether $\lambda_1 \equiv \lambda_2 \equiv 2 \ mod\ 4$ or, $\lambda_1 \equiv \lambda_2 \equiv 0 \ mod\ 4$, respectively.

Proof: Any $f \in \mathcal{F}_n$ can be written as

$$((0,0,1,1) \otimes g_1) \oplus ((0,1,0,1) \otimes g_2) \oplus ((0,0,0,1) \otimes g_3) \oplus ((1,1,1,1) \otimes g_4)$$

where $g_1, g_2, g_3, g_4 \in F_{n-2}$. Then by Example 3.1.3 b), it follows that $S_1(f) = 2(w(g_1) + w(g_1 \oplus g_3))$, $S_2 = 2(w(g_2) + w(g_2 \oplus g_3))$, we have

$$\lambda_1 = 4w(g_1) + 2w(g_3) - 4w(g_1 g_3)$$

$$\lambda_2 = 4w(g_2) + 2w(g_3) - 4w(g_2 g_3)$$

After fixing g_3, the number of ways of choosing g_1 and g_2 satisfying above conditions are $\binom{2^{n-2}-w(g_3)}{\frac{\lambda_1-2w(g_3)}{4}} \cdot 2^{w(g_3)}$ and $\binom{2^{n-2}-w(g_3)}{\frac{\lambda_2-2w(g_3)}{4}} \cdot 2^{w(g_3)}$, respectively. Weight of g_3 can take any value between 0 and 2^{n-2}, and g_4 can be choosen arbitrarily. So, the number of possible ways of constructing $f \in \mathcal{F}_n$ with $S_1(f) = \lambda_1$, $S_2(f) = \lambda_2$ is

$$2^{2^{n-2}} \sum_{t=0}^{2^{n-2}} \binom{2^{n-2}}{t} \binom{2^{n-2}-t}{\frac{\lambda_1-2t}{4}} \binom{2^{n-2}-t}{\frac{\lambda_2-2t}{4}} 2^{2t}.$$

□

An immediate consequence of this theorem is

Corollary 3.2.8.

$$S(n,2) = 2^{2^{n-2}} \sum_{t=0}^{2^{n-2}} \binom{2^{n-2}}{t} \binom{2^{n-2}-t}{\frac{2^{n-1}-2t}{4}}^2 2^{2t}$$

where t ranges over only even integers.

Chapter 4

Constructions

In this chapter we focus on the construction of Boolean functions satisfying strict avalanche criterion.

4.1 Constructions Using Kronecker Product

Theorem 4.1.1. *Given $g_1, g_2 \in \mathcal{F}_{n-2}$ such that $S_i(g_1) + S_i(g_2) = 2^{n-2}$ for all $i \in \{1, 2, \ldots, n-2\}$, then $f \in \mathcal{F}_n$ defined by*

$$f = ((1,0,0,0) \otimes g_1) \oplus ((0,1,0,0) \otimes g_1) \oplus \ ((0,0,1,0) \otimes g_2) \oplus ((0,0,0,1) \otimes \bar{g_2})$$

satisfies strict avalanche criterion.

Proof: By Example 3.1.3 a) we obtain,

$$\begin{aligned} S_1(f) &= 2[w(g_1 \oplus g_2) + w(g_1 \oplus \bar{g_2})] = 2[w(g_1 \oplus g_2) + (2^{n-2} - w(g_1 \oplus g_2))] \\ &= 2^{n-1}, \\ S_2(f) &= 2[w(g_1 \oplus g_1) + w(g_2 \oplus \bar{g_2})] = 2 \cdot 2^{n-2} = 2^{n-1}, \\ S_i(f) &= 2S_{i-2}(g_1) + S_{i-2}(g_2) + S_{i-2}(\bar{g_2}) = 2^{n-1}, \ \ i = 3, \ldots, n. \end{aligned}$$

□

Corollary 4.1.2. *Given a balanced function $g \in \mathcal{F}_{n-2}$ satisfying strict avalanche criterion, then $f \in \mathcal{F}_n$ defined by*

$$f = ((1,0,0,0) \otimes g) \oplus ((0,1,0,0) \otimes g) \oplus ((0,0,1,0) \otimes g) \oplus ((0,0,0,1) \otimes \bar{g})$$

is also a balanced function satisfying strict avalanche criterion.

Proof: According to the Example 3.1.3 a) $w(f) = w(g)+w(g)+w(g)+w(\bar{g})$. Therefore, $w(f) = 2^{n-1}$ since $w(g) = w(\bar{g}) = 2^{n-3}$. That is f is balanced. On the other hand, since $S_i(g) + S_i(\bar{g}) = 2^{n-2}$ for all $i \in \{1, 2, \ldots, n-2\}$, then by above theorem f satisfies strict avalanche criterion. □

Example 4.1.3. Let $T_g = (1,0,1,0,1,1,0,0)$. Then $g \in \mathcal{F}_3$ is a balanced function with difference distribution vector $S(g) = (4,4,4)$. $f \in \mathcal{F}_5$ defined by above theorem has a truth table

$$T_f = (\mathbf{1,0,1,0,1,1,0,0}, 1,0,1,0,1,1,0,0, \mathbf{1,0,1,0,1,1,0,0}, 0,1,0,1,0,0,1,1)$$

is also a balanced function satisfying strict avalanche criterion.

