Boolean Functions and S-boxes

New Bounds on the Number of Functions Satisfying the Strict Avalanche Criterion

A. M. Youssef[†] , T.W. Cusick[††], P. Stănică[††] and S.E. Tavares[†]

[†]Department Of Electrical and Computer Engineering

Queen's University, Kingston, Ontario, Canada, K7L 3N6

[††] Department of Mathematics, SUNY at Buffalo, 106 Diefendorf Hall

Buffalo, New York 14214–3093

Key words: Cryptography; Strict Avalanche Criterion; Boolean functions; Enumeration; Combinatorial problems

Abstract:

In this paper we present asymptotic expressions for the number of functions satisfying the Strict Avalanche Criterion (SAC) with respect to one and two variables, previously developed by O'Connor. Cusick recently gave a conjecture for a lower bound on the number of functions satisfying the SAC. Here, we give a constructive proof for this conjecture. Moreover, we provide an improved lower bound.

1. Introduction

The Strict Avalanche Criterion (SAC) was introduced by Webster and Tavares [11] in a study of design criteria for certain cryptographic functions. A boolean function $f : Z_2^n \rightarrow Z_2$ is said to satisfy the SAC if complementing a single input bit results in changing the output bit with probability exactly one half.

The SAC was intended to combine two earlier criteria for cryptographic applications due to [6] and [4]. Forré [5] extended the concept by defining higher order SAC. A boolean function on n variables is said to satisfy the SAC of order k, $0 \leq k \leq n-2$, if whenever k input bits are fixed arbitrarily, the resulting function of $n-k$ variables satisfies the SAC. It is easy to see [7] that if a function satisfies the SAC of order k, then it also satisfies the SAC of order j for any $j = 0, 1, \ldots, k-1$.

As in the case with any criterion of cryptographic significance, it is of interest to count the functions which satisfy the criterion. Many recent papers (for example [7], [2]) have been concerned with counting functions that satisfy the SAC of various orders. It is easier to count the functions satisfying the SAC of the largest order, because relatively few functions exist which satisfy these stringent criteria.

2. Main Results

O'Connor [9] gave an upper bound for the number of functions $f(\mathbf{x})$, where $\mathbf{x} = (x_1, ..., x_n)$ is in Z_2^n, satisfying the SAC. Let $S(n,k)$ denote the number of functions for which the output changes with probability $1/2$ if any one of the input bits $x_1, ..., x_k$ is complemented. He also gave [9] explicit formulas for $S(n,1)$ and $S(n,2)$; of course these are upper bounds for the number of functions satisfying the SAC. In this paper we give asymptotics for the size of $S(n,1)$ and $S(n,2)$, thus quantifying the upper bound for the number of SAC function given in [9].

Cusick [1] gave a lower bound for the number of functions satisfying the SAC. He also gave a conjecture that provided an improvement of the lower bound. In this paper, we give a constructive proof for this conjecture. Moreover, we provide an improved lower bound. We also give a lower bound for the number of balanced functions that satisfies the SAC.

Notation:

Throughout this paper, let

$f_n : Z_2^n \rightarrow Z_2$ describes a boolean function with n input variables.

$V = \{\mathbf{v}_i \mid 0 \leq i \leq 2^n - 1\}$: denote the set of vectors in Z_2^n in lexicographical order. A boolean function $f_n(\mathbf{x})$ is specified by $f_n(\mathbf{x}) = [b_0, b_1, ..., b_{2^n-1}]$, where $b_i = f_n(\mathbf{v}_i)$.

$\mathbf{e}$: denotes any element of Z_2^n with hamming weight 1. Let $\grave{\mathbf{e}}$, $\grave{\mathbf{v}}_i$ denote the $n-1$ least significant bits of $\mathbf{e}$ and $\mathbf{v}_i$ respectively.

$\mathbf{a}$: denotes any element of Z_2^{n-1} with odd hamming weight.

$g_n : Z_2^n \rightarrow Z_2$: denotes the boolean function $\mathbf{1} \cdot \mathbf{x} \oplus b$, $b \in Z_2$. It is easy see that g_n satisfies

$$g_n(\mathbf{x}) = \overline{g}_n(\mathbf{x} \oplus \mathbf{a}) . \qquad (1)$$

$MSB(\cdot)$ denotes the most significant bit of the enclosed argument.

