Fast algorithm for computation the parameters of s-boxes that determine the security of SNOW 2.0-like stream ciphers against correlation attacks over extension fields

Mykhailo Poremskyi

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Institute of Special Communication and Information Security, Kyiv, Ukraine

Abstract: The security of SNOW 2.0-like stream ciphers against a wide class of correlation attacks can be evaluated by values of some numerical parameters of s-boxes used in these ciphers. We propose fast algorithm that compute the values of these parameters. The proposed algorithm is based on the fast Hadamard transform and has significantly lower time complexity compared to the previously known one. We also show experimentally, using Monte Carlo method, that for most random 8×8 s-boxes the values of the considered parameters ensure an appropriate security level of SNOW 2.0-like stream ciphers against known correlation attacks.

Keywords: algorithmic cryptanalysis, SNOW 2