New Analysis Methods on Strict Avalanche Criterion of S-Boxes

Phyu Phyu Mar, Khin Maung Latt

Abstract—**S**-boxes (Substitution boxes) are keystones of modern symmetric cryptosystems (block ciphers, as well as stream ciphers). **S**-boxes bring nonlinearity to cryptosystems and strengthen their cryptographic security. They are used for confusion in data security An **S**-box satisfies the strict avalanche criterion (SAC), if and only if for any single input bit of the **S**-box, the inversion of it changes each output bit with probability one half. If a function (cryptographic transformation) is complete, then each output bit depends on all of the input bits. Thus, if it were possible to find the simplest Boolean expression for each output bit in terms of the input bits, each of these expressions would have to contain all of the input bits if the function is complete. From some important properties of **S**-box, the most interesting property **SAC** (Strict Avalanche Criterion) is presented and to analyze this property three analysis methods are proposed.

Keywords— S-boxes, cryptosystems, strict avalanche criterion, function, analysis methods**.**

I. Introduction

SBOX -boxes are lookup tables that map **n** bits to **m** bits (see Figure 1). There are several ways of constructing good S-boxes for ciphers, as well as several ways of measuring them. Many block ciphers are based on the old Shannon idea of the sequential application of ***confusion*** and ***diffusion***. Typically, confusion is provided by some form of substitution ("**S**-boxes"). So the obvious question is whether some substitutions are better than others. The obvious answer is "Yes," because one possible substitution maps every value onto itself, just as though there were no substitution at all. So the hunt was on for measures which would distinguish between "bad" and "good" substitutions, and for techniques to construct "good" substitutions. But since weakness measures are related to attacks, new attacks often imply a need for new measures.

S-boxes are quite important components of modern symmetric cryptosystems (in particular, block ciphers) in the sense that **S**-boxes bring nonlinearity to block ciphers and strengthen their cryptographic security. An **S**-box is said to satisfy the strict avalanche criterion(SAC), if and only if for any single input bit of the **S**-boxes, the inversion of it changes each output bit with probability one half. In DES-like cryptosystems (for example: DES, FEAL[First Encryption Algorithm], Multi2, LOKI, etc) , a substitution box (**S**-box) is implemented by a logic circuit or a table lookup memory and a permutation is implemented by a once-to-once wiring, **S**-boxes bring nonlinearity to cryptosystems and strengthen their cryptographic security. This means that an **S**-box plays the most important role in DES-like cryptosystems.

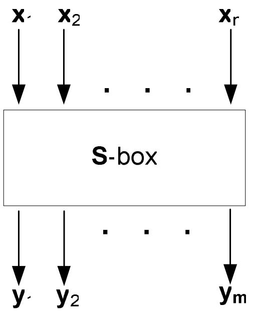

Fig 1. Substitution box (S-box)

II. Notation and Basic Definition

Z	- the set of integers
Z_2^n	- the **n**-dimensional vector space over the finite field Z_2 = **GF**(2)
⊕	- the addition over Z_2^n, or, the bit-wise exclusive-or (XOR)
wt (.)	- Hamming weight function
|.|	- the cardinality of a set or the absolute value of a real number

For a positive integer **n**, define $\mathbf{C}_1^{(n)}, \mathbf{C}_2^{(n)}, \ldots, \mathbf{C}_n^{(n)} \in \mathbf{Z}_2^n$ by

$$\mathbf{C}_1^{(n)} = [0, 0, \ldots, 0, 0, 1]$$

$$\mathbf{C}_2^{(n)} = [0, 0, \ldots, 0, 1, 0]$$

$$\ldots$$

$$\mathbf{C}_n^{(n)} = [1, 0, \ldots, 0, 0, 0]$$

Intuitively, $C_i^{(n)}$ means an **n** dimensional vector with Hamming weight 1 at the **i**-th position.

X.**W** denotes the dot product of **X** and **W**, defined as $\mathbf{X \cdot W = x_1 w_1 \oplus x_2 w_2 \oplus \ldots x_n w_n}$.