Definition 1 [11]: A boolean function $f_n : Z_2^n \rightarrow Z_2$ is said to satisfy SAC if complementing a single input bit results in changing the output bit with probability exactly one half, i.e.,

$$\sum_{i=0}^{2^n-1} f_n(\mathbf{v}_i) \oplus f_n(\mathbf{v}_i \oplus \mathbf{e}) = 2^{n-1}. \qquad (2)$$

Definition 2 [3], [8]: A linear structure of a boolean function $f_n : Z_2^n \rightarrow Z_2$ is identified as a vector $\mathbf{c} \neq \mathbf{0} \in Z_2^n$ such that $f_n(\mathbf{v}_i \oplus \mathbf{c}) \oplus f_n(\mathbf{v}_i)$ takes the same value (0 or 1) for all $i, 0 \leq i \leq 2^n - 1$.

The results of O'connor [9] are quantified by the following two Lemmas.

Lemma 1

$$S(n,1) \sim 2\,\pi^{-1}\,2^{2^n - n/2}. \tag{3}$$

Proof: Lemma 1 of [9] states

$$S(n,1) = \binom{2^{n-1}}{2^{n-2}} 2^{2^{n-1}}. \tag{4}$$

Applying Stirling's formula, $n! \sim (2\pi n)^{1/2}(n/e)^n$, to the binomial coefficient proves the Lemma. □

Lemma 2

For $n \geq 2$,

$$S(n,2) > 2^{2^n - n}. \tag{5}$$

Proof: Lemma 2 of [9] gives the formula

$$S(n,2) = \sum_{i=0}^{2^{n-3}} \binom{2^{n-2}}{2i} 2^{3 \cdot 2^{n-2} - 4i} \sum_{j=0}^{i} \binom{2i}{2j}\binom{2j}{j}\binom{2i-2j}{i-j}. \tag{6}$$

Expanding the binomial coefficients shows that the inner sum is equal to the binomial coefficient sum $m(i)$ given by

$$m(i) = \sum_{j=0}^{i} \binom{i}{j}^2 \binom{2i}{i} = \binom{2i}{i}^2. \tag{7}$$

It is easy to prove by induction that $m(i) > 2^{4i-2}/i$ for $i \geq 2$. Thus we have

$$S(n,2) > \sum_{i=0}^{2^{n-3}} \frac{1}{i} \binom{2^{n-2}}{2i} 2^{3 \cdot 2^{n-2} - 2}. \tag{8}$$

By noting that

$$\begin{aligned} &\sum_{i=0}^{[M/2]} (2i+1)^{-1} \binom{M}{2i} x^{2i+1} \\ &\qquad = \frac{1}{2}(M+1)^{-1}\Big((1+x)^{M+1} - (1-x)^{M+1}\Big) \end{aligned} \tag{9}$$

and taking $M = 2^{n-2}$ and $x = 1$, we have

$$\sum_{i=0}^{2^{n-3}} \frac{1}{i} \binom{2^{n-2}}{2i} > 2 \sum_{i=0}^{2^{n-3}} (2i+1)^{-1} \binom{2^{n-2}}{2i} = (2i+1)^{-1} 2^{2^{n-2}+1} \tag{10}$$

which proves the Lemma. □

If we use Lemma 1 and Lemma 2 in the inequality (8) of [9], we have that the fraction of functions satisfying the SAC is asymptotically less than

$$2\pi^{-1/2}\, n^{-1}\, 2^{-n/2}. \tag{11}$$

Now we turn to the problem of lower bounds.

The following conjecture was given in [1] without proof. This conjecture implies that there are at least $2^{2^{n-1}}$ boolean functions of n variables which satisfy the SAC.

Conjecture [1]: Given any choice of the values $f_n(\mathbf{v}_i)$, $\ 0 \leq i \leq 2^{n-1}-1$, there exists a choice of $f_n(\mathbf{v}_i)$, $2^{n-1} \leq i \leq 2^n - 1$, such that the resulting function $f_n(\mathbf{x})$ satisfies the SAC.

For $n = 1$, it is trivial to show that if $f_1(1) = f_1(0) \oplus 1$ then the resulting function satisfies the SAC. In the following Lemma we prove that, for $n \geq 2$, there exist at least two choices for $f_n(\mathbf{v}_i)$, $2^{n-1} \leq i \leq 2^n - 1$, such that the resulting function satisfies the SAC.

Lemma 3:

Let $f_n = [h_{n-1} \ \ [h_{n-1} \oplus g_{n-1}]]$ where h_{n-1} is an arbitrary boolean function with $n-1$ input variables, $n \geq 2$, and g_{n-1} is constructed as above to satisfy equation (1), then f_n satisfies the SAC.

