Chapter 5: Analysis of Confusion in S-Boxes through SAC Test: 1 Bit Alteration †

The security of a block cipher, in general, depends on the characteristics of Substitution box (S-box) because it is the only non-linear component of a block cipher. The design and characteristics of S-boxes of a block cipher are central measures of resistance against all cryptanalytical techniques. So, analysis of an S-box before it could be implemented is very much important. This chapter involves one such novel bit-level confusion analysis test for S-boxes. The method has been used to test the strengths of S-boxes of DES and AES using Strict Avalanche Criterion (SAC) matrix.

5.1 Introduction

The strength of an encryption algorithm lies in the confusion and diffusion properties in it. The only non-linear component included in many block ciphers, called Substitution box (S-box), provides this functionality to the algorithm. The qualities of the algebraic and statistical properties of an S-box define the strength of an algorithm and thus are a vital source of analysis.

This chapter discusses a novel approach for statistical analysis of the properties of an S-box. The approach includes the analysis of Strict Avalanche Criterion (SAC) matrix using statistical measures like coefficient of variance, correlation, standard deviation and mean of standard deviation.

5.2 Strict Avalanche Criterion (SAC)

The Strict Avalanche Criterion (SAC) was introduced by A.F. Webster and S. E. Tavares [7] and according to them "If a function is to satisfy the strict avalanche criterion, then each of its output bits should change with a probability of one half whenever a single input bit x is complemented to $\bar{x}$ **[7]**. SAC can be mathematically represented as:

$$\forall x, y | H(x, y) = 1,$$
$$avg(H(F(x), F(y))) = {}^{n}/_{2}$$

i.e., a very small change in input produces reasonable change in output. Hence, if F has the avalanche effect then the Hamming Distance between the input and the output generated by changing one of its bits should be close to $n/2$.

† This chapter is referenced from the published research paper: "A Novel Technique for Analysing Confusion in S-Boxes, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 6, June 2016, ISSN(Online): 2320-9801, ISSN (Print) : 2320-9798, DOI: 10.15680/IJIRCCE.2016. 0406253".

Avalanche Effect: Feistal **[27]** has proposed a function $\boldsymbol{f}: \mathbb{Z}_2^{\boldsymbol{n}} \rightarrow \mathbb{Z}_2^{\boldsymbol{m}}$ that defines the avalanche effect if and only if

$$\sum_{x \in \mathbb{Z}_2^n} wt\big(f(X) \oplus f(X \oplus C_i^n)\big) = m2^{n-1} \text{ for all } i\ (1 < i < n).$$

Completeness: Kam and Davida **[28]** proposed the completeness condition that each output bit depends on all input bits of substitution by a function $\boldsymbol{f}: \mathbb{Z}_2^{\boldsymbol{n}} \rightarrow \mathbb{Z}_2^{\boldsymbol{m}}$ is complete if and only if

$$\sum_{x \in \mathbb{Z}_2^n} \boldsymbol{f}(\boldsymbol{X}) \oplus \boldsymbol{f}(\boldsymbol{X} \oplus \boldsymbol{C_i^n}) > (0,0,\ldots,0) \text{ for all } i\ (1 < i < n).$$

5.3 Proposed Method

The steps involved in the proposed technique [47] include:

- Generate SAC matrix by altering each bit of the input to the S-box.
- Analysis of Coefficient of Variance of the generated SAC matrix.
- Analysis of frequency of various avalanche effects from the generated SAC matrix.
- Analysis of frequency of various differential values from the generated SAC matrix.
- Analysis of Hamming Weights according to bit positions from the generated SAC matrix.

5.4 Algorithm

Algorithm – Bit Level Confusion Analysis of S-Box:

Input: S-box with length *m*, where *m* is the number of bits.

Step 1: Choose a random number $P \in \mathbb{Z}_2^m$. Find the corresponding output value of S-box:

$$C = S(P)$$

Step 2: Change the $P_i's$ to find their corresponding output values $C_i's$. P_i is the new P by altering i^{th} bit to its complement.