Proposition 4.1.4. *Let $f \in \mathcal{F}_n$. If $deg(f, x_i) = 2$ for all $i \in \{1, 2, \ldots, n\}$, then f satisfies strict avalanche criterion.*

Proof: For any $x_i \in \{x_1, \ldots, x_n\}$ we have $deg(\frac{\partial f}{\partial x_i}) = 1$. Thus $\frac{\partial f}{\partial x_i}$ is a nonconstant affine function, hence balanced. This completes the proof. □

Proposition 4.1.5. *Number of quadratic functions satisfying strict avalanche criterion is at least $2^{n+1} \sum_{k=0}^{n+1} (-1)^k \binom{n}{k} 2^{\binom{n-k}{2}}$.*

Proof: Number of functions with $deg(f) \leq 2$ and $deg(f, x_{i_1}), \ldots, deg(f, x_{i_k}) \leq 1$, for some fixed $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ is $2^{\binom{n-k}{2}+n+1}$. Since $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ can be chosen in $\binom{n}{k}$ distinct ways, the result directly follows from principle of inclusion-exculsion. □

Theorem 4.1.6. *Let $f \in \mathcal{F}_n$, $g \in \mathcal{F}_m$. If f, g satisfy strict avalanche criterion then $h \in \mathcal{F}_{n+m}$ defined by $(1_m \otimes f) \oplus (g \otimes 1_n)$ also satisfies strict avalanche criterion.*

Proof: $h(x) = (1_m \otimes f) \oplus (g \otimes 1_n)$ where $x = (y, z)$ with $y = x_1, \ldots, x_m$, $z = x_{m+1}, \ldots, x_{n+m}$. Then for $i = \{1, \ldots, m\}$,

$S_i(h) = \sum_x h(x) \oplus h(x \oplus e_i)$

$= \sum_x [(1_m \otimes f) \oplus (g \otimes 1_n)](x) \oplus [(1_m \otimes f) \oplus (g \otimes 1_n)](x \oplus e_i)$

$= \sum_x [(1_m(y) \cdot f(z)) \oplus (1_m(y \oplus e_i) \cdot f(z))] \oplus [(g(y) \cdot 1_n(z)) \oplus (g(y \oplus e_i) \cdot 1_n(z))]$

$= \sum_x [1_m(y) \oplus (1_m(y \oplus e_i))] \cdot f(z) \oplus [(g(y) \oplus g(y \oplus e_i))] \cdot 1_n(z)$

$= \sum_x g(y) \oplus g(y \oplus e_i) = 2^n \sum_y g(y) \oplus g(y \oplus e_i) = 2^n \cdot 2^{m-1} = 2^{n+m-1}$.

For $i = \{m+1, \ldots, m+n\}$,

$S_i(h) = \sum_x h(x) \oplus h(x \oplus e_i)$

$= \sum_x [(1_m(y) \cdot f(z)) \oplus (1_m(y) \cdot f(z \oplus e_i))] \oplus [(g(y) \cdot 1_n(z)) \oplus (g(y) \cdot 1_n(z \oplus e_i))]$

$= \sum_x 1_m(y)[f(z) \oplus f(z \oplus e_i)] \oplus g(y)[1_n(z) \oplus 1_n(z \oplus e_i)]$

$= \sum_x f(z) \oplus f(z \oplus e_i) = 2^m \sum_z f(z) \oplus f(z \oplus e_i) = 2^m \cdot 2^{n-1} = 2^{n+m-1}$. □

Example 4.1.7. Let $f \in \mathcal{F}_3$ and $g \in \mathcal{F}_2$ satisfy strict avalanche criterion, and have the truth tables $T_f = (0,1,0,1,0,0,1,1)$, $T_g = (1,0,0,0)$. Then, $F = (1_2 \otimes f) \oplus (g \otimes 1_3)$, having the truth table

$$T_F = (\mathbf{1,0,1,0,1,1,0,0}, 0,1,0,1,0,0,1,1, \mathbf{0,1,0,1,0,0,1,1}, 0,1,0,1,0,0,1,1)$$

also satisfies strict avalanche criterion.

Corollary 4.1.8. *Let $f \in \mathcal{F}_n$. If f satisfies strict avalanche criterion, then $(f \otimes \bar{f}) \oplus (\bar{f} \otimes f)$ also satisfies strict avalanche criterion.*

Proof: The proof follows from the above theorem, and the fact that $(f \otimes \bar{f}) \oplus (\bar{f} \otimes f) = (1_n \otimes f) \oplus (f \otimes 1_n)$. □

4.2 SAC and Nonlinearity

Now we give two constructions defined in [7] by Jennifer Seberry, Xian-Mo Zhang. This constructions yield SAC satisfying bent functions [1]. We restate the theorems and give an alternative notation.