Proof:

Case 1: $MSB(\mathbf{e}) \ = \ 0$:

$$\begin{aligned}
&\sum_{i=0}^{2^n-1} f_n(\mathbf{v}_i) \oplus f_n(\mathbf{v}_i \oplus \mathbf{e}) \\
&\quad = \sum_{i=0}^{2^{n-1}-1} f_n(\mathbf{v}_i) \oplus f_n(\mathbf{v}_i \oplus \mathbf{e}) + \sum_{i=2^{n-1}}^{2^n-1} f_n(\mathbf{v}_i) \oplus f_n(\mathbf{v}_i \oplus \mathbf{e}) \\
&\quad = \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i) \oplus h_{n-1}(\grave{\mathbf{v}}_i \oplus \grave{\mathbf{e}}) \\
&\qquad + \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i) \oplus h_{n-1}(\grave{\mathbf{v}}_i \oplus \grave{\mathbf{e}}) \oplus g_{n-1}(\grave{\mathbf{v}}_i) \oplus g_{n-1}(\grave{\mathbf{v}}_i \oplus \grave{\mathbf{e}}) \\
&\quad = \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i) \oplus h_{n-1}(\grave{\mathbf{v}}_i \oplus \grave{\mathbf{e}}) + \sum_{i=0}^{2^{n-1}-1} \overline{(h_{n-1}(\grave{\mathbf{v}}_i) \oplus h_{n-1}(\grave{\mathbf{v}}_i \oplus \grave{\mathbf{e}}))} \\
&\quad = 2^{n-1}.
\end{aligned}$$

Case 2: $MSB(\mathbf{e}) = 1$:

$$\begin{aligned}
&\sum_{i=0}^{2^n-1} f_n(\mathbf{v}_i) \oplus f_n(\mathbf{v}_i \oplus \mathbf{e}) \\
&\qquad = 2 \sum_{i=0}^{2^{n-1}-1} f_n(\mathbf{v}_i) \oplus f_n(\mathbf{v}_i \oplus \mathbf{e}) \\
&\qquad = 2 \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i) \oplus h_{n-1}(\grave{\mathbf{v}}_i) \oplus g_{n-1}(\grave{\mathbf{v}}_i) \\
&\qquad = 2 \sum_{i=0}^{2^{n-1}-1} g_{n-1}(\grave{\mathbf{v}}_i) \\
&\qquad = 2^{n-1}.
\end{aligned}$$

which proves the Lemma. □

From Lemma 3 above, and by noting that we have two choices for g_n, we conclude that, for $n \geq 2$, the number of function satisfying the SAC is lower bounded by $2^{2^{n-1}+1}$. Using the following Lemma, one can provide some improvement to the above bound.

Lemma 4:

Let $f_n = [h_{n-1} \;\; [l_{n-1} \oplus g_{n-1}]]$ where h_{n-1} is an arbitrary boolean function with $n-1$ input variables, $l_{n-1}(\mathbf{x}) = h_{n-1}(\mathbf{x} \oplus \mathbf{a})$, $n \geq 2$, and g_{n-1} is constructed as above to satisfy equation (1), then f_n satisfies the SAC.

Proof:

Case 1: $MSB(\mathbf{e}) = 0$:

$$\begin{aligned}
&\sum_{i=0}^{2^n-1} f_n(\mathbf{v}_i) \oplus f_n(\mathbf{v}_i \oplus \mathbf{e}) \\
&\qquad = \sum_{i=0}^{2^{n-1}-1} f_n(\mathbf{v}_i) \oplus f_n(\mathbf{v}_i \oplus \mathbf{e}) + \sum_{i=2^{n-1}}^{2^n-1} f_n(\mathbf{v}_i) \oplus f_n(\mathbf{v}_i \oplus \mathbf{e}) \\
&\qquad = \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i) \oplus h_{n-1}(\grave{\mathbf{v}}_i \oplus \grave{\mathbf{e}}) \\
&\qquad\quad + \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i \oplus \mathbf{a}) \oplus h_{n-1}(\grave{\mathbf{v}}_i \oplus \mathbf{a} \oplus \grave{\mathbf{e}}) \oplus g_{n-1}(\grave{\mathbf{v}}_i) \oplus g_{n-1}(\grave{\mathbf{v}}_i \oplus \grave{\mathbf{e}}) \\
&\qquad = \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i) \oplus h_{n-1}(\grave{\mathbf{v}}_i \oplus \grave{\mathbf{e}}) + \sum_{i=0}^{2^{n-1}-1} \overline{(h_{n-1}(\grave{\mathbf{v}}_i) \oplus h_{n-1}(\grave{\mathbf{v}}_i \oplus \grave{\mathbf{e}}))} \\
&\qquad = 2^{n-1}.
\end{aligned}$$