Step 3: $X_i = C_i \oplus C$ is stored in the i^{th} row of a matrix of size $x \times y$ where x is the number of bits of P and y is the number of bits of C.

Step 4: Find the number of 1s in each column (j).

As it is evident from the algorithm that there are n bits in the block and at every instance only one bit has been changed, as a result there are n blocks such that $H(P, P_i[i] = 1)$, where H denotes the Hamming Distance. The test finds the number of times a particular bit has changed when each of the n newly generated S-box outputs with respect to the S-box output of the original input. Ideally each bit should change

n/2 times, because if a particular bit has changed with very high or very low frequency, it might become a vulnerable target for an attack.

The number of times a particular bit has changed is referred to as the vulnerability factor of the bit. An extremely low or extremely high vulnerability factor associated with a particular bit may act as an area of exploitation for the attackers.

The time taken to construct the X matrix is $O(n^2)$ and time taken to determine the bit-vulnerability factors is also $O(n^2)$. So, the time complexity of the algorithm is $O(n^2)$.

5.5 Experimental Results

Coefficient of Variance Analysis of generated SAC of S-boxes of DES:

S-boxes are the only non-linear elements in DES design. There are 8 S-boxes with the structure of matrix 4 × 16. Each of the unique selection function $S_1, S_2, S_3, \ldots, S_8$, takes a 6-bit block as input and yields a 4-bit block as output. The structure of S-box input and output of DES is given in Figure 5.1 **[29]**

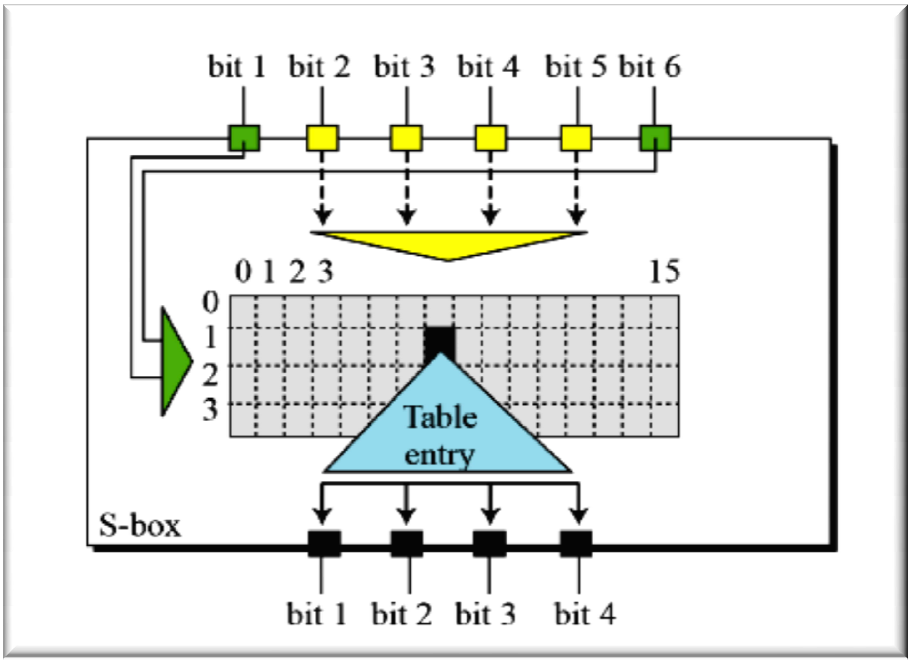

Figure 5.1: Structure of S-box input and output of DES

Evaluation of output is:

Step1: S= matrix 4 × 16, values from 0 to 15.

Step2: B (6-bit input) = $\boldsymbol{b_1}b_2b_3b_4b_5\boldsymbol{b_6}$

a. $\boldsymbol{b_1b_6}$ ⟶ r = row of the matrix (2 bits: 0, 1, 2, 3).

b. $b_2b_3b_4b_5$ ⟶ c = column of the matrix (4 bits: 0,1,2,….,15).

