Using search algorithm of affine equivalent S-boxes set for their quality assessment

Nikolay Pavlovich Borisenko

Tóm tắt— **Bài viết này tập trung vào các vấn đề đánh giá chất lượng của các S-hộp bằng cách đánh giá độ phức tạp tìm kiếm của tập các S-hộp tương đương affine với một S-hộp đã biết. Các S-hộp tuyến tính với các hàm thành phần tuyến tính và các hàm liên quan đến nó là có tính chất mật mã yếu. Các S-hộp tốt là các S-hộp có các hàm Bool thành phần phải có khoảng cách lớn so với tập các hàm Bool tuyến tính. Do vậy các S-hộp sở hữu độ phi tuyến cực đại là cần được lưu ý về số lượng nhiều (hay ít) các S-hộp mà tương đương affine với nó.**

Abstract— **The article focuses on the problem of quality assessment of S-boxes by evaluating the search complexity of the set of S-boxes that are affine equivalent to the specified S-box. It is known, that S-boxes with linear and close to them coordinate functions are considered to be cryptologically weak and vice versa. That is, in good S-boxes coordinate Boolean functions must be maximum remote from linear ones. However, S-boxes possessing maximum nonlinearity are notable for the strength of their set of affine equivalence class that under other equal conditions enables to present the given S-box as a large (or less) number of S-boxes that are affine equivalent to it.**

Keyword— affine equivalent S-box set; Hamming distance; S-box nonlinearity; mapping; transformation; component (coordinate) Boolean functions.

I. INTRODUCTION

In modern telecommunications systems symmetric block ciphers have found wide application as ciphering algorithms whose cryptographic security is greatly determined by their nonlinear elements. Currently, one of the most prospective techniques of implementation of nonlinear transformations is using substitution boxes or S-boxes.

Choosing quality S-box can be considered to be one of the most critical aspects of ciphering algorithms, that's why their choice is paid keen interest to. Practical experience shows that at present the set of techniques available for quality assessment of substitution boxes does not allow to choose the best S-box with respect to various techniques of cipher analysis and to methods of their hardware and software realizations [2].

To assess the quality of S-boxes it is often necessary to establish the fact of belonging of several S-boxes to one class of affine equivalence. The analyzed problem is solved by various methods. The most popular is the algorithm developed in [1]. In the process of studying the behavior of the given algorithm it became clear that for simple S-boxes (linear or weakly linear) it quickly converges to "representative", as for complex (substantionally nonlinear) the speed of convergence to "representative" increases substantionally. However, an attempt to build an S-box complexity scale through search complexity of its representative by means of the given algorithm proved erroneous due to difficulty of introducing criterion of S-box quality assessment. Besides, the given algorithm requires additional computational resources for testing membership of S-boxes received during recurrent iteration, initial class of affine equivalence.

Notwithstanding the listed drawbacks, the algorithm described above allows to come to conclusion that available criteria for assessing S-boxes are not sufficient for adequate substantiation of their quality. For example, S-boxes with equal index numbers of their nonlinearity have different algorithm order of convergence. Thus on the basis of data received during the algorithm analysis described in [1], as one of the methods of S-box quality assessment, using the complexity of the algorithm of searching for the representative of affine equivalent S-boxes set is proposed.

II. PROBLEM STATEMENT

Initial data:

S-box of arbitrary structure.

It is necessary to develop the algorithm of searching for a representative of a set of affine equivalent S-boxes, having no drawbacks listed above and on its basis:

- To formulate criteria of S-boxes quality assessment;
- To compare characteristics of different S-boxes on the basis of different criteria, including the proposed one.

III. S-BOX NONLINEARITY

One of the most important S-box quality criteria is its nonlinearity which can be determined as the minimum Hamming distance between component functions defining S-box as well as their linear combinations and the whole set of their affine functions [2].

$$nl(\Phi) = \min(\mathrm{dist}(f, A_n)) \qquad (1)$$

where Φ is mapping defined by specified *S-box*, $f \in \langle f_S \rangle$ is case of all linear combinations of S-box component functions, A_n is set of all affine functions.