Case 2: $MSB(\mathbf{e}) = 1$:

$$\begin{aligned}
&\sum_{i=0}^{2^n-1} f_n(\mathbf{v}_i) \oplus f_n(\mathbf{v}_i \oplus \mathbf{e}) \\
&\quad = 2 \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\dot{\mathbf{v}}_i) \oplus h_{n-1}(\dot{\mathbf{v}}_i \oplus \mathbf{a}) \oplus g_{n-1}(\dot{\mathbf{v}}_i) \\
&\quad = \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\dot{\mathbf{v}}_i) \oplus h_{n-1}(\dot{\mathbf{v}}_i \oplus \mathbf{a}) \oplus g_{n-1}(\dot{\mathbf{v}}_i) \\
&\qquad + \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\dot{\mathbf{v}}_i \oplus \mathbf{a}) \oplus h_{n-1}(\dot{\mathbf{v}}_i) \oplus g_{n-1}(\dot{\mathbf{v}}_i \oplus \mathbf{a}) \\
&\quad = \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\dot{\mathbf{v}}_i) \oplus h_{n-1}(\dot{\mathbf{v}}_i \oplus \mathbf{a}) \oplus g_{n-1}(\dot{\mathbf{v}}_i) \\
&\qquad + \sum_{i=0}^{2^{n-1}-1} \overline{h_{n-1}(\dot{\mathbf{v}}_i \oplus \mathbf{a}) \oplus h_{n-1}(\dot{\mathbf{v}}_i) \oplus g_{n-1}(\dot{\mathbf{v}}_i)} \\
&\quad = 2^{n-1}.
\end{aligned}$$

which proves the Lemma. □

Note that if the function f_{n-1} does not have any linear structures, then all the functions generated by $l_{n-1} \oplus g_{n-1}$ will be unique for all the 2^{n-2} choices of $\mathbf{a}$. From Lemma 3 and Lemma 4 we have $2^{n-1} + 2$ distinct choices for $f_{n-1}(\mathbf{v}_i)$, $2^{n-1} \leq i \leq 2^n - 1$. Thus we have the following corollary:

Corollary 1:

The number of functions satisfying the SAC is lower bounded by

$$\left(2^{2^{n-1}} - \mathcal{LS}^{n-1}\right)\left(2^{n-1} + 2\right) + 2\mathcal{LS}^{n-1} \tag{16}$$

where $\mathcal{LS}^{n-1}$ is the number of functions with $n-1$ input bits having any linear structure. An exact count for $\mathcal{LS}^n$ is given in [10]. It can also be shown [10] that $\mathcal{LS}^n$ is asymptotic to $(2^n - 1)2^{2^{n-1}+1}$. One should note that while this bound provides some improvement over the proved bound in [1], exhaustive search (see Table 1) shows that the quality of this bound degrades as n increases. One can improve this bound slightly by identifying special classes of functions $f_n(\mathbf{v}_i)$, $0 \leq i \leq 2^{n-1} - 1$ for which there is a large number of choices for $f_n(\mathbf{v}_i)$, $2^{n-1} \leq i \leq 2^n - 1$ such that the resulting function, f_n, satisfies the SAC. For example, if the function h_{n-1} satisfies the SAC, then the function $f_n = [h_{n-1}[h_{n-1} \oplus \mathbf{c}{\cdot}\mathbf{x} \oplus b]]$, $b \in Z_2$ also satisfies the SAC. Thus our bound is slightly improved to

$$\left(2^{2^{n-1}} - \mathcal{LS}^{n-1} - \mathcal{SAC}^{n-1}\right)\left(2^{n-1} + 2\right) + 2^n\mathcal{SAC}^{n-1} + 2\mathcal{LS}^{n-1} \tag{17}$$

where $\mathcal{SAC}^{n-1}$ is the number of functions with $n-1$ input bits that satisfy the SAC.

We now give a lower bound on the number of balanced functions that satisfy the SAC.