Step3: C (4-bit output) = Binary representation S(r,c).

A SAC matrix has been generated from every possible input of every S-box and 1s of every column of every output of every S-box have been calculated. Coefficients of Variance of all S-boxes have been calculated on the number of 1s available in each column. Some of the generated SAC matrices have been given as examples in Table 5.2 and Table 5.3. For all possible 64 inputs and outputs of all 8 S-boxes, the sum of 1s of every column is termed as V-vector.

SAC Matrix of Input – 0 of S-box 0			
1	0	1	0
1	1	0	1
1	1	0	0
0	0	1	1
1	0	1	0
1	1	1	0
V-vector of Input – 0			
5	3	4	2

Table 5.2. SAC Matrix of input 0 of S-box 0 of DES

SAC Matrix of Input – 0 S-box 1			
1	1	1	1
0	1	1	0
1	0	0	1
0	1	1	1
1	1	1	0
1	1	0	0
V-vector of Input – 0			
4	5	4	3

Table 5.3. SAC Matrix of input 0 of S-box 1 of DES

5.5.1 Experimental Results for DES S-boxes

The results of the proposed test on Data Encryption Standard (DES) **[30]** S-boxes are depicted in Table 5.4.

S-box#	Observed Mean	Variance	Standard Deviation	Coefficient of Variance
0	3.71875	1.436523	1.198551	0.322299
1	3.796875	1.036865	1.018266	0.268185
2	3.9375	1.214844	1.1022	0.279924
3	3.6875	1.027344	1.01358	0.274869
4	3.796875	1.224365	1.10651	0.291427
5	3.90625	1.209961	1.099982	0.281595
6	3.9375	1.417969	1.190785	0.302422
7	3.75	0.953125	0.976281	0.260342

Table 5.4. Experimental Results for DES S-box

The column graph of co-variance (CV) for every S-box of DES is shown in Figure 5.3. The column graph of Standard Deviation for every S-box of DES is shown in Figure 5.4.

