AN IMPACT OF S-BOX BOOLEAN FUNCTION PROPERTIES TO STRENGTH OF MODERN SYMMETRIC BLOCK CIPHERS

Roman Oliynykov, Oleksandr Kazymyrov

Abstract: S-boxes are among the most important parts in modern symmetric ciphers. Generation of substitutions with good cryptographic properties is one of the meaningful components in the new ciphers' design. Boolean function properties of several modern block ciphers' S-boxes are analyzed. An overview of existing requirements to S-box generation is given, and the most significant for the strength of symmetric block ciphers are chosen.

Keywords: Boolean function, S-box, differential cryptanalysis, linear cryptanalysis, XSL attack

Introduction

Modern symmetric block ciphers uses an idea of interleaving of linear and non-linear layers. For most of ciphers, S-boxes are the only one non-linear component determined the cryptographic strength to various analytical attacks. Up to date an important question of generation of substitutions with optimal characteristics to prevent all known cryptanalytical attacks remains open. There exist several approaches to generation of S-boxes with good cryptographic properties, and methods based on the analysis of Boolean function properties are among widespread ones. Another method for generation of S-boxes with perfect differential and linear properties based on the power function in the finite field was proposed in[1]. Such type of S-boxes was used in many ciphers, including AES/Rijndael. However, further analysis have shown that this approach leads to existence of overdefined system of equations with low degree for S-boxes, and potential vulnerability of the cipher to XSL attack[2]. Further analysis of Boolean function properties have shown that such type of S-boxes uses the single non-linear Boolean function with different linear transformation of its arguments[3]. In this article there are identified most important properties of S-boxes Boolean function properties with respect to strength of symmetric block ciphers to known cryptanalytical attacks, and correlation of such properties with other known S-box characteristics.

Cryptographic properties of Boolean functions

Let $V_2^n = GF(2)^n$ be the N-dimensional vector space of binary N-tuples. An N-variable Boolean function $f(x)$ is a mapping from $f(x) : V_2^N \to V_2$ where $x = (x_{n-1}, x_{x-2}, ..., x_0)$.

Let $f(x)$ be N-variable Boolean function. Then Hadamard weight is calculated as follows:

$$hw(f) = \sum_{x=0}^{2^n-1} f(x) \tag{1}$$

Let $f(x)$ and $g(x)$ are N-variable Boolean functions. Then hamming distance between this functions can be calculated as follows:

$$hd(f,g) = \sum_{x=0}^{2^n-1} f(x) \oplus g(x) \tag{2}$$

The algebraic normal form (ANF) of an N-variable Boolean function $f(x)$ is written as follows:

$$\begin{aligned} f(x) = a \oplus a_0 x_0 \oplus a_1 x_1 \oplus \ldots \oplus a_{n-1} x_{n-1} \oplus a_{01} x_0 x_1 \oplus a_{02} x_0 x_2 \oplus \ldots \oplus \\ \oplus a_{(n-2)(n-1)} x_{n-2} x_{n-1} \oplus \ldots \oplus a_{012\ldots n-1} x_0 x_1 x_2 \ldots x_{n-1} \end{aligned} \tag{3}$$

The algebraic degree of a Boolean function $f(x)$, denoted by $deg(f)$, is defined to be the number of variables in the largest product term of the function's ANF having a non-zero coefficient.

The Walsh Hadamard transform (WHT) of the truth table of a Boolean function $f(x)$, denoted by $W(\omega)$, is a measure of the correlation between a function and the set of all linear functions. It is defined by

$$W(x) = \sum_{x=0}^{2^n-1} (-1)^{f(x) \oplus l_\omega(x)} \tag{4}$$

where Boolean function of the form $l_\omega(x) = \omega \cdot x = \omega_0 x_0 \oplus \omega_1 x_1 \oplus \ldots \oplus \omega_{n-1} x_{n-1}$ is a linear function of N variables.

Function $f(x)$ of N-variable is called balanced if

$$hw(f) = 2^{n-1} \tag{5}$$

The nonlinearity of an N-variable Boolean function $f(x)$, denoted by $NL(f)$, is the minimum distance to the set of all N-variable affine functions. Following equation describes relationship between the nonlinearity of Boolean function $f(x)$ and the Walsh Hadamard transform:

$$NL(f) = \frac{1}{2}\left(2^n - W_{\max}\right) \tag{6}$$

The autocorrelation function of the truth table of a Boolean function $f(x)$, denoted by $r_f(\alpha)$, is a measure of the directional derivative of a function for an input shift in the direction of α overall $x \in V_2^n$. It is defined by

$$r_f(\alpha) = \sum_{x=0}^{2^n-1} (-1)^{f(x) \oplus f(x \oplus \alpha)} \tag{7}$$