Lemma 5

Let $f_n = [h_{n-1} \;\; [l_{n-1} \oplus g_{n-1}]]$ where h_{n-1} is an arbitrary boolean function with $n-1$ input variables that satisfies $\sum\limits_{wt(\mathbf{v}_i)\ odd} h_{n-1}(\mathbf{v}_i) = 2^{n-3}$, $l_{n-1}(\mathbf{x}) = h(\mathbf{x} \oplus \mathbf{a})$, $n \geq 2$, and g_{n-1} is constructed as above to satisfy equation (1), then f_n is a balanced function that satisfies the SAC.

Proof:

From Lemma 5, it follows that f_n satisfies the SAC. Here we will prove that f_n is a balanced function.

$$\begin{aligned}
\sum_{i=0}^{2^n-1} f_n(\mathbf{v}_i) &= \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i) + \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i \oplus \mathbf{a}) \oplus g_{n-1}(\grave{\mathbf{v}}_i) \\
&= \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i) + \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i) \oplus g_{n-1}(\grave{\mathbf{v}}_i \oplus \mathbf{a}) \\
&= \sum_{i=0}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i) + \sum_{i=0}^{2^{n-1}-1} \overline{h_{n-1}(\grave{\mathbf{v}}_i) \oplus 1 \cdot \grave{\mathbf{v}}_i} \\
&= \sum_{wt(\grave{\mathbf{v}}_i)\ even}^{2^{n-1}-1} \left(h_{n-1}(\grave{\mathbf{v}}_i) + \overline{h_{n-1}(\grave{\mathbf{v}}_i)}\right) \quad + 2 \sum_{wt(\grave{\mathbf{v}}_i)\ odd}^{2^{n-1}-1} h_{n-1}(\grave{\mathbf{v}}_i) \\
&= 2^{n-2} + 2 \cdot \; 2^{n-3} \\
&= 2^{n-1}.
\end{aligned}$$

which proves the Lemma. □

Similarly, one can also show that the function $f_n = [h_{n-1} \;\; [h_{n-1} \oplus g_{n-1}]]$ where h_{n-1} is an arbitrary boolean function that satisfies $\sum\limits_{wt(\mathbf{v}_i)\ even} h_{n-1}(\mathbf{v}_i) = 2^{n-3}$ is a balanced function that satisfies the SAC. From the Lemma above, it follows that the number of balanced SAC functions is lower bounded by

$$\binom{2^{n-2}}{2^{n-3}} 2^{2^{n-2}+1}. \tag{19}$$

n	2	3	4	5
$\mathcal{LS}^n$	4	8	128	4,992
Old Bound [1]	2	4	16	256
New Bound (exp. (16))	8	64	1,536	1,099,776
New Bound (exp. (17))	8	64	1,920	1,157,568
Exact Number	8	64	4,128	27,522,560

Table 1 : Exact number of functions satisfying SAC versus the derived lower bounds.

References

[1] T. W. Cusick. Bounds on the number of functions satisfying the Strict Avalanche Criterion. *Information Processing Letters, 57, pp. 261–263*, 1996.

[2] T.W. Cusick. Boolean functions satisfying a higher order strict avalanche criterion. *Advances in Cryptology: Proc. of EUROCRYP '93, Springer-Verlag, pp. 102–117*, 1994.

[3] J.H. Evertse. Linear structures in block ciphers. *Advances in Cryptology: Proc. of EUROCRYPT '87, Springer-Verlag, Berlin, pp.249–266*, 1988.

[4] H. Feistel. Cryptography and computer privacy. *Scientific American, 228, pp. 15–23*, 1973.

[5] R. Forré. The strict avalanche criterion: Spectral properties of boolean functions and an extended definition. *Advances in Cryptology: Proc. of CRYPTO '88, Springer-Verlag, pp. 450–468*, 1989.

[6] J.B. Kam and G.I. Davida. Structured design of substitution-permutation encryption networks. *IEEE Trans. Comp. C-28, pp.747–753*, 1979.

[7] S. Lloyd. Counting functions satisfying a higher order strict avalanche criterion. *Advances in Cryptology: Proc. of EUROCRYP '89, Springer-Verlag, pp.63–74*, 1990.

[8] W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. *Advances in Cryptology: Proc. of EUROCRYPT' 89, Springer-Verlag, pp. 549–562*, 1990.

[9] L.J. O'Connor. An upper bound on the number of functions satisfying the Strict Avalanche Criterion. *Information Processing Letters, 52, pp.325–327*, 1994.

[10] L.J. O'Connor and A. Klapper. Algebraic nonlinearity and its application to cryptography. *Journal of Cryptology, Vol.7, pp. 213–227*, 1994.