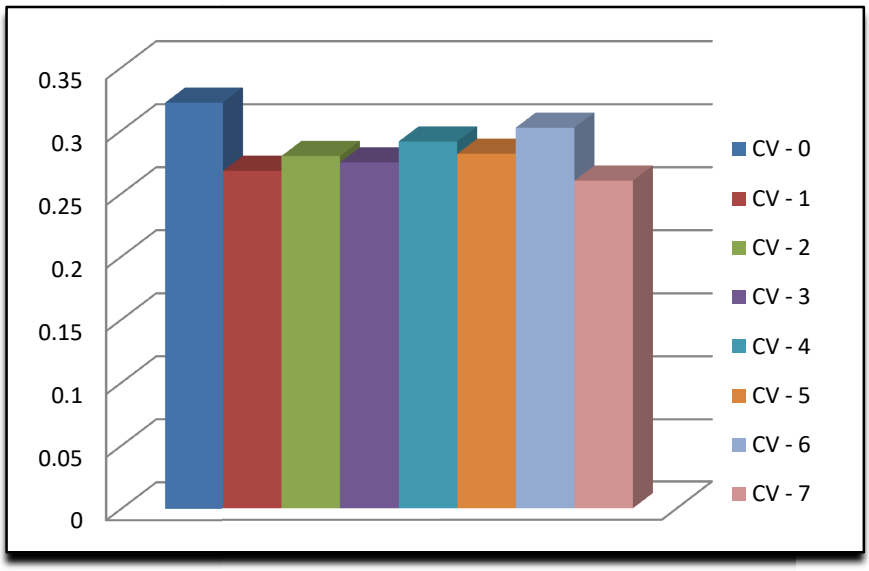

Figure 5.2: Analysis of Co-variance for S-boxes of DES

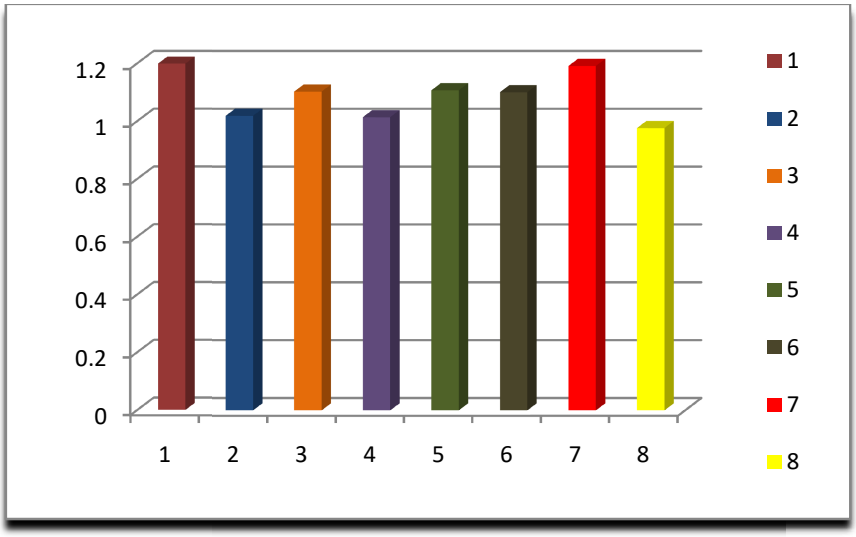

Figure 5.3: Analysis of Standard Deviation for S-boxes of DES

The Comparison line graph of Standard Deviation (SD) and Coefficient of Variance (CV) of every S-box of DES is shown in Figure 5.4.

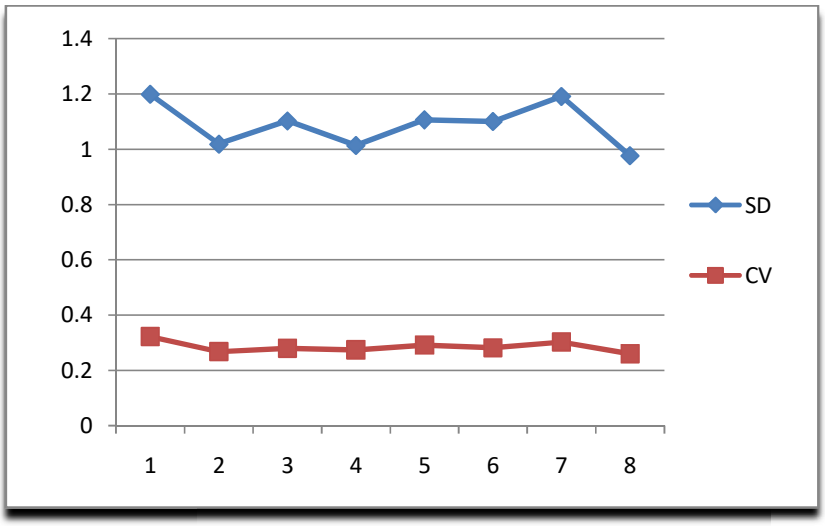

Figure 5.4: Comparison of SD and CV for S-boxes of DES

Coefficient of Variance Analysis of generated SAC of S-box of AES:

The single S-box is the only non-linear element in AES design. The S-box has the matrix structure of 16×16 and elements of the s-box are individually an hex value. For each of the unique inputs of 8-bit block it yields an 8-bit blocks as output and evaluation of output is:

Step1: S= matrix 16×16, in hex, values from 0 to F.

Step2: B (8-bit input) = $b_1 b_2 b_3 b_4 \mathbf{b_5 b_6 b_7 b_8}$

a. $b_1 b_2 b_3 b_4$ ⟶ r = row of the matrix for output.

b. $\mathbf{b_5 b_6 b_7 b_8}$ ⟶ c = column of the matrix for output

Step3: C (8-bit output) = Binary representation of hex value of S(r, c).