Let $|AC|_{\max}$ denote the maximum absolute autocorrelation value derived from the autocorrelation function $r_f(\alpha)$, then:

$$|AC|_{\max} = \max_{\alpha} |r_f(\alpha)| \tag{8}$$

Let σ denote the sum-of-square indicator, also derived from the autocorrelation function $r_f(\alpha)$, then:

$$\sigma = \sum_{\alpha=0}^{2^n-1} r_f^2(\alpha) \tag{9}$$

An *N*-variable Boolean function *f*(*x*) is said to satisfy strict avalanche criterion (SAC) if for every *s* is true that:

$$\begin{cases} hw(s) = 1; \\ \sum_{x=0}^{2^n-1} f(x) \oplus f(x \oplus s) = 2^{n-1}. \end{cases} \tag{10}$$

An *N*-variable Boolean function *f*(*x*) is said to satisfy the propagation criteria of degree *k*, *PC*(*k*), accounting a non-zero vector $\alpha \in V_2^n$ if

$$\begin{cases} 1 \le hw(\alpha) \le k; \\ \sum_{x=0}^{2^n-1} f(x) \oplus f(x \oplus \alpha) = 2^{n-1}. \end{cases} \tag{11}$$

An *N*-variable Boolean function *f*(*x*) is m^{th}-order correlation immunity, denoted by *CI*(*m*), if for every ω is true that

$$\begin{cases} 1 \le hw(\omega) \le m; \\ W(\omega) = 0. \end{cases} \tag{12}$$

An *N*-variable Boolean function, *f*(*x*), which is both balanced and has m^{th}-order correlation immunity, is known as an *m*-resilient Boolean function.

Cryptographic properties of substitution

Let $S = (f_M, f_{M-1}, \ldots, f_1)$ be an *N*x*M* substitution (S-box) where f_i ($i = M, M-1, \ldots, 1$) are *N*-variable Boolean functions. Let g_j be the set of linear combinations of f_i ($i = M, M-1, \ldots, 1$) (which includes the functions f_i). Then:

- the nonlinearity of *S*, denoted by $N(S_{N,M})$ can be expressed as follows:

$$N(S_{N,M}) = \min_g \{NL(g_j)\} \quad \left(j = 1,2,\ldots,2^{M-1}\right) \tag{13}$$

- the algebraic degree of *S*, denoted by $\deg(S_{N,M})$, is defined as:

$$\deg(S_{N,M}) = \min_g \{\deg(g_j)\} \quad \left(j = 1,2,..,2^{M-1}\right) \tag{14}$$

- the maximum absolute autocorrelation value of S, denoted by $\left|AC(S_{N,M})\right|_{\max}$, is defined as:

$$\left|AC(S_{N,M})\right|_{\max} = \max_g \{r_{g_j}(\alpha)\} \tag{15}$$

where $\alpha \in \left\{1,2,..,2^{N-1}\right\}$ and $\left(j = 1,2,..,2^{M-1}\right)$.

- *S* is said to satisfy strict avalanche criterion if every g_j $\left(j = 1,2,..,2^{M-1}\right)$ satisfies SAC.

- *S* is said to satisfy propagation criteria of order *k*, if every g_j $\left(j = 1,2,..,2^{M-1}\right)$satisfies *PC*(*k*).

- *S* is t^{th}-order correlation immunity if all g_j $\left(j = 1,2,..,2^{M-1}\right)$ are t^{th}-order correlation immunity Boolean functions.

- *S* is a *t*-resilient s-box, if all g_j $\left(j = 1,2,..,2^{M-1}\right)$are *t*-resilient Boolean functions.

- *S* is balanced if every g_j $\left(j = 1,2,..,2^{M-1}\right)$ is balanced.

Let *S* be an *N*x*M* S-box. Let δ be the largest value in the differential distribution table of the S-box (not taking into account first row and column). Then

$$\delta = \max_{\alpha \in V_2^N, \alpha \neq 0, \beta \in V_2^M} \max \#\left\{x \mid S(x) \oplus S(x \oplus \alpha) = \beta\right\} \tag{16}$$

Let *S* be an *N*x*M* S-box. Let λ be the largest value in the linear distribution table of the S-box (not taking into account first row and column). Then

$$\lambda = \max_{\alpha \in V_2^N, \alpha \neq 0, \beta \in V_2^M} \left| \#\left\{x \mid \bigoplus_{s=0}^{N}\left(x[s] \cdot \alpha[s]\right) = \bigoplus_{t=0}^{N}\left(S(x)[t] \cdot \beta[t]\right)\right\} \right| \tag{17}$$

where $\xi[s]$ is a s^{th} bit of ξ.