[11] A.F. Webster and S.E. Tavares. On the design of S-boxes. *Advances in Cryptology : Proc. of CRYPTO '85 , Springer-Verlag, pp. 523–534*, 1986.

Difference Distribution Table of a Regular Substitution Box

Xian-Mo Zhang
Department of Computer Science
The University of Wollongong
Wollongong, NSW 2522, AUSTRALIA
E-mail: xianmo@cs.uow.edu.au

Yuliang Zheng
School of Computing and Information Technology
Monash University
Melbourne, VIC 3199, AUSTRALIA
E-mail: yzheng@fcit.monash.edu.au

July 15, 1996

This short paper reports an interesting property of the difference distribution table of an S-box or substitution box, which has been discovered by the authors while studying relationships between differential and other cryptographic characteristics of an S-box. Namely, an $n \times m$ S-box is regular if and only if the sum of the entries in a column in the difference distribution table of the S-box is 2^{2n-m}.

Denote by V_n the vector space of n tuples of elements from $GF(2)$. An $n \times m$ S-box is a mapping from V_n to V_m, i.e., $F = (f_1, \ldots, f_m)$, where n and m are integers with $n \geqq m \geqq 1$ and each component function f_j is a function from V_n to $GF(2)$ (or on V_n for short).

The *Sylvester-Hadamard matrix (or Walsh-Hadamard matrix)* of order 2^n, denoted by H_n, is generated by the recursive relation

$$H_n = \begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{bmatrix}, \; n = 1, 2, \ldots, \; H_0 = 1.$$

Each row (column) of H_n is a linear sequence of length 2^n.

In cryptography we are mainly concerned with *regular* S-boxes. An S-box $F = (f_1, \ldots, f_m)$ is said to be regular if $F(x)$ runs through each vector in V_m 2^{n-m} times while x runs through V_n once. It is well-known that a regular S-box can be characterized by the balance of the linear combinations of its component functions. The following is a re-statement of Corollary 7.39 of [1]:

Lemma 1 *Let $F = (f_1, \ldots, f_m)$ be a mapping from V_n to V_m, where n and m are integers with $n \geqq m \geqq 1$ and each $f_j(x)$ is a function on V_n. Then F is regular if and only if every non-zero linear combination of $f_1, \ldots, f_m$, $f(x) = \bigoplus_{j=1}^{m} c_j f_j(x)$, is balanced.*

Now we introduce three notations: $k_j(\alpha)$, $\Delta_j(\alpha)$ and η_j associated with $F = (f_1, \ldots, f_m)$.

Definition 1 *Let $F = (f_1, \ldots, f_m)$ be an $n \times m$ S-box, $\alpha \in V_n$, $j = 0, 1, \ldots, 2^m - 1$ and $\beta_j = (b_1, \ldots, b_m)$ be the vector in V_m that corresponds to the binary representation of j. In addition, set $g_j = \bigoplus_{u=1}^{m} b_u f_u$ be the jth linear combination of the component functions of F. Then we define*

1. *$k_j(\alpha)$ as the number of times $F(x) \oplus F(x \oplus \alpha)$ runs through $\beta_j \in V_m$ while x runs through V_n once.*
2. *$\Delta_j(\alpha)$ as the auto-correlation of g_j with shift α.*
3. *η_j as the sequence of g_j.*

Using the three notations we introduce three matrices in the following:

Definition 2 For $F = (f_1, \ldots, f_m)$, set

$$K = \begin{bmatrix} k_0(\alpha_0) & k_1(\alpha_0) & \ldots & k_{2^m-1}(\alpha_0) \\ k_0(\alpha_1) & k_1(\alpha_1) & \ldots & k_{2^m-1}(\alpha_1) \\ & \vdots & & \\ k_0(\alpha_{2^n-1}) & k_1(\alpha_{2^n-1}) & \ldots & k_{2^m-1}(\alpha_{2^n-1}) \end{bmatrix},$$

$$D = \begin{bmatrix} \Delta_0(\alpha_0) & \Delta_1(\alpha_0) & \ldots & \Delta_{2^m-1}(\alpha_0) \\ \Delta_0(\alpha_1) & \Delta_1(\alpha_1) & \ldots & \Delta_{2^m-1}(\alpha_1) \\ & \vdots & & \\ \Delta_0(\alpha_{2^n-1}) & \Delta_1(\alpha_{2^n-1}) & \ldots & \Delta_{2^m-1}(\alpha_{2^n-1}) \end{bmatrix}$$