It is necessary to note, that the given distance is closely related with c Walsh-Hadamard coefficients, calculated for the considered S-box. It is obvious that nonlinearity distance is greater when these coefficients are modulo [2].

$$nl(f) = 2^{n-1} - \tfrac{1}{2}\max_{\alpha \in V_n}\left|W_f(\alpha)\right| \qquad (2)$$

In the extreme case, when all Walsh-Hadamard conversion coefficients are minimum and equal modulo, and nonlinearity distance is maximum, component Boolean functions are termed bent-functions.

When comparing S-box used in State Standard P 34.11–2012, State Standard P 34.12–2015 with AES-similar S-box of the same dimension on the given criteria, it should be noted that AES-similar S-box possesses the best characteristics (it has nonlinearity value equal to 112, while for S-box – State Standard P 34.11–2012, State Standard P 34.12–2015 the distance to the nearest linear function is 100).

IV. MODIFICATION OF ALGORITHM TO SEARCH FOR A REPRESENTATIVE OF AFFINE EQUIVALENT S- BLOCK SET

To search for the representative in the algorithm described in [1] one uses the following approach:

The initial S-box having n inputs is presented in the form of truth tables making up its component functions, Table 1. Each line of the table is considered as a number, which makes it possible to solve the problem of searching for the representative by permutation of the lines of the given table. As it was mentioned above, the given approach to searching the representative has serious drawbacks.

To modify the search algorithm of a representative of affine equivalent S-box set, the following model was used. Let be the representative of affine equivalent S-box class. Any S-box of the class can be defined by two transformations of A and B from additive transformation group *(Affine General Linear Group, AGL).*

$$S = B \circ R_0 \circ A. \qquad (3)$$

TABLE 1. RANDOM S-BOX

№ п/п	X8	X4	X2	X1		Y8	Y4	Y2	Y1
0	0	0	0	0		0	0	0	0
1	0	0	0	1		0	0	0	1
2	0	0	1	0		1	0	0	1
3	0	0	1	1		1	1	1	0
4	0	1	0	0		1	1	0	1
5	0	1	0	1		1	0	1	1
6	0	1	1	0		0	1	1	1
7	0	1	1	1		0	1	1	0
8	1	0	0	0		1	1	1	1
9	1	0	0	1		0	0	1	0
A	1	0	1	0		1	1	0	0
B	1	0	1	1		0	1	0	1
C	1	1	0	0		1	0	1	0
D	1	1	0	1		0	1	0	0
E	1	1	1	0		0	0	1	1
F	1	1	1	1		1	0	0	0

Knowing R_0 one can obtain any *S-box* from the given set. Accordingly, each *S-box* of the given set can be matched with unique combination of nonsingular *A* and *B* matrixes. Having grouped *S-boxes* received from R_0 by affine equivalent transformations by value of *A* matrix, let us make up representation model of affine equivalent S-box set (Figure 1).

At the Figure 1, the whole affine equivalent S- block set is shown with the dotted line. Rhombs inside the given set show subsets of S-boxes having the same *A* matrixes obtained from affine equivalent S- block set and all possible values of *B* matrix, that is, adjacent classes of affine equivalent substitutes. Points bellow rhombs are minimum representatives of the given subsets or local representatives of the given adjacent class. Point below the Figure is sought-for R_0 representative of the whole S^0 set.

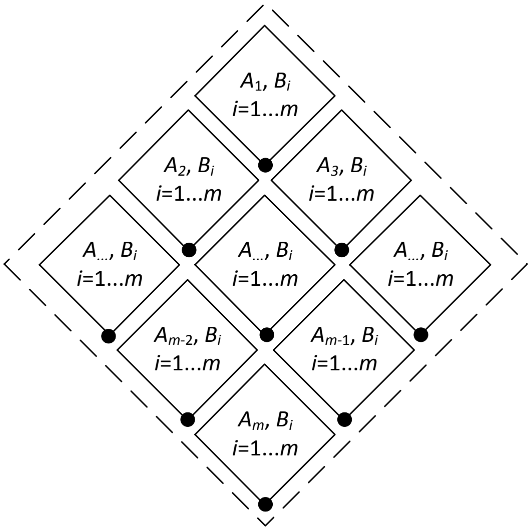


Figure 1. Representation model of affine equivalent S-box set.