A SAC matrix has been generated from every possible input of every S-box and 1s of every column of every output of every S-box have been calculated. Coefficients of Variance of all S-box have been calculated on the number of 1s available in each column. Some of the generated SAC matrices are given as examples in Table 5.5 and Table 5.6. For all 6 inputs and outputs of S-box, the sum of 1s of every column is termed as V-vector.

Input : 11000011							
Original Output : 00101110							
1	1	1	0	0	1	0	1
0	0	0	1	0	0	1	1
1	1	1	0	1	1	1	0
1	0	0	1	1	0	0	1
1	1	1	0	0	0	0	0
0	0	1	1	1	0	0	1
1	0	0	0	0	1	1	1
1	1	0	1	1	0	1	0
V-vector of input 11000011 - $(195)_{10}$							
6	4	4	4	4	3	4	5

Table 5.5. SAC Matrix of input 11000011 of AES S-box

Input : 10101010							
Original Output : 10101100							
0	0	0	1	1	0	1	0
0	1	1	1	1	0	0	0
1	0	0	0	0	0	0	1
0	0	0	0	1	0	1	1
1	1	0	0	0	1	0	1
0	0	0	1	1	0	1	1
0	0	1	1	1	1	0	1
1	0	0	1	1	1	0	1
V-vector of input 10101010 – $(170)_{10}$							
3	2	2	5	6	3	3	6

Table 5.6. SAC Matrix of input 10101010 of AES S-box

5.5.2 Experimental Results for AES S-box

The results of the proposed test on the Advanced Encryption Standard (AES) **[31]** S-box are depicted in Table 5.7.

Input	Observed Mean	Variance	Standard Deviation	Coefficient of Variance
$(195)_{10}$	4.25	0.6875	0.829156	0.195096
$(170)_{10}$	3.75	2.4375	1.561249	0.416333
$(204)_{10}$	4.125	1.60937	1.268611	0.307542
$(105)_{10}$	3.125	1.60937	1.268611	0.405956
$(45)_{10}$	3.625	3.98437	1.99609	0.550645
$(210)_{10}$	4.375	1.23437	1.111024	0.253948

Table 5.7. Experimental Results for AES S-box

The column graph of co-variance of every input for S-box of AES is shown in Figure 5.5. The column graph of Standard Deviation of every input for S-box of AES is shown in Figure 5.6.