and

$$P = \begin{bmatrix} \langle \eta_0, \ell_0 \rangle^2 & \langle \eta_1, \ell_0 \rangle^2 & \cdots & \langle \eta_{2^m-1}, \ell_0 \rangle^2 \\ \langle \eta_0, \ell_1 \rangle^2 & \langle \eta_1, \ell_1 \rangle^2 & \cdots & \langle \eta_{2^m-1}, \ell_1 \rangle^2 \\ & \vdots & & \\ \langle \eta_0, \ell_{2^n-1} \rangle^2 & \langle \eta_1, \ell_{2^n-1} \rangle^2 & \cdots & \langle \eta_{2^m-1}, \ell_{2^n-1} \rangle^2 \end{bmatrix},$$

where ℓ_i is the ith row of H_n, $i = 0, 1, \ldots, 2^n - 1$. The three $2^n \times 2^m$ matrices K, D and P are called *difference distribution table, auto-correlation distribution table and correlation immunity distribution table* of the S-box F respectively.

In designing a strong S-box, many cryptographic criteria should be examined not only against component functions, but also against their linear combinations. Such criteria include those related to nonlinearity, propagation characteristics and difference distribution tables. The matrix K characterizes the differential characteristics of an S-box. The matrix D indicates the auto-correlation of all linear combinations of the component functions. While the matrix P represents the inner product between the sequence of each linear combination of the component functions and each linear sequence. P is helpful in studying the correlation immunity, as well as the nonlinearity, of each linear combination of the component functions (see [2]).

As one immediately expects, the three matrices K, D and P are closely related. In particular the following result has been proven in [4]:

Theorem 1 *Let $F = (f_1, \ldots, f_m)$ be a mapping from V_n to V_m, where n and m are integers with $n \geqq m \geqq 1$ and each $f_j(x)$ is a function on V_n. Set $g_j = \bigoplus_{u=1}^{m} c_u f_u$ where $(c_1, \ldots, c_m)$ is the binary representation of integer j, $j = 0, 1, \ldots, 2^m - 1$. Then* *(i)* $D = KH_m$,

(ii) $P = H_nD$,

(iii) $P = H_nKH_m$.

Using Theorem 1, we now show that a regular S-box can be completely characterized by its difference distribution table.

Corollary 1 *Let $F = (f_1, \dots, f_m)$ be a mapping from V_n to V_m, where n and m are integers with $n \geqq m \geqq 1$ and each f_j is a function on V_n. Then F is regular if and only if the sum of a column in the difference distribution table is 2^{2n-m}, i.e., $\sum_{\alpha \in V_n} k_i(\alpha) = 2^{2n-m}$, $i = 0, 1, \dots, 2^m - 1$.*

Proof. Compare the first rows in both sides of the formula in (iii) of Theorem 1,

$$(\sum_{\alpha\in V_n} k_0(\alpha), \sum_{\alpha\in V_n} k_1(\alpha), \dots, \sum_{\alpha\in V_n} k_{2^m-1}(\alpha))H_m = (\langle\eta_0, \ell_0\rangle^2, \langle\eta_1, \ell_0\rangle^2, \dots, \langle\eta_{2^m-1}, \ell_0\rangle^2). \quad (1)$$

Obviously, if $\sum_{\alpha\in V_n} k_i(\alpha) = 2^{2n-m}$, $i = 0, 1, \dots, 2^m - 1$. then $\langle\eta_1, \ell_0\rangle^2 = \cdots = \langle\eta_{2^m-1}, \ell_0\rangle^2 = 0$. Note that ℓ_0 is the all-one sequence of length 2^n. Hence $g_1, \dots, g_{2^m-1}$ are balanced, where $g_1, \dots, g_{2^m-1}$ are defined in Theorem 1. By Lemma 1, F is regular.

Conversely, suppose F is regular. By Lemma 1, $g_1, \dots, g_{2^m-1}$ are balanced. Hence $\langle\eta_1, \ell_0\rangle^2 = \cdots = \langle\eta_{2^m-1}, \ell_0\rangle^2 = 0$. Note that $\langle\eta_0, \ell_0\rangle^2 = 2^{2n}$. Rewrite (1) as

$$2^m(\sum_{\alpha\in V_n} k_0(\alpha), \sum_{\alpha\in V_n} k_1(\alpha), \cdots, \sum_{\alpha\in V_n} k_{2^m-1}(\alpha)) = (2^{2n}, 0, \dots, 0)H_m.$$

This proves that $\sum_{\alpha\in V_n} k_i(\alpha) = 2^{2n-m}$, $i = 0, 1, \dots, 2^m - 1$. □

Corollary 1 has also been obtained independently by Tapia-Recillas, Daltabuit and Vega [3].