For a function **f**: $\mathbf{Z}_2^n \to \mathbf{Z}_2^m$, denoted by $\mathbf{f_i}$ $(1 \le \mathbf{j} \le \mathbf{m})$ the function $\mathbf{Z}_2^n \to \mathbf{Z}_2$ such that

Phyu Phyu Mar is a PhD candidate and Demonstrator of Department of Engineering Physics, Mandalay Technological University, Mandalay, Union of Myanmar.

Professor Dr Khin Maung Latt is Head of Department of Engineering Physics, Mandalay Technological University, Mandalay, Union of Myanmar.

$$\mathbf{f(X) = (f_m(X), f_{m-1}(X), ..., f_2(X), f_1(X))} .$$

III. Strict Avalanche Criterion (SAC)

S-box in general is considered to be a table lookup memory or a Boolean function from $\mathbf{Z_2^n}$ to $Z_2^m(n \geq m)$ as shown in Figure 1. Feistel [5] has proposed one important criterion to design cryptographic function.

A.*Definition 1 (Avalanche effect)*. A function $f : Z_2^n \rightarrow Z_2^m$ exhibits the avalanche effect if and only if

$$\sum_{\mathbf{x} \in \mathbf{Z_2^n}} \mathbf{wt(f(X) \oplus f(X \oplus C_i^n)) = m2^{n-1}}$$

for all **i** ($1 \leq \mathbf{i} \leq \mathbf{n}$).

This means that an average of one half of the output bits change whenever a single input bit is complemented.

Kam and Davida [6] proposed the completeness condition that each output bit depends on all input bits of the substitution.

B.*Definition 2 (Completeness)*. A function **f**: $f : Z_2^n \rightarrow Z_2^m$ is complete if and only if

$$\sum_{\mathbf{x} \in \mathbf{Z_2^n}} \mathbf{f(X) \oplus f(X \oplus C_i^n)} > (0,0,...,0)$$

for all $\mathbf{i} (1 \leq \mathbf{i} \leq \mathbf{n})$ where both the summation (∑) and the greater-than (>) are component-wise over $\mathbf{Z^m}$.

This means that each output bit depends on all of the input bits. Thus, if it were possible to find the simplest Boolean expression for each output bit in terms of the input bits, each of these expressions would have to contain all of the input bits if the function is complete.

Ayoub [2] suggested the probabilistic completeness of substitution-permutation encryption networks. Webster and Tavares introduced the Strict Avalanche Criterion (SAC) in order to combine the notions of the completeness and the avalanche effect.

C.*Definition 3 (SAC, Strong S-box)*. We say that a function $\boldsymbol{f : Z_2^n \rightarrow Z_2^m}$ satisfies the SAC, or **f** is a strong **S**-box, if for all
i ($1 \leq \mathbf{i} \leq \mathbf{n}$) there hold the following equations:

$$\sum_{\mathbf{x} \in \mathbf{Z_2^n}} \mathbf{f(X) \oplus f(X \oplus C_i^n)} = (2^{\mathbf{n}-1}, 2^{\mathbf{n}-1}, ..., 2^{\mathbf{n}-1}).$$

In particular, if $\boldsymbol{f : Z_2^n \rightarrow Z_2^m}$ satisfies the SAC, **f** is called a Boolean strong **S**-box.

If a function satisfies the SAC, each of its output bits should change with a probability of one half whenever a single input bit is complemented. Clearly, a strong **S**-box is complete and exhibits the avalanche effect.

If some output bits depend on only a few input bits, then, by observing a significant number of input-output pairs such as chosen plaintext attack, a cryptanalyst might be able to detect these relations and use this information to aid in the search for the key. Strong **S**-boxes play significant roles in cryptography.

IV. Proposed Methods

In this section three analysis methods for strict avalanche criterion (SAC) of **S**-box are proposed:

1) Analysis of the frequency of various hamming weight (*Avalanche effect*);
2) Analysis of the frequency of various differential value ΔY (*completeness*);
3) Analysis of hamming weights according to the bit position (*Strong S-box*).

A. Analysis of the frequency of various hamming weight

Input: S-box with length m, where m is number of bits.
Output: frequency of various hamming weight.

Algorithm.