As in the algorithm described in [1], *S-box* having *n* inputs, is presented in the form of the truth tables making up its component functions. The main difference of the algorithms consists in the approach to treatment of S-boxes: in this case the truth table forming *S-box*, Table 1, is considered not by lines, but by columns, i.e., as *n* numbers of numerical system nonagenary

$$M = 2^{2^n} \tag{4}$$

At the next stage for component functions making up *S-box*, set of linear combinations and their inversions are built (*L*). The process of building the given set consists of the following stages:

– initialization of the initial empty set *L* = ∅;

– $\forall i,\ 1 \le i < 2^n$:

- *i* is presented in binary form

$$i = (i_1 i_2 \ldots i_n), \forall i_q \in GF(2), q = 1..n; \tag{5}$$

- Components of the set of linear combinations are calculated by a formula:

$$z_i = i_1 p_1 \oplus i_2 p_2 \oplus \ldots \oplus i_n p_n = \left(a_0, a_1, \ldots, a_{2^n-1}\right) \tag{6}$$

$$z_{2^n+i} = z_i \oplus (1,0,\ldots,0) = \left(a_0 \oplus 1, a_1, \ldots, a_{2^n-1}\right) \tag{7}$$

where z_i, z_{2^n+i} are vector indices obtained by means of expressions (6) и (7), *p*1...*pn* – component functions forming *S-box*;

- Obtained components of *L* linear combinations set are added to the set

$$L = L \cup \{z_i, z_{2^n+i}\} \tag{8}$$

As a result, the strength of *L* linear combinations set will be equal to

$$\#L = 2^{n+1} - 2 \tag{9}$$

– For every element of the given set the weight is calculated in accordance with the considerations described above;

– Then from the obtained set *n* minimum linear independent functions for building a new S-box are chosen. Before choosing the recurrent minimum function from the *L* set it is necessary to exclude linear combinations and their inversions for minimum functions chosen earlier, at the expense of which linear independence of component Boolean functions making up the new *S-box* is achieved;

– The weight of the obtained S-box is calculated.

V. DETERMINATION OF THE REPRESENTATIVES NUMBER

Next we search for all possible values of *A* nonsingular matrix, for each obtained value we search for a representative.

The number of *different* representatives of the obtained ones is proposed to be used as criteria of quality assessment of S-box. Let us designate this number as N_{lm}, denoting the number of local representatives.

Theoretically probable number of subsets of affine equivalent *S-boxes*, and, therefore, representatives, is calculated by a formula.

$$N_{lm} \le \prod_{i=1}^{n} (2^{n+1} - 2^i). \tag{10}$$

For different *n* values we shall calculate maximum possible number of representatives and introduce them into Table 2.

TABLE 2. MAXIMUM NUMBER OF REPRESENTATIVES (ADJACENT CLASSES)

N	$\max N_{lm}$
4	322560
5	319979520
6	1290157424640
7	20972799094947840
8	1369104324918194995200
16	5,10259969886794669593834275239 76e+71

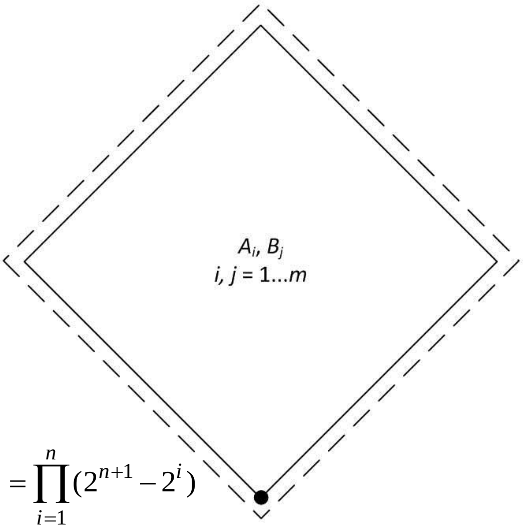
On the basis of the developed algorithm as a criterion to assess S-box quality, the number of received local representatives is proposed to be used. Thus, it was found out, that the affine equivalent S-box case, built for substitution block of n = 4 dimension, according to State Standard P 34.12–2015 has much more local representatives than *AES*-like *S-box* of the same dimension. The experimental results for different affine non equivalent *S-boxes* are presented in Table 3.