Theorem 4.2.1. *Let $g_1, g_2 \in \mathcal{F}_{2k}$ such that g_1 is a bent function, and g_2 is a non-constant affine function, then $f \in \mathcal{F}_{2k+1}$ defined by*

$$f = ((1,0) \otimes g_1) \oplus ((0,1) \otimes (g_1 \oplus g_2))$$

is a balanced function satisfiying strict avalanche criterion.

Proof: Let $g_1 \in \mathcal{F}_{2k}$ be a bent function, then without loss of generality we may assume that $w(g_1) = 2^{2k-1} - 2^{k-1}$. It is known that $(g_1 \oplus g_2)$ is also a bent for any non-constant affine function g_2. Therefore, we have $w(g_1 \oplus g_2) = 2^{2k-1} + 2^{k-1}$. Then by using Example 3.1.2 a), $w(f) = w(g_1) + w(g_1 \oplus g_2) = 2^{2k}$. Hence, f is balanced.
We complete the proof showing that $S_i(f) = 2^{2k}$ for all $i \in \{1, \ldots, 2k+1\}$. By using Example 3.1.2 a), it follows that

$$S_1(f) = 2w(g_2) = 2 \cdot 2^{2k-1} = 2^{2k},$$

$$S_i(f) = S_{i-1}(g_1) + S_{i-1}(g_1 \oplus g_2) = 2^{2k-1} + 2^{2k-1} = 2^{2k} \ for \ i \in \{2, \ldots, 2k+1\}.$$

□

Theorem 4.2.2. *Let $g, h_1, h_2, h_3 \in \mathcal{F}_{2k-2}$ such that g is a bent function, and h_1, h_2, h_3 are non-constant affine functions, then $f \in \mathcal{F}_{2k}$ defined by*

$$f = [(1,0,0,0) \otimes g] \oplus [(0,1,0,0) \otimes (g \oplus h_1)] \oplus [(0,0,1,0) \otimes (g \oplus h_2)] \oplus [(0,0,0,1) \otimes (g \oplus h_3)]$$

is a balanced function satisfiying strict avalanche criterion .

[1] In this work we are concentrated on the strict avalanche criterion, for definitions concerning nonlinearity and bent functions reader may see [6]

Proof: Let $g \in \mathcal{F}_{2k-2}$ be a bent function. Since $(g \oplus h_1)$, $(g \oplus h_2)$ and $(g \oplus h_1)$ are also bent functions for any non-constant affine functions $h_1, h_2, h_3 \in \mathcal{F}_{2k-2}$, without loss of generality, we can assume that $w(g) = w(g \oplus h_1) = 2^{2k-3} - 2^{k-2}$ and $w(g \oplus h_2) = w(g \oplus h_3) = 2^{2k-3} + 2^{k-2}$. Note that by Example 3.1.3 a), we have $w(f) = 2^{2k-1}$. So f is balanced.
Applying the results in Example 3.1.3 a), we have

$$S_1(f) = 2w(h_2) + 2w(h_1 \oplus h_3) = 2^{2k-1},$$

$$S_2(f) = 2w(h_1) + 2w(h_2 \oplus h_3) = 2^{2k-1},$$

$S_i(f) = S_{i-2}(g)+S_{i-2}(g\oplus h_1)+S_{i-2}(g\oplus h_2)+S_{i-2}(g\oplus h_3) = 2^{2k-1}$ *for* $i \in \{3, \ldots, 2k\}$.

□

Lemma 4.2.3. *Let* $f, g \in \mathcal{F}_n$ *and* $x_i \in \mathcal{A}_n$. *If* $f(x) = g(x) \oplus x_i$, *then* $S_i(f) = 2^n - S_i(g)$.

Proof: If $f(x) = g(x) \oplus x_i$, then $S_i(f) = \sum_x (g(x) \oplus x_i) \oplus (g(x \oplus e_i) \oplus x_i \oplus 1) = \sum_x g(x) \oplus g(x \oplus e_i) \oplus 1$. Hence, $S_i(f) = 2^n - S_i(g)$. □

Theorem 4.2.4. *Let* $f, g \in \mathcal{F}_n$. *If* g *satisfies strict avalanche criterion, then* f *defined by* $f(x) = g(x) \oplus x_i$ *for any* $i \in \{1, 2, \ldots, n\}$ *also satisfies strict avalanche criterion.*

Proof: Since g satisfies strict avalanche criterion we have $S_i(g) = 2^{n-1}$ and by above lemma $S_i(f) = 2^n - S_i(g)$. Thus $S_i(f) = 2^{n-1}$. □

Theorem 4.2.5. *Let* $g \in \mathcal{F}_{n-1}$. *The function* $f \in \mathcal{F}_n$ *defined by*

$$f = ((1, 0) \otimes g) \oplus ((0, 1) \otimes h)$$

where $h = g \oplus x_1 \oplus x_2 \cdots \oplus x_{n-1} \oplus c$ *for* $c \in \{0, 1\}$, *satisfies strict avalanche criterion.*