Step 1: Choose a random number $\mathbf{x} \in \mathbf{Z_2^m}$. Find corresponding output value of **S**-box;

$$\mathbf{y = S(x)}.$$

Step 2: Choose another random number $\mathbf{x}' \in \mathbf{Z_2^m}$. Find corresponding output value of **S**-box;

$$\mathbf{y' = S(x')}.$$

Step 3: Compute the differential value of outputs;

$$\Delta\mathbf{y = y \oplus y'}.$$

Step 4: Find hamming weight **w** in the differential value **ΔY** of outputs.

Step 5: Repeat again from step 1 to step 4 for necessary count of testing.

Step 6: Analyze the frequency of various hamming weight **w**.

Example 1. (Let us assume, that testing count is equal 10)

The length of **S**-box: **n** = 4.

The contents of input **S**-box = { 7, 0, 4, 12, 11, 8, 2, 1, 15, 3, 5, 6, 9, 10, 13, 14 }.

1) $\mathbf{x} = 6$, $\mathbf{x}' = 10$;
$\mathbf{y} = \mathbf{S}(6) = 2$, $\mathbf{y}' = \mathbf{S}(10) = 5$, $\Delta\mathbf{y}_1 = 2 \oplus 5 = 7 = 0111$; $\mathbf{w} = 3$.
2) $\mathbf{x} = 3$, $\mathbf{x}' = 2$;
$\mathbf{y} = \mathbf{S}(3) = 12$, $\mathbf{y}' = \mathbf{S}(2) = 4$, $\Delta\mathbf{y_2} = 8 = 1000$; $\mathbf{w} = 1$.
3) $\mathbf{x} = 7$, $\mathbf{x}' = 2$;
$\mathbf{y} = \mathbf{S}(7) = 1$, $\mathbf{y}' = \mathbf{S}(2) = 4$, $\Delta\mathbf{y_3} = 5 = 0101$; $\mathbf{w} = 2$.
4) $\Delta\mathbf{y_4} = 15 = 1111$; $\mathbf{w} = 4$.
5) $\Delta\mathbf{y_5} = 0 = 0000$; $\mathbf{w} = 0$.
6) $\Delta\mathbf{y_6} = 6 = 0110$; $\mathbf{w} = 2$.
7) $\Delta\mathbf{y_7} = 15 = 1111$; $\mathbf{w} = 4$.
8) $\Delta\mathbf{y_8} = 2 = 0010$; $\mathbf{w} = 1$.
9) $\Delta\mathbf{y_9} = 11 = 1011$; $\mathbf{w} = 3$.
10) $\Delta\mathbf{y_{10}} = 9 = 1001$; $\mathbf{w} = 2$.

TABLE I
FREQUENCY TABLE FOR HAMMING WEIGHT

weights	frequency

0	1
1	2
2	3
3	2
4	2

Table (1) shows hamming weights and their frequency. Their relation can be displayed by bar charts as shown in figure 2a. The figure 2b shows the result of a good S-box calculation from DES S-box, while the figure 2c shows the result of a poor S-box from user defined S-box. The result of methods 1 shows that if the frequency are variously random, this result shows the testing S-box is poor, if the frequency of testing result is as like as figure b, this result shows the S-box has a good properties of Strict avalanche criterion.