TABLE 3. THE EXPERIMENTAL RESULTS FOR DIFFERENT AFFINE NON EQUIVALENT S-BOXES

S-box	nonlinearity	Representative number
D2781EB45AF0963C	0	1
0FA5C369872D4BE1	0	1
01C86F4E3DBA2975	2	5376
019EDB76F2C5A438(*AES*)	4	5376
0123468A5BCF7E9D	4	80640
C462A5B9E8D703F1(ГОСТ)	4	322560

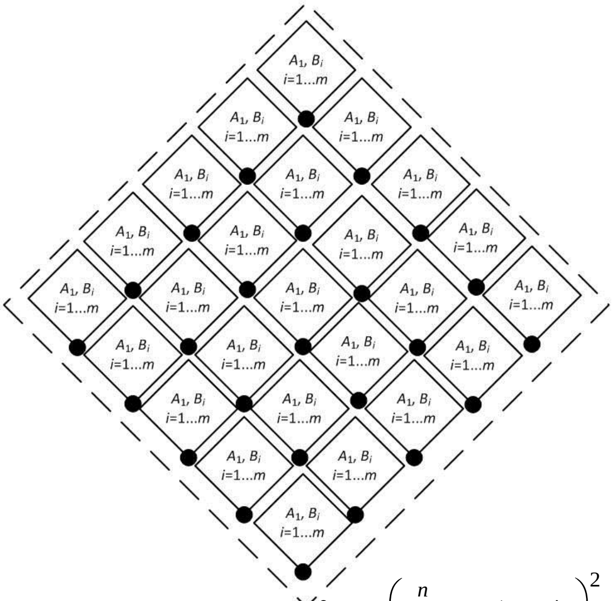
The fact that *S-boxes* having similar nonlinearity indices, may have considerably differing characteristics on some other criteria, for example, the number of local representatives enables us to come to conclusion that the nonlinearity distance for *S-boxes* cannot be considered to be exhaustive quality criteria, and using as a complexity index of algorithm to search for representative of affine equivalent S-boxes set is a vital and perspective task.

On the basis of the developed criteria it is possible to present a model of affine equivalent S-boxes set obtained from the initial linear and S-boxes, having strong nonlinearity (Figure 2 and 3 accordingly).



$$S^0 = \prod_{i=1}^{n}(2^{n+1} - 2^i)$$

Figure 2. Model of presenting affine equivalent S-boxes set



$$S^0_{\max} = \left(\prod_{i=1}^{n}(2^{n+1} - 2^i)\right)^2$$

Figure 3. Model of presenting affine equivalent S-boxes set obtained from essentially nonlinear *S-box*

The developed algorithm makes it possible to verify random S-boxes for affine equivalence the same way as the algorithm described in [1], additional computational resources to check on S-box membership of the initial affine equivalence class after recurrent line permutation being not required.

Moreover, it can be used to estimate the number of adjacent N_{lm} classes in S^0 set, and by comparing the obtained value with its maximum value $\max N_{lm}$ presented in Table 1 to draw a conclusion concerning S-box quality. The less the obtained difference, the better the S-box.

It is appropriate to introduce the definition of coefficient of using substitution set received in accordance with the expression (3), k_{ib}, as an expression

$$k_{ib} = \frac{N_{lm}}{\max N_{lm}}. \quad (11)$$

VI. AN EXAMPLE OF AN APPLICATION OF ALGORITHM TO SEARCH FOR A REPRESENTATIVE OF AFFINE EQUIVALENT S-BOXES SET

Let us take the S-box considered in the previous section as the initial one (Table 1).

We shall determine the weight for each coordinate function and put it into Table 4 (the function *Y*1 is in the bottom line).

TABLE 4. THE INITIAL S-BOX

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	Weight
0	0	1	1	1	1	0	0	1	0	1	0	1	0	0	1	15529
0	0	0	1	1	0	1	1	1	0	1	1	0	1	0	0	7092
0	0	0	1	0	1	1	1	1	1	0	0	1	0	1	0	6090
0	1	1	0	1	1	1	0	1	0	0	1	0	0	1	0	28306

At the next stage, taking into account (4–9), we shall build the set of linear combinations and their inversions (L) making up S-box, for component functions. Its strength is equal to:

$$\# L = 2^{n+1} - 2 = 30 .$$

For each element of the given set in accordance with the above described considerations, the value weight is estimated and put into Table 5.