Proof: By Example 3.1.2 a) we obtain,

$$S_1(f) = 2w(g \oplus h) = 2w(x_1 \oplus x_2 \cdots \oplus x_{n-1} \oplus c) = 2^{n-1},$$

$$S_i(f) = S_{i-1}(g) + S_{i-1}(h) \ for \ i \in \{2, \ldots, n\}.$$

Also, by using above lemma, it is easy to verify that
$S_{i-1}(g) + S_{i-1}(h) = S_{i-1}(g) + 2^{n-1} - S_{i-1}(g) = 2^{n-1} \ for \ i \in \{2, \ldots, n\}.$ □

Chapter 5

Statistical Observations

There is no explicit formula to compute $S(n,k)$ for $k \geq 3$ and to compute $[a_1, \ldots, a_n]$ is even a much more difficult task for arbitrary integers $a_1, \ldots, a_n$. For the cases $n = 1, 2, 3, 4, 5$, since the number of all Boolean functions remain in a reasonable range, computations can be performed by direct counting. For these cases we give tables showing all related information about the difference distribution table. For $n = 5, 6, 7$ and 8 we present some statistical results.

5.1 Table of Difference Distribution Vectors for $n \leq 5$

In the below tables, the first column shows the difference distribution vectors; the first row shows the weights of the functions. In the other boxes, we give the number of functions having the corresponding difference distribution vector in the first column and with the weight in the first row. For example, consider the following row. The bold face **4** indicates the number of functions of weight 2 and having the difference distribution vector $0, 4, 4$ or any of its permutations.

	0	1	2	3	4	5	6	7	8
0 4 4			**4**				4		

a-) **n=1**

	0	1	2
0	1		1
2		2	

Table 5.1: Number of Difference Distribution Vectors for n=1

b-) **n=2**

	0	1	2	3	4
0 0	1				1
2 2		4		4	
0 4			2		
4 4			2		

Table 5.2: Number of Difference Distribution Vectors for n=2

c-) **n=3**

	0	1	2	3	4	5	6	7	8
0 0 0	1								1
2 2 2		8						8	
0 4 4			4				4		
4 4 4			16		32		16		
2 2 6				8		8			
2 6 6				8		8			
6 6 6				8		8			
0 0 8					2				
4 4 8					8				
0 8 8					2				
8 8 8					2				

Table 5.3: Number of Difference Distribution Vectors for n=3

d-) **n=4**
For $n = 4$, using the facts in chapter 3.2., we give only the half of the table. In the following example, the bold face **112** shows the number of functions of weight 4 or 12 having the difference distrubition vector $(04, 04, 08, 08)$ or one of its permutations.

	0-16	1-15	2-14	3-13	4-12	5-11	6-10	7-9	8
04 04 08 08					**112**		96		32

	0-16	1-15	2-14	3-13	4-12	5-11	6-10	7-9	8
00 00 00 00	1								
02 02 02 02		16							
00 04 04 04			8						
04 04 04 04			88						
02 02 06 06				16					
02 06 06 06				64					
06 06 06 06				208		400		48	
04 04 04 08					64				
00 00 08 08					4				
04 04 08 08					112		96		32
00 08 08 08					16				32
04 08 08 08					144		288		288
08 08 08 08					228		1152		1368
02 02 06 10						16			
02 06 06 10						16		48	
06 06 06 10						320		336	
02 02 10 10						16			
02 06 10 10						32		32	
06 06 10 10						176		480	
02 10 10 10						48		16	
06 10 10 10						112		544	
10 10 10 10						128		528	
00 04 04 12							8		
04 04 04 12							32		112
04 04 08 12							48		32
04 08 08 12							128		192
08 08 08 12							288		576
00 04 12 12							8		
04 04 12 12							40		96
04 08 12 12							32		64
08 08 12 12							96		256
00 12 12 12							8		
04 12 12 12							48		80
08 12 12 12							16		96
12 12 12 12							56		64

	0-16	1-15	2-14	3-13	4-12	5-11	6-10	7-9	8
02 02 02 14								16	
02 06 06 14								16	
06 06 06 14								64	
02 06 10 14								16	
06 06 10 14								64	
02 10 10 14								16	
06 10 10 14								64	
10 10 10 14								64	
02 02 14 14								16	
06 06 14 14								16	
06 10 14 14								16	
10 10 14 14								16	
02 14 14 14								16	
14 14 14 14								16	
00 00 00 16									2
04 04 04 16									16
00 08 08 16									8
08 08 08 16									64
04 04 12 16									16
04 12 12 16									16
12 12 12 16									16
00 00 16 16									2
08 08 16 16									8
00 16 16 16									2
16 16 16 16									2

Table 5.4: Number of Difference Distribution Vectors for n=4

e-) **n=5**

For the case $n = 5$, since the table becomes very large, we give it without considering weights of the functions. In the following table, the number of functions with a difference distribution vector in the indicated class is given.