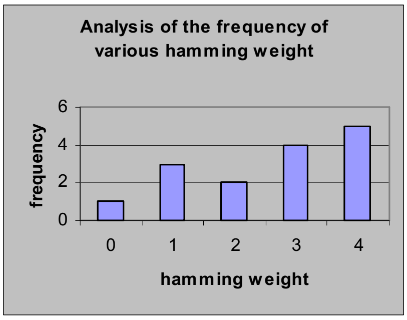

Fig. 2 (a). Testing result of example 1

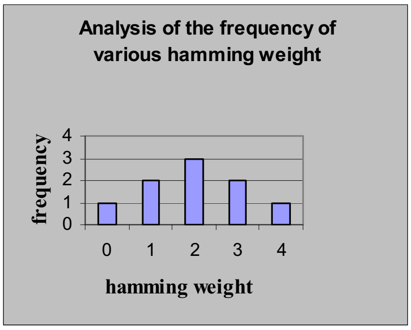

Fig 2 (b). Testing result of a good S-Box

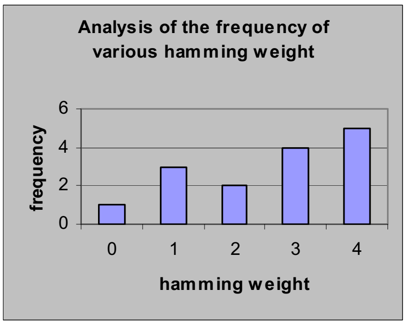

Fig. 2 (c). Testing result of a poor S-Box

B. Analysis of the frequency of various differential values ΔY

Input: **S**-box with length **m**, where **m** is number of bits.
Output: frequency of various differential value **ΔY**.

Algorithm.

Step 1: Choose a random number $\mathbf{x} \in \mathbf{Z}_2^{\mathbf{m}}$. Find corresponding output value of **S**-box;

$$\mathbf{y} = \mathbf{S}(\mathbf{x}).$$

Step 2: Choose another random number $\mathbf{x}' \in \mathbf{Z}_2^{\mathbf{m}}$. Find corresponding output value of **S**-box;

$$\mathbf{y}' = \mathbf{S}(\mathbf{x}').$$

Step 3: Compute the differential value of outputs;

$$\Delta\mathbf{y} = \mathbf{y} \oplus \mathbf{y}'.$$

Step 4: Repeat again from step 1 to step 3 for necessary count of testing.

Step 5: Analyze the frequency of various differential values **ΔY**.

Example 2. (Let us assume, that testing count is equal 20)

The length of **S**-box: **n** = 4.

The contents of input **S**-box = {7, 0, 4, 12, 11, 8, 2, 1, 15, 3, 5, 6, 9, 10, 13, 14}.

$\mathbf{x} = 6$, $\mathbf{x}' = 10$; $\mathbf{y} = 2$, $\mathbf{y}' = 5$, $\Delta\mathbf{y}_1 = 7$.
$\mathbf{x} = 3$, $\mathbf{x}' = 2$; $\mathbf{y} = 12$, $\mathbf{y}' = 4$, $\Delta\mathbf{y}_2 = 8$.
$\mathbf{x} = 7$, $\mathbf{x}' = 2$; $\mathbf{y} = 1$, $\mathbf{y}' = 4$, $\Delta\mathbf{y}_3 = 5$.
$\Delta\mathbf{y}_4 = 15$, $\Delta\mathbf{y}_5 = 0$, $\Delta\mathbf{y}_6 = 6$, $\Delta\mathbf{y}_7 = 8$, $\Delta\mathbf{y}_8 = 2$,
$\Delta\mathbf{y}_9 = 11$, $\Delta\mathbf{y}_{10} = 9$, $\Delta\mathbf{y}_{11} = 3$, $\Delta\mathbf{y}_{12} = 4$,
$\Delta\mathbf{y}_{13} = 10$, $\Delta\mathbf{y}_{14} = 2$, $\Delta\mathbf{y}_{15} = 1$, $\Delta\mathbf{y}_{16} = 13$,
$\Delta\mathbf{y}_{17} = 14$, $\Delta\mathbf{y}_{18} = 9$, $\Delta\mathbf{y}_{19} = 12$, $\Delta\mathbf{y}_{20} = 5$.

TABLE 2
FREQUENCY TABLE FOR DIFFERENTIAL OUTPUT

differential values	frequency
0	1
1	1
2	2
3	1
4	1
5	2
6	1
7	1
8	2
9	2
10	1
11	1
12	1
13	1
14	1
15	1

Table (2) shows differential values and their frequency. Their relation can be displayed by bar chart as shown in figure 3a. The figure 3b shows the result of a good S-box from the DES S-box, while the figure 3c shows the result of poor S-box from the user defined S-box. The result of methods 2 shows that if the frequencies of differential output are random, the testing S-box is poor, if the frequency of testing result is as like as figure b, this result shows the S-box has good properties of completeness properties.