TABLE 5. THE SET OF LINEAR COMBINATIONS AND THEIR COMPONENT FUNCTIONS INVERSIONS OF S-BOX

31	30	…	12	11	10	9	8	7	6	5	4	3	2	1
1	1	…	0	0	0	0	0	0	0	0	0	0	0	0
0	1	…	0	1	0	1	0	1	0	1	0	1	0	1
1	0	…	1	0	1	0	1	1	0	1	0	1	0	1
0	0	…	0	0	0	1	1	0	0	1	1	1	1	0
0	1	…	0	0	1	0	1	0	1	0	1	1	0	1
0	1	…	1	1	0	0	1	0	1	1	0	0	1	1
0	1	…	1	0	1	1	0	1	0	0	1	0	1	1
1	1	…	1	1	1	0	0	0	0	1	1	1	1	0
1	0	…	0	1	0	0	1	1	0	0	1	0	1	1
0	0	…	0	1	1	0	0	1	1	0	0	1	1	0
1	1	…	0	1	1	1	1	1	1	1	1	0	0	0
1	0	…	1	1	0	1	0	0	1	0	1	1	0	1
1	1	…	1	0	0	1	1	1	1	0	0	1	1	0
0	0	…	1	0	0	0	0	1	1	1	1	0	0	0
1	0	…	0	0	1	1	0	0	1	1	0	0	1	1
0	0	…	1	1	1	1	1	0	0	0	0	0	0	0
41402	53032	…	10013	17905	11107	21051	15529	25324	3198	29990	7092	31064	6090	28306

Then the appropriate functions are sorted out according to their weight growth (Table 6).

TABLE 6. CONCORDANCE OF THE LINEAR COMBINATION NUMBER AND WEIGHT

№	6	2	4	12	10	14	8	11	13	9
Weight	3198	6090	7092	10013	11107	12503	15529	17905	18831	21051
№	15	7	1	5	3	19	21	17	23	31
Weight	24133	25324	28306	29990	31064	34471	35545	37229	40211	41402
№	25	29	27	24	30	26	28	20	18	22
Weight	44484	46704	47630	50006	53032	54428	55522	58443	59445	62337

Let us choose 4 linear independent component functions according to minimum weight value:

1) №6 is chosen; № 6, 22 are excluded;

2) № 2 is chosen; № 2, 4, 18, 20 are excluded;

3) № 12 is chosen; № 8, 10, 12, 14, 24, 26, 28, 30 are excluded;

4) № 11 is chosen.

Let us build minimum weight S-box which is the representative of the given adjacent class of affine equivalent S-box (Table 7).

TABLE 7. OBTAINED S-BOX (REPRESENTATIVE OF AFFINE EQUIVALENT SUBSTITUTION SET)

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	Weight
0	0	0	0	1	1	0	0	0	1	1	1	1	1	1	0	3198
0	0	0	1	0	1	1	1	1	1	0	0	1	0	1	0	6090
0	0	1	0	0	1	1	1	0	0	0	1	1	1	0	1	10013
0	1	0	0	0	1	0	1	1	1	1	1	0	0	0	1	17905

Further all possible values of A nonsingular matrix are sorted out, a representative for each obtained values being sought for.

The number of *different* representatives proved to be equal 5376; in comparison with the maximum value N_{lm} from Table 3 gives us the difference in 317184 (the number of the missing adjacent classes), but the S-box is built on the basis of the algorithm to *AES*-like. Having substituted the obtained values in (11) we have

$$k_{ib} = \frac{N_{lm}}{\max N_{lm}} = \frac{5376}{322560} \approx 0{,}017 .$$

V. CONCLUSION

To estimate the obtained results let us consider hardware implementation of affine equivalent transformations (Figure 4). In compliance with (3), on the left and on the right from the initial $S1$ S-box A matrix and summation with a vector, B matrix and summation with b vector accordingly.

