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 $f : \mathrm{Z}_2^n \to \mathrm{Z}_2$ 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

$f_n : \mathrm{Z}_2^n \to \mathrm{Z}_2$ describes a boolean function with n input variables.

$V = \{\mathbf{v}_i \mid 0 \le i \le 2^n - 1\}$: 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.

$g_n : \mathrm{Z}_2^n \to \mathrm{Z}_2$: denotes the boolean function $\mathbf{1} \cdot \mathbf{x} \oplus b$, $b \in \mathrm{Z}_2$. It is easy see that g_n satisfies

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

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

Definition 1 [11]: A boolean function $f_n : \mathrm{Z}_2^n \to \mathrm{Z}_2$ is said to satisfy SAC if complementing a single input bit results in changing the output bit with probability exactly one half, i.e.,

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

Definition 2 [3], [8]: A linear structure of a boolean function $f_n : \mathrm{Z}_2^n \to \mathrm{Z}_2$ is identified as a vector $\mathbf{c} \neq \mathbf{0} \in \mathrm{Z}_2^n$ 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 $i, 0 \le i \le 2^n - 1$.

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 $f_n(\mathbf{v}_i),\ \ 0 \le i \le 2^{n-1}-1$, there exists a choice of $f_n(\mathbf{v}_i),\ 2^{n-1} \le i \le 2^n - 1$, 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 $n \ge 2$, there exist at least two choices for $f_n(\mathbf{v}_i),\ 2^{n-1} \le i \le 2^n - 1$, 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, $n \ge 2$, 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: $MSB(\mathbf{e}) = 1$:

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

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 $f_n = [h_{n-1}[h_{n-1} \oplus \mathbf{c}{\cdot}\mathbf{x} \oplus b]]$, $b \in \mathrm{Z}_2$ 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 $\sum_{wt(\mathbf{v}_i) \ odd} h_{n-1}(\mathbf{v}_i) = 2^{n-3}$, $l_{n-1}(\mathbf{x}) = h(\mathbf{x} \oplus \mathbf{a})$, $n \geq 2$, 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 $\sum_{wt(\mathbf{v}_i) \ even} h_{n-1}(\mathbf{v}_i) = 2^{n-3}$ 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.