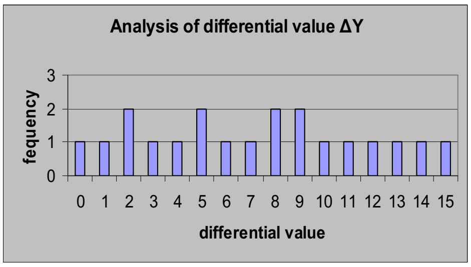

Fig. 3 (a). Testing result of example 2

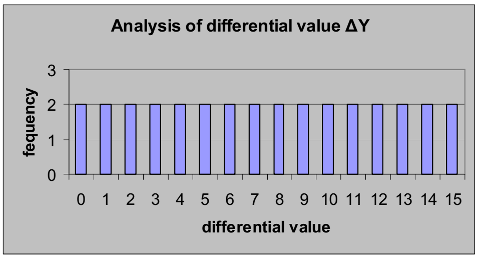

Fig. 3 (b). Testing result of a good S-Box

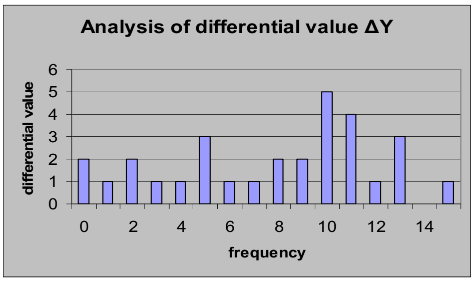

Fig. 3 (c). Testing result of a poor S-Box

C. Analysis of hamming weights according to the bit position
Input: **S**-box with length **m**, where **m** is number of bits.
Output: hamming weights according to the bit position.
Algorithm.

Step 1: Choose a random number $\mathbf{x} \in \mathbf{Z}_2^{\mathbf{m}}$. Find corresponding output value of **S**-box;

$$\mathbf{y} = \mathbf{S}(\mathbf{x}).$$

Step 2: Choose another random number $\mathbf{x}' \in \mathbf{Z}_2^{\mathbf{m}}$. Find corresponding output value of **S**-box;

$$\mathbf{y}' = \mathbf{S}(\mathbf{x}').$$

Step 3: Compute the differential value of outputs;

$$\Delta\mathbf{y} = \mathbf{y} \oplus \mathbf{y}'.$$

Step 4: Repeat again from step 1 to step 3 for necessary count of testing.

Step 5: Analyze the hamming weight according to the bit position of resulting differential values **ΔY**.

Example 3. ***(****Let us assume, that testing count is equal 10****)***

The length of **S**-box: **n** = 4.

The contents of input **S**-box = {7, 0, 4, 12, 11, 8, 2, 1, 15, 3, 5, 6, 9, 10, 13, 14}.

1) $\mathbf{x} = 6$, $\mathbf{x}' = 10$;
 $\mathbf{y} = 2$, $\mathbf{y}' = 5$, $\Delta\mathbf{y_1} = 7 = 0111$.
2) $\mathbf{x} = 3$, $\mathbf{x}' = 2$;
 $\mathbf{y} = 12$, $\mathbf{y}' = 4$, $\Delta\mathbf{y_2} = 8 = 1000$.
3) $\mathbf{x} = 7$, $\mathbf{x}' = 2$;
 $\mathbf{y} = 1$, $\mathbf{y}' = 4$, $\Delta\mathbf{y_3} = 5 = 0101$.
4) $x = 12$, $\mathbf{x}' = 2$, $\Delta\mathbf{y_4} = 15 = 1111$.
5) $x = 9$, $\mathbf{x}' = 9$, $\Delta\mathbf{y_5} = 0 = 0000$.
6) $x = 6$, $\mathbf{x}' = 2$, $\Delta\mathbf{y_6} = 6 = 0110$.
7) $x = 8$, $\mathbf{x}' = 7$, $\Delta\mathbf{y_7} = 15 = 1111$.
8) $x = 1$, $\mathbf{x}' = 3$, $\Delta\mathbf{y_8} = 2 = 0010$.
9) $x = 10$, $\mathbf{x}' = 1$, $\Delta\mathbf{y_9} = 11 = 1011$.
10) $x = 8$, $\mathbf{x}' = 1$, $\Delta\mathbf{y_{10}} = 9 = 1001$.