Let *r* be the number of equations of a system, which described an *N*x*M* S-box. Then:

$$r = \sum_{x=0}^{d} \binom{N+M}{d} - Rank(A) \tag{18}$$

where A is a matrix, which contains all input and output bits, and all combinations up to degree d.

We said that S-box is good, if $Rank(A) = 2^N$. It is easy to see, that if $Rank(A) = 2^N$ then d has a maximum value[4]. Increasing of d up to the maximum value increases the system dimension and consequently more memory is required for processing the system of equations.

Next three properties were given in works[5, 6]. Let S be an NxN S-box. Let η be a number of inversions in substitution. Then

$$\left|\eta - \frac{N(N-1)}{4}\right| \le a\frac{\sqrt{N^3}}{6} \tag{19}$$

where a=1 in general case.

Let S be an NxN S-box. Let ς be a number of cycles in substitution. Then

$$\left|\varsigma - \ln(N)\right| \le a\sqrt{\ln(N)} \tag{20}$$

where a=1 in general case.

Let S be an NxN S-box. Let θ be a number of increases in substitution. Then

$$\left|\theta - \frac{N}{2}\right| \le a\sqrt{\frac{n}{12}} \tag{21}$$

where a=1 in general case.

Analysis of criteria for generation of substitutions

After publishing papers on differential[7] and linear[8] cryptanalysis there were suggested criteria for S-box selection for protection of such types of attacks. In [1] there were proposed a method for generation of perfect non-linear S-boxes with the minimal achievable probability of non-trivial difference transformation and linear approximation. The algebraic attack is based on possibility to describe all cipher by a system of equations. In contrast to differential and linear cryptanalysis, attacker needs several pairs of plaintext and corresponding ciphertext[9, 10, 11, 12].

The relationship between difference distribution table and table of Boolean functions' autocorrelation is showed in[13]. In paper only the boundary between autocorrelation and maximum of differential table is described

$$\delta > 2^{N-M} + 2^{-M} AC(S_{N,M})_{\max}. \tag{22}$$

This relationship suggests that the rate of the autocorrelation is important for choosing a S-box, but now there is no method for determining direct correspondence between the autocorrelation of Boolean functions and a the maximum of differential table.

Nonlinearity of Boolean functions is directly related with maximum value of approximating table as $NL(S_{N,M} = 2^{N-1} - \lambda)$. Clearly, these two criteria are commutative. Thus, if we know the nonlinearity of Boolean functions, we can uniquely determine the maximum of approximating table and vice versa.

The avalanche effect is an important property of S-boxes. It shows how many of output bits of S-box change with changing of a subset of the input bits. This property is necessary to prevent statistical attacks and connected with differential properties of S-box. If a substitution satisfies the strict avalanche criterion then changing of a single input bit changes exactly half of output bits. Propagation criterion includes strict avalanche criterion. From equation (11) is easily seen that when *k*=1 PC becomes SAC. Thus, we should choose only propagation criterion while generating substitution.

The correlation immunity criteria of Boolean functions are used in primary to protect stream ciphers against the correlation attacks. In [14] χ^2-attack on ciphers RC-5 and RC-6 is shown. But till now there are no papers showed application of such type of attacks to block ciphers.

In addition, in recent papers there was showed that the value of the differential does not depend on non-linear S-box properties after sufficient number of rounds[15]. However, it is noted that the S-box influence on dynamic indicators, i.e. how quickly the cipher comes to the theoretical average maximum of differential table. This means that S-box with good properties is much more effective than randomly generated.

The table 1 contains results of calculating the number of equations describing substitution, properties of Boolean functions, the maximum of differential and linear table. Values were calculated for substitutions that are used in ciphers AES, Camellia, Labyrinth and Kalina[16, 17].

Table 1 – S-box characteristics

	AES	Cam.	Lab.	K0	K1	K2	K3	K7
Balance of BF	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NL	112	112	112	96	98	96	96	96
AC	32	32	32	88	88	88	104	96
SSI	133120	133120	133120	244480	252928	259456	291712	251392
PC	0	0	0	0	0	0	0	0

CI	0	0	0	0	0	0	0	0
Resilient	0	0	0	0	0	0	0	0
Degree	7	7	7	7	7	6	7	7
Number of eq.	39	39	39	441	441	441	441	441
Degree of eq.	2	2	2	3	3	3	3	3
Inversions	Yes	No	Yes	No	No	No	No	Yes
Cycles	Yes	No	Yes	Yes	Yes	No	Yes	Yes
Increses	Yes	No	Yes	No	No	Yes	Yes	Yes
MDT	4	4	4	8	8	8	8	8
MLT	16	16	16	32	30	32	32	32
Bijective	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Results given in the table 1 show that S-boxes of modern symmetric block ciphers does not satisfy to many well-known requirements for Boolean functions. Substitutions from current standards of encryption (such as AES, Camellia) do not satisfy propagation criterion (including the SAC), correlation immunity, but have the minimal value of non-trivial transformation in differential and linear approximating tables.