16 16 16 16 16	27522560	08 08 12 12 16	145024	04 08 12 12 16	8448
12 16 16 16 16	14528512	08 08 12 12 12	135296	02 14 14 14 14	8256
14 14 14 14 14	13062656	06 10 10 14 14	81152	04 08 08 16 16	8192
12 12 16 16 16	8563712	08 08 08 16 16	73536	06 06 10 10 14	7232
12 12 12 16 16	4920576	04 12 16 16 16	68096	06 06 06 14 14	6400
10 14 14 14 14	4205056	08 08 08 12 12	56512	04 08 08 08 08	6016
12 12 12 12 16	2959104	06 10 10 10 10	46080	06 06 06 06 06	5760
08 16 16 16 16	2681344	04 12 12 16 16	43264	06 06 06 10 10	5312
12 12 12 12 12	2055296	06 10 10 10 14	40256	04 08 08 12 12	4608
10 10 14 14 14	1510144	08 08 08 12 16	39296	04 08 08 12 16	4352
08 12 16 16 16	1456640	08 08 08 08 08	37344	00 16 16 16 16	4128
08 12 12 16 16	947712	04 12 12 12 12	33472	02 10 14 14 14	3456
08 12 12 12 16	607104	06 06 14 14 14	32768	02 10 10 10 10	2496
10 10 10 14 14	594368	04 12 12 12 16	32320	04 04 12 16 16	2304
08 12 12 12 12	453248	08 08 08 08 12	29184	02 10 10 14 14	2176
06 14 14 14 14	404992	04 08 16 16 16	26112	04 04 12 12 12	1984
08 08 16 16 16	391808	06 06 10 10 10	15360	06 06 06 06 10	1600
10 10 10 10 14	292288	04 08 12 16 16	15104	04 04 08 08 08	1472
10 10 10 10 10	261824	06 06 10 14 14	11968	06 06 06 10 14	1344
08 08 12 16 16	217600	04 08 12 12 12	10176	00 12 12 12 12	1312
06 10 14 14 14	161152	04 04 16 16 16	9216	04 04 04 16 16	1280
04 16 16 16 16	149504	08 08 08 08 16	8928	00 08 16 16 16	1152
02 06 14 14 14	1152	04 04 08 08 12	384	04 04 08 12 16	128
04 08 08 08 12	1152	04 04 12 12 16	384	00 00 16 16 16	64
04 08 08 08 16	1152	02 02 10 10 10	256	02 02 02 02 02	64
02 10 10 10 14	1088	02 02 14 14 14	256	02 02 02 14 14	64
04 04 08 12 12	960	04 04 04 08 08	256	02 02 06 06 06	64
04 04 04 04 04	832	02 06 06 10 14	192	02 02 06 10 10	64
02 06 06 06 06	704	06 06 06 06 14	192	02 06 06 10 10	64
02 06 10 10 10	512	00 08 08 08 08	176	00 04 04 04 04	32
02 06 10 14 14	512	00 04 12 12 12	128	00 04 04 12 12	32
04 04 08 16 16	512	00 08 08 08 16	128	00 00 08 08 08	16
00 08 08 16 16	448	02 06 10 10 14	128	00 00 00 16 16	8
04 04 04 12 16	448	04 04 04 12 12	128	00 00 00 00 00	2
02 06 06 14 14	384	04 04 08 08 16	128		

Table 5.5: Number of Difference Distribution Vectors for n=5

5.2 Table of Difference Distribution Vectors for $n = 5, 6, 7$ *and* 8

To have some idea about $[a_1, \ldots, a_n]$ for $n = 5, 6, 7, 8$ a series of statistics have been performed. A complete table for $n = 5$ is presented in the previous section. But, to verify the method, by comparing the actual and statistical results, the same statistics is also performed for $n = 5$. Hence, we used the case $n = 5$ as the control set of the statistics.

For each case, a number of data sets is used for statistics. Below table gives the size of data set and also the number of functions in each data set.

n	Size of data set	Number of functions in each data set
5	40	10^6
6	40	10^7
7	40	10^8
8	30	10^9

Table 5.6: Size and Cardinality of Data Sets

We explain the method for $n = 8$, the other cases are similar.

For each data set, 1.000.000.000 random functions are generated allowing repitations. For each of these functions the difference distribution vector is computed and the occurence number of each particular difference distribution vector is registered. Facts stated in Section 3.2 are used to collect certain combinations in a class. Namely, all permutations of a certain combination are regarded as a single class. Moreover, the combinations $[a_1, \ldots, a_8]$ and $[b_1, \ldots, b_8]$ are counted in the same class if $a_i = b_i$ or $a_i = 256 - b_i$ for $i = 1, \ldots, 8$. Then, respecting these gatherings, the difference distribution vectors are ordered up to their frequencies. For example, suppose that the combination

$$(112, 112, 120, 120, 124, 128, 128, 128)$$

is observed $325,185$ times for 1.000.000.000 functions in a data set. This vector

represents $\frac{8!}{3!2!2!}2^5 = 960$ difference distribution vectors, that is $\frac{8!}{3!2!2!}$ for distinct permutations and 2^5 for the cases a_i and $256 - a_i$. So we obtain the number $\frac{325.185}{960} \approx 338$ as occurence number of any vector in the same class with $(112, 112, 120, 120, 124, 128, 128, 128)$. Then, for example the probability of having the difference distribution vector $(112, 112, 120, 120, 124, 128, 128, 128)$ or $(112, 128, 120, 124, 120, 128, 128, 112)$ or $(128, 112, 132, 128, 136, 112, 128, 120)$ etc. is expected to be close to 338 out of 1.000.000.000.