TABLE 3

HAMMING WEIGHT FOR DIFFERENTIAL OUTPUT ACCORDING TO BIT POSITION

ΔY				
decimal	binary			
7	0	1	1	1
8	1	0	0	0
5	0	1	0	1
15	1	1	1	1
0	0	0	0	0
6	0	1	1	0
15	1	1	1	1
2	0	0	1	0
11	1	0	1	1
9	1	0	0	1
hamming weight	5	5	6	6

Table (3) shows differential values and their frequency. Their relation can be displayed by bar charts as shown in figure 4a. The figure 4b shows the result of a good S-box from DES S-box, while the figure 4c shows the result of poor S-box from User defined S-box. The result of methods 3 shows that if the frequencies of hamming weight of differential output according to the bit position are random, the testing S-box is poor, if the frequency of testing result is as like as figure b, this result shows the S-box has good properties of strong S-box.

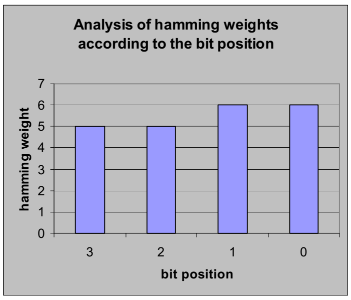

Fig. 4(a). Testing result of example 3

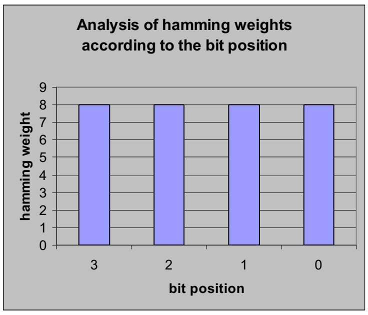

Fig. 4 (b). Testing result of a good S-Box

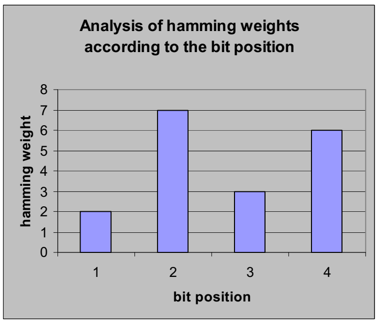

Fig. 4(c). Testing result of a poor S-Box

V. CONCLUSION

The three methods, as mentioned above, it can be seen that the new analysis methods are simpler and easier than the old methods. Because the old methods used the mathematical equations and required repeated test in order to make sure their results provide a secure data diffusion in the S-box. This new methods used only simple calculation and its result can easily be evaluated by the bar graphs. This bar graph represents whether the testing S-box is poor or good. This new methods allow to be implemented in high-level software such as C^{++}, Matlab and likewise. The three main properties of a strong S-box were tested in this work and the results show that the present S-boxes are applicable in data security algorithm.

REFERENCES

[1] C. Adams, S. Tavares*, "The Structured Design of Cryptographically Good S-boxes*", To appear in J. of Cryptology, 1990.

[2] F. Ayoub, *"Probabilistic Completeness of Substitution-Permutation Encryption Network",* IEEE, Vol.129, E, 5, pp195-199, Sep., 1982.

[3] E.F. Brickell, J.H. Moore, M.R. Purtill,*"Structures in the S-boxes of the DES*", Proc. of CRYPTO'86, Springer-Verlag, pp. 3-8, 1986.

[4] J. Daemen, V. Rijmen, "AES Proposal: Rijndael", Document version 2, 03-09-99, *http://csrc.nist.gov/CryptoToolkit/aes/rijndael/Rijndael.pdf*.

[5] H. Feistel, *"Cryptography and Computer Privacy",* Scientific American, Vol.228, No.5, pp 15-23, 1973.

[6] J.B. Kam, G.I. Davida, *"Structured Design of Substitution-Permutation Encryption Network",* IEEE Trans. on Compute. Vol.C-28, No.10, pp.747-753, Oct., 1979.

[7] Kwangjo KIM, *"A Study on the Construction and Analysis of Substitution Boxes for Symmetric Cryptosystems*", Dissertation submitted to the Division of Electrical and Computer Engineering for the Degree of Doctor of Philosophy, December 25, 1990, http://citeseer.ist.psu.edu/336097.html.