The following corollary shows the uniqueness of the first column of the difference distribution table of a regular mapping.

Corollary 2 *Let $F = (f_1, \dots, f_m)$ be a mapping from V_n to V_m, where n and m are integers with $n \geqq m \geqq 1$ and each f_j is a function on V_n. Then F is regular if and only if the sum of the leftmost column is 2^{2n-m}, i.e., $\sum_{\alpha\in V_n} k_0(\alpha) = 2^{2n-m}$.*

Proof. Multiply both sides of the equality in (iii) of Theorem 1 by e^T where, e denotes the all-one sequence of length 2^m. Hence we have

$$H_n \begin{bmatrix} k_0(\alpha_0) & k_1(\alpha_0) & \dots & k_{2^m-1}(\alpha_0) \\ k_0(\alpha_1) & k_1(\alpha_1) & \dots & k_{2^m-1}(\alpha_1) \\ & \vdots & & \\ k_0(\alpha_{2^n-1}) & k_1(\alpha_{2^n-1}) & \dots & k_{2^m-1}(\alpha_{2^n-1}) \end{bmatrix} \begin{bmatrix} 2^m \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \sum_{j=0}^{2^m-1}\langle\eta_j, \ell_0\rangle^2 \\ \sum_{j=0}^{2^m-1}\langle\eta_j, \ell_1\rangle^2 \\ \vdots \\ \sum_{j=0}^{2^m-1}\langle\eta_j, \ell_{2^n-1}\rangle^2 \end{bmatrix}$$

and hence

$$2^mH_n \begin{bmatrix} k_0(\alpha_0) \\ k_0(\alpha_1) \\ \vdots \\ k_0(\alpha_{2^n-1}) \end{bmatrix} = \begin{bmatrix} \sum_{j=0}^{2^m-1}\langle\eta_j, \ell_0\rangle^2 \\ \sum_{j=0}^{2^m-1}\langle\eta_j, \ell_1\rangle^2 \\ \vdots \\ \sum_{j=0}^{2^m-1}\langle\eta_j, \ell_{2^n-1}\rangle^2 \end{bmatrix}. \quad (2)$$

Compare the two sides of equality (2), obtaining

$$2^m \sum_{i=0}^{2^n-1} k_0(\alpha_i) = \sum_{j=0}^{2^m-1} \langle \eta_j, \ell_0 \rangle^2. \tag{3}$$

Since g_0 is the constant zero, η_0 is the all-one sequence of length 2^n and hence $\langle \eta_0, \ell_0 \rangle^2 = 2^{2n}$.

If $\sum_{\alpha \in V_n} k_0(\alpha) = 2^{2n-m}$, hen from (3), $\langle \eta_1, \ell_0 \rangle^2 = \cdots = \langle \eta_{2^m-1}, \ell_0 \rangle^2 = 0$. Note that ℓ_0 is the all-one sequence of length 2^n. Hence $g_1, \ldots, g_{2^m-1}$ are balanced, where $g_1, \ldots, g_{2^m-1}$ are defined in Theorem 1. By Lemma 1, F is regular.

Conversely, if F is regular, then by Corollary 1 $\sum_{\alpha \in V_n} k_0(\alpha) = 2^{2n-m}$. □

From Corollaries 1 and 2, we conclude that (1) an S-box is regular, (2) the sum of the first column in its difference distribution table is 2^{2n-m}, and (3) the sum of each column in the difference distribution table is 2^{2n-m}, are all equivalent statements.

References

[1] LIDL, R., AND NIEDERREITER, H. *Finite Fields, Encyclopedia of Mathematics and Its Applications.* Cambridge University Press, 1983.

[2] SEBERRY, J., ZHANG, X. M., AND ZHENG, Y. On constructions and nonlinearity of correlation immune functions. In *Advances in Cryptology - EUROCRYPT'93* (1994), vol. 765, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, New York, pp. 181–199.

[3] TAPIA-RECILLAS, H., DALTABUIT, E., AND VEGA, G. Some results on regular mappings, 1996. (preprint).

[4] ZHANG, X. M., AND ZHENG, Y. Relationships between differential and other cryptographic characteristics of an S-box, 1996. (submitted for publication).