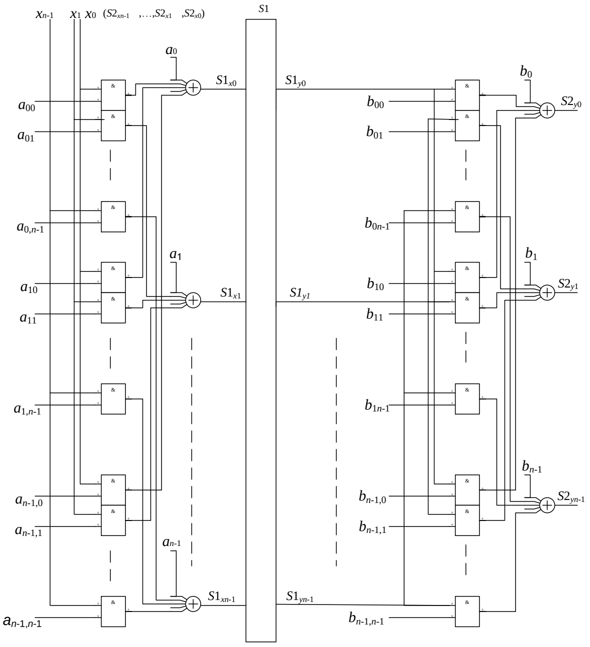


Figure 4. Scheme of hardware implementation of affine equivalent transformations (*n* block dimension)

If one turns the scheme to the right 90° ("transpose"), one obtains classical *SP* block cipher structure (Figure 5).

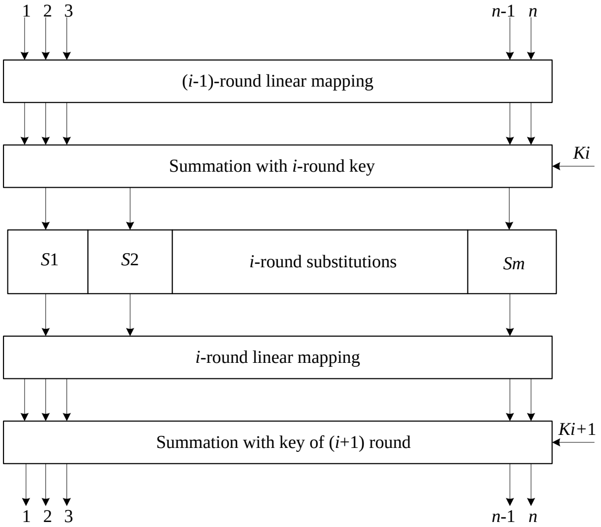

1 2 3
n-1 n
(i-1)-round linear mapping
Summation with i-round key
Ki
S1
S2
i-round substitutions
Sm
i-round linear mapping
Summation with key of (i+1) round
Ki+1
1 2 3
n-1 n

Figure 5. “Transposed” scheme of affine equivalent transformations

One can come to the following conclusion: Only such block cipher structure may be considered optimal from the point of view of maximization $k_{ib} = \dfrac{N_{lm}}{\max N_{lm}}$, which allows implementing one of the set of substitutions $S^0 = \left(\prod_{i=1}^{n}(2^{n+1} - 2^i)\right)^2$ at every round and every S-box of *n* dimension and with equal probability.

REFERENCES

[1]. Alex Birykov, Christophe De Cannere, An Braeken, and Barn Prenell, “A Toolbox for Cryptanalysis: Linear and Affine Equivalent Algorithms”, EUROCRYPTO’2003. Springer, vol. 2656, pp. 33-50, 2003.

[2]. O. A.Logachev, A. A. Salnikov, S. V. Smyshlyaev [and another], “Boolean functions in coding theory and cryptology”, Moscow : LENAND, pp. 576, 2015.

AUTHORS PROFILE

Nikolay Pavlovich Borisenko

Workplace: The Academy of Federal Guard Service of the Russian Federation, 35, Priborostroitelnaya Street, Orel, 302034, Russia.

Email:

science@academ.msk.rsnet.ru

The education process: Borisenko received his Ph.D. degree in Engineering Sciences in Military Academy of communications S.M.Budenogo, St.-Peterburg in 1974.

Research today: Information security, cryptography.