Since the number of all difference distribution vectors is very large, we considered only those with $100 \leq S_i \leq 156$. For each class, the associated probability is in fact the probability of any difference distribution vector in that class. For each class, the 30 sample probability values obtained from 30 data sets are used to estimate the actual probability and an interval of confidence (with $\alpha = 0.001$) is given.

Following tables are the statistical results we obtained.

n=5	Difference Distribution Vector	Probability	Interval of Confidence
1	016 016 016 016 016	0.006421	0.006403 - 0.006438
2	012 016 016 016 016	0.003385	0.003381 - 0.003389
3	014 014 014 014 014	0.003041	0.003038 - 0.003043
4	012 012 016 016 016	0.001993	0.001991 - 0.001995
5	012 012 012 016 016	0.001144	0.001144 - 0.001145
6	010 014 014 014 014	0.000979	0.000979 - 0.000980
7	012 012 012 012 016	0.000690	0.000689 - 0.000690
8	008 016 016 016 016	0.000625	0.000623 - 0.000627
9	012 012 012 012 012	0.000479	0.000478 - 0.000480
10	010 010 014 014 014	0.000352	0.000352 - 0.000352

Table 5.7: Statistical Results for n=5

n=6	Difference Distribution Vector	Probability	Interval of Confidence
1	032 032 032 032 032 032	0.000301	0.000299 - 0.000302
2	028 032 032 032 032 032	0.000226	0.000225 - 0.000226
3	030 030 030 030 030 030	0.000197	0.000197 - 0.000197
4	028 028 032 032 032 032	0.000171	0.000171 - 0.000172
5	028 028 028 032 032 032	0.000131	0.000130 - 0.000131
6	026 030 030 030 030 030	0.000114	0.000114 - 0.000114
7	028 028 028 028 032 032	0.000100	0.000100 - 0.000100
8	024 032 032 032 032 032	0.000097	0.000097 - 0.000098
9	028 028 028 028 028 032	0.000077	0.000077 - 0.000077
10	024 028 032 032 032 032	0.000075	0.000075 - 0.000075

Table 5.8: Statistical Results for n=6

n=7	Difference Distribution Vector	Probability	Interval of Confidence
1	064 064 064 064 064 064 064	7.10E-06	7.03E-06 - 7.17E-06
2	060 064 064 064 064 064 064	6.24E-06	6.22E-06 - 6.26E-06
3	062 062 062 062 062 062 062	5.65E-06	5.64E-06 - 5.65E-06
4	060 060 064 064 064 064 064	5.46E-06	5.45E-06 - 5.46E-06
5	060 060 060 064 064 064 064	4.78E-06	4.77E-06 - 4.78E-06
6	058 062 062 062 062 062 062	4.32E-06	4.32E-06 - 4.32E-06
7	060 060 060 060 064 064 064	4.18E-06	4.18E-06 - 4.18E-06
8	056 064 064 064 064 064 064	4.17E-06	4.16E-06 - 4.19E-06
9	060 060 060 060 060 064 064	3.66E-06	3.66E-06 - 3.66E-06
10	056 060 064 064 064 064 064	3.65E-06	3.65E-06 - 3.65E-06

Table 5.9: Statistical Results for n=7

n=8	Difference Distribution Vector	Probability	Interval of Confidence
1	128 128 128 128 128 128 128 128	8.47E-08	8.22E-08 - 8.72E-08
2	124 128 128 128 128 128 128 128	8.02E-08	7.90E-08 - 8.13E-08
3	126 126 126 126 126 126 126 126	7.51E-08	7.49E-08 - 7.53E-08
4	124 124 128 128 128 128 128 128	7.49E-08	7.46E-08 - 7.52E-08
5	124 124 124 128 128 128 128 128	7.05E-08	7.04E-08 - 7.06E-08
6	124 124 124 124 128 128 128 128	6.60E-08	6.59E-08 - 6.61E-08
7	122 126 126 126 126 126 126 126	6.60E-08	6.59E-08 - 6.60E-08
8	120 128 128 128 128 128 128 128	6.59E-08	6.54E-08 - 6.64E-08
9	124 124 124 124 124 128 128 128	6.18E-08	6.18E-08 - 6.19E-08
10	120 124 128 128 128 128 128 128	6.18E-08	6.17E-08 - 6.20E-08

Table 5.10: Statistical Results for n=8

n=5	Difference Distribution Vector (Balanced Functions)	Probability	Interval of Confidence
1	016 016 016 016 016	0.011545	0.011518 - 0.011571
2	016 016 016 016 020	0.007284	0.007275 - 0.007293
3	012 016 016 016 016	0.005062	0.005054 - 0.005070
4	016 016 016 020 020	0.004809	0.004804 - 0.004815
5	012 016 016 016 020	0.003798	0.003794 - 0.003801
6	016 016 020 020 020	0.003004	0.002999 - 0.003009
7	012 016 016 020 020	0.002530	0.002528 - 0.002533
8	012 012 016 016 016	0.002460	0.002456 - 0.002464
9	012 012 016 016 020	0.001948	0.001946 - 0.001950
10	016 020 020 020 020	0.001903	0.001897 - 0.001908