According to strength to known cryptanalytical attacks, there can be selected the following criteria for S-boxes of symmetric block ciphers:

- nonlinearity of Boolean functions (or maximum of approximating table);
- maximum of differential table;
- maximum degree of the system representing the S-box.
- propagation criterion.

It also should be mentioned that it is necessary to research the relationships between these criteria.

Conclusions

There exist several different approaches to generation of cryptographically strength S-boxes for symmetric block ciphers. An approach based on analysis of Boolean function properties takes into consideration requirements of non-linear components of stream ciphers for construction of S-boxes for block ciphers. It is shown, that S-boxes of widely used modern symmetric block ciphers does not satisfy to several well-known requirements to Boolean

functions, like propagation criterion and correlation immunity. But other properties of S-box Boolean functions, like nonlinearity and autocorrelation, connected to strength of the cipher to differential and linear attacks.

Notes:

1. Kaisa Nyberg,"Perfect nonlinear S-boxes". *Advances in Cryptology, EUROCRYPT*, 1991, 378-386.
2. Nicolas T. Courtois and Gregory V. Bard, "Algebraic Cryptanalysis of the Data Encryption Standard". *Lecture Notes in Computer Science, Volume 4887/2007*, 2007, 152-169.
3. J.Fuller and W. Millan, "Linear Redundancy in S-Boxes". *Lecture Notes in Computer Science, Volume 2887/2003*, 2003, 74-86.
4. R. Oliynykov and O. Kazymyrov, "The choice of S-boxes for symmetric cryptographic algorithms based on the analysis of algebraic properties". *Journal of Kharkov National University. Mathematical simulation. Information Technology. Control Systems.*, 2010, 177-179.
5. I.V. Lysytska, A.S. Koriak, et al. "The selection criteria of random substitution tables for symmetric enciphering algorithms" / *Abstracts of XXVIth General Assembly. Toronto, Ontario Canada* 1999, 204.
6. V.I. Dolgov, Lysytsky K.E. "Random substitutions in cryptography" / *Radioelectronic and computer systems, №5* 2010, 79-85.
7. E. Biham and A. Shamir, "Differential Cryptanalysis of DES-like Cryptosystems" *Advances in Cryptology, Springer-Verlag* 1990, 2–21.
8. Mitsuru Matsui, "Linear cryptanalysis method for DES cipher". *In Workshop on the theory and application of cryptographic techniques on Advances in cryptology (EUROCRYPT '93), Tor Helleseth (Ed.). Springer-Verlag New York, Inc., Secaucus, NJ, USA* 1993, 386-397.
9. R.V Oliynykov and O.V. Kazymyrov, "Algebraic properties of the key schedule of block symmetric cipher "Kalina"". *Radioelectronic and computer systems, №5*, 2010, 61-66.
10. R.V. Oliynykov and O.V. Kazymyrov, "Construction of the overdefined system of equations to describe symmetric block cipher "Labyrinth"". *Applied electronics, Volume 8, №3*, 2009, 247-251.
11. N. Courtis and J. Pieprzyk, "Cryptanalysis of Block Cipher with Overdefined System of Equations". *Lecture Notes in Computer Science, Volume 2501/2002* 2002, 267-287.
12. N. Courtois, A. Klimov, J. Patarin and A. Shamir, "Efficient Algorithms for solving Overdefined System of Multivariate Polynomial Equations". *Lecture Notes in Computer Science, Volume 1807/2000* 2000, 392-407.
13. X. Zhang, Y. Zheng and H. Imai, "Relating Differential Distribution Tables to Other Properties of of Substitution Boxes". *Designs, Codes and Cryptography, Volume 19, Number 1*, 1998, 45-63.
14. T. Shimoyama, K. Takeuchi and J. Hayakawa. "Correlation Attack to the Block Cipher RC5 and the Simplified Variants of RC6". *In Proceedings of the 3rd AES Conference (AES3)*, 2000.
15. V.I. Dolgov, I.V. Lysytska adn O. Kazymyrov, "Variations on a theme cipher Rijndael". *Applied electronics, Volume 9, №3*, 2010, 321-325.

16. I.D. Gorbenko, V.I.Dolgov, R.V.Oliynykov et al., “Advanced symmetric block cipher "Kalina"- the basic main aspects and specifications”. *Applied electronics, Volume 6, №2*, 2007, 195-208.

17. S. Golovashich, “Specification of block symmetric encryption algorithm "Labyrinth"”. *Information processing systems, Volume 4*, 2007, 230-240.