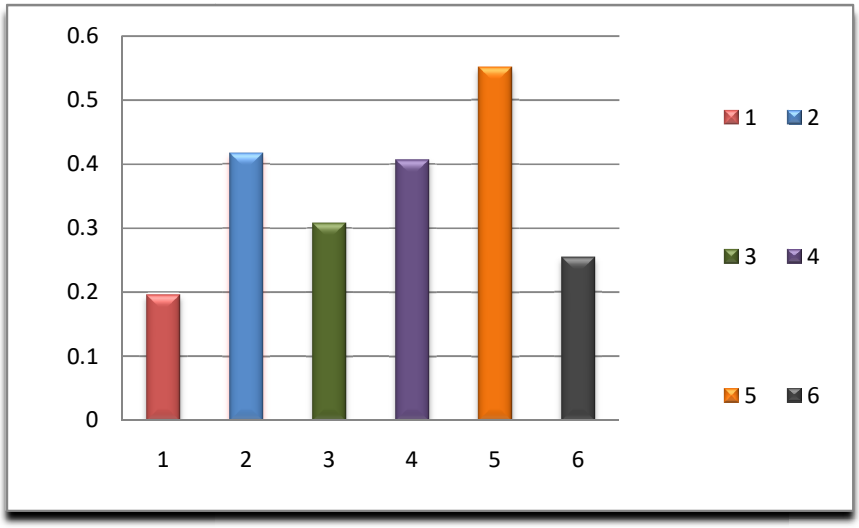

Figure 5.5: Analysis of Co-variance of Inputs of S-box of AES

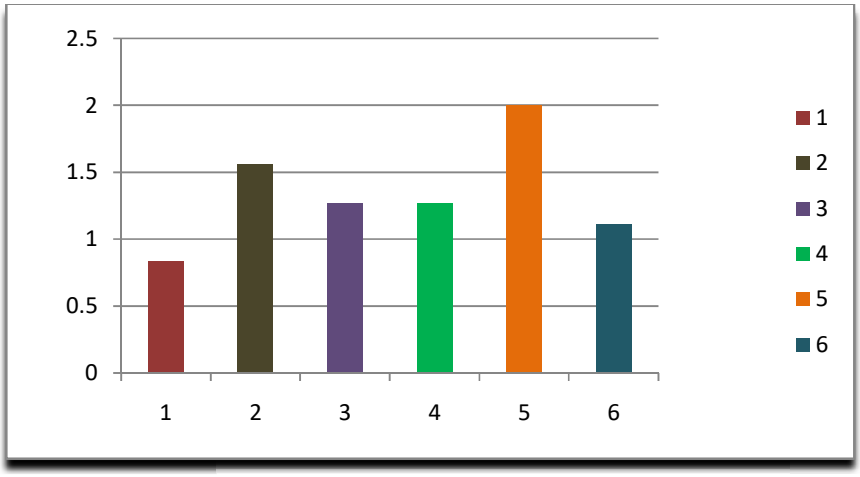

Figure 5.6: Analysis of Standard Deviation of Inputs of S-box of AES

The Comparison line graph of Standard Deviation (SD) and Coefficient of Variance (CV) of every input for S-box of AES is shown in Figure 5.7.

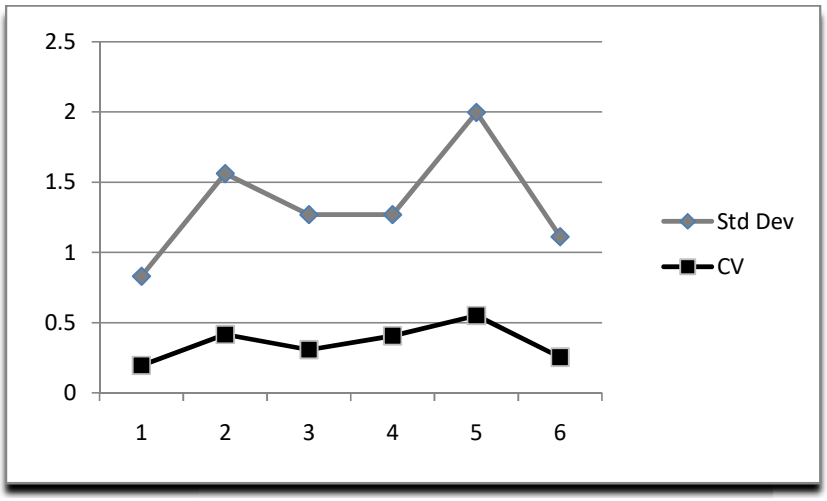

Figure 5.7: Comparison of SD and CV for S-box of AES

5.6 Discussion

After computing results of the proposed test algorithm on every possible input for every S-box of DES and six numbers of random inputs for single S-box of AES, coefficient of variance has been calculated as a statistical measure of the dispersion of data point in a data series around the mean. The coefficient of variance (CV) is calculated as $CV = Std.Deviation\ /\ Mean$.

By the analysis of coefficient of variance in case of S-boxes of DES, it is closer to 0.3 that is lower end of the spectrum and indicates that S-boxes of DES perform well with respect to the test.

On the other hand the result obtained from the single S-box of AES varies from 0.2 to 0.55 which is also lower end of the spectrum and indicates that S-box of AES perform well with respect to the proposed test algorithm.

5.7 Conclusion

As discussed in section 5.1, confusion and diffusion are the two fundamental aspects of a block cipher for analyzing its cryptographic strength. There are many methods to test diffusion but this proposed test could very serve well to analyze the confusion characteristic too, and may be included as a part of a comprehensive test suite for analyzing cryptographic strength of block ciphers.