Table 5.11: Statistical Results of Balanced Functions for n=5

n=6	Difference Distribution Vector (Balanced Functions)	Probability	Interval of Confidence
1	032 032 032 032 032 032	0.000555	0.000552 - 0.000557
2	032 032 032 032 032 036	0.000456	0.000455 - 0.000457
3	028 032 032 032 032 032	0.000391	0.000390 - 0.000392
4	032 032 032 032 036 036	0.000369	0.000369 - 0.000370
5	028 032 032 032 032 036	0.000325	0.000325 - 0.000325
6	032 032 032 036 036 036	0.000298	0.000298 - 0.000299
7	028 028 032 032 032 032	0.000273	0.000273 - 0.000273
8	028 032 032 032 036 036	0.000267	0.000267 - 0.000267
9	032 032 036 036 036 036	0.000240	0.000239 - 0.000240
10	028 028 032 032 032 036	0.000231	0.000231 - 0.000231

Table 5.12: Statistical Results of Balanced Functions for n=6

n=7	Difference Distribution Vector (Balanced Functions)	Probability	Interval of Confidence
1	064 064 064 064 064 064 064	1.38E-05	1.37E-05 - 1.38E-05
2	064 064 064 064 064 064 068	1.24E-05	1.24E-05 - 1.25E-05
3	060 064 064 064 064 064 064	1.16E-05	1.16E-05 - 1.16E-05
4	064 064 064 064 064 068 068	1.12E-05	1.12E-05 - 1.13E-05
5	060 064 064 064 064 064 068	1.05E-05	1.05E-05 - 1.05E-05
6	064 064 064 064 068 068 068	1.02E-05	1.02E-05 - 1.02E-05
7	060 060 064 064 064 064 064	9.75E-06	9.73E-06 - 9.77E-06
8	060 064 064 064 064 068 068	9.54E-06	9.53E-06 - 9.55E-06
9	064 064 064 068 068 068 068	9.14E-06	9.13E-06 - 9.15E-06
10	060 060 064 064 064 064 068	8.90E-06	8.89E-06 - 8.91E-06

Table 5.13: Statistical Results of Balanced Functions for n=7

n=8	Difference Distribution Vector (Balanced Functions)	Probability	Interval of Confidence
1	128 128 128 128 128 128 128 128	1.67E-07	1.63E-07 - 1.70E-07
2	128 128 128 128 128 128 128 132	1.59E-07	1.58E-07 - 1.61E-07
3	124 128 128 128 128 128 128 128	1.54E-07	1.53E-07 - 1.55E-07
4	128 128 128 128 128 128 132 132	1.51E-07	1.51E-07 - 1.52E-07
5	124 128 128 128 128 128 128 132	1.47E-07	1.47E-07 - 1.48E-07
6	128 128 128 128 128 132 132 132	1.45E-07	1.44E-07 - 1.45E-07
7	124 124 128 128 128 128 128 128	1.42E-07	1.41E-07 - 1.43E-07
8	124 128 128 128 128 128 132 132	1.40E-07	1.40E-07 - 1.40E-07
9	128 128 128 128 132 132 132 132	1.37E-07	1.37E-07 - 1.38E-07
10	124 124 128 128 128 128 128 132	1.35E-07	1.35E-07 - 1.36E-07

Table 5.14: Statistical Results of Balanced Functions for n=8

5.3 Observations

Now we list some properties observed from the statistical tables.

Observation-1. Let $f, g \in \mathcal{F}_n$ are balanced. Suppose that $S_i(f) + S_i(g) = 2^n$ for $i \in \{1, \ldots, n\}$. Reorder $S(f)$ and $S(g)$ by arranging the components in decreasing order to obtain $S'(f)$ and $S'(g)$. In this rearrangement say that the first unequal components of $S'(f)$ and $S'(g)$ are at position j and if $S'_j(f) > S'_j(g)$ then $[S(f)]_B > [S(g)]_B$, where $[S]_B$ denotes the number of balanced functions having difference distribution vector S.

Observation-2. For any nonnegative integers $a_1, \ldots, a_k$,

$$[a_1, \ldots, a_{k-1}, a_k, 2^{n-1}, 2^{n-1}, \ldots, 2^{n-1}] \leq [a_1, \ldots, a_{k-1}, 2^{n-1}, \ldots, 2^{n-1}].$$

Equality holds if and only if $a_k = 2^{n-1}$.

Observation-3. Ordering properties of difference distribution vectors.

From the statistical results, we observe that the below descending order holds for $n = 5, 6, 7$ and 8 for the functions of even weight.

1)-$[2^{n-1}, 2^{n-1}, \ldots, 2^{n-1}]$

2)-$[2^{n-1} + 4, 2^{n-1}, \ldots, 2^{n-1}]$

3)-$[2^{n-1} + 4, 2^{n-1} + 4, 2^{n-1}, \ldots, 2^{n-1}]$

4)-$[2^{n-1} + 4, 2^{n-1} + 4, 2^{n-1} + 4, 2^{n-1}, \ldots, 2^{n-1}]$

5)-$[2^{n-1} + 4, 2^{n-1} + 4, 2^{n-1} + 4, 2^{n-1} + 4, 2^{n-1}, \ldots, 2^{n-1}]$

6)-$[2^{n-1} + 8, 2^{n-1}, \ldots, 2^{n-1}]$

7)-$[2^{n-1} + 4, 2^{n-1} + 4, 2^{n-1} + 4, 2^{n-1} + 4, 2^{n-1} + 4, 2^{n-1}, \ldots, 2^{n-1}]$

8)-$[2^{n-1} + 8, 2^{n-1} + 4, 2^{n-1}, \ldots, 2^{n-1}]$

The above table can be represented as follows, where each component is the deviation from 2^{n-1}.

1)-$[0, 0, 0, 0, 0, 0, \ldots, 0]$

2)-$[4, 0, 0, 0, 0, 0, \ldots, 0]$

3)-$[4, 4, 0, 0, 0, 0, \ldots, 0]$

4)-$[4, 4, 4, 0, 0, 0, \ldots, 0]$

5)-$[4, 4, 4, 4, 0, 0, \ldots, 0]$

6)-$[8, 0, 0, 0, 0, 0, \ldots, 0]$

7)-$[4, 4, 4, 4, 4, 0, \ldots, 0]$

8)-$[8, 4, 0, 0, 0, 0, \ldots, 0]$

Keeping the same notation as above we list some other tables.

The below descending order holds for $n = 5, 6, 7$ and 8 for the functions of odd weight.

1)-$[2, 2, 2, 2, 2, \ldots, 2]$

2)-$[6, 2, 2, 2, 2, \ldots, 2]$

3)-$[6, 6, 2, 2, 2, \ldots, 2]$

4)-$[6, 6, 6, 2, 2, \ldots, 2]$

5)-$[10, 2, 2, 2, 2 \ldots, 2]$

6)-$[6, 6, 6, 6, 2, \ldots, 2]$

7)-$[10, 6, 2, 2, 2, \ldots, 2]$

Similarly, the below descending order holds for $n = 5, 6, 7$ and 8 for balanced functions.

1)-$[0, 0, 0, 0, 0, \ldots, 0]$

2)-$[4, 0, 0, 0, 0, \ldots, 0]$

3)-$[-4, 0, 0, 0, 0, \ldots, 0]$

4)-$[4, 4, 0, 0, 0, \ldots, 0]$

5)-$[4, -4, 0, 0, 0, \ldots, 0]$

6)-$[4, 4, 4, 0, 0, \ldots, 0]$

Chapter 6

Conclusion

In this thesis, we deal with the Boolean functions having particular difference distribution vectors and we give the ordering properties of the difference distribution vectors by using statistics. Furthermore, some new constructions of the functions satisfying strict avalanche criterion are proposed in Chapter 4. Also two known constructions, by Jennifer Seberry, Xian-Mo Zhang [7], are given in an alternative notation. In Chapter 5, some statistical results are presented, and three important observations are given. These observation will be proved for any n in our further studies.

References

[1] Feistel H., *Cryptography and computer privacy*, Scienctific American, 228(5): 15-23, 1973.

[2] Forré R., *The strict avalanche criterion: Spectral properties of Boolean functions and extended definition*, Advances in Cryptology - CRYPTO'88 (Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, Newyork) 403: 450-468, 1988.

[3] Kam J.B. and Davida G.I., *Structured design of substitution permutation encryption networks.*, IEEE Transactions on Computers, C-28(10):747-753, 1979.

[4] O'Connor L., *An upper bound on the number of functions satisfying the Strict Avalanche Criterion*, Information Processing Letters 52(6): 325-327, 1994.

[5] Preneel B., Leekwijck W.V., Linden L.V., Govaerts R., and Vandewalle J., Propagation characteristics of Boolean functions, Advances in Cryptology - EUROCRYPT90 (Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, New York 1991) 437: 155-165, 1990.

[6] Sağdıçoğlu S., *Cryptological viewpoint of Boolean functions*, M. Sc. Thesis, The Department of Mathematics, Middle East Technical University, Ankara, Turkey, 2003.

[7] Seberry J. and Zhang X.M., *Highly Nonlinear 0-1 Balanced Boolean Functions Satisfying Avalanche Criterion*, Advances in Cryptology -

AUSCRYPT'92 Proceedings, Lecture Notes in Computer Science 718. Springer-Verlag, 1993.

[8] Siegenthaler T., *Correlation-immunity of nonlinear combining functions for cryptographic applications*, IEEE Transactions on Information Theory, IT-30, No. 5: 776-779, 1984.

[9] Webster A.F. and Tavares S.E., *On the design of S-boxes*, Advances in Cryptology - CRYPTO'85 ed. H.C. Williams (Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, Newyork) 218: 523-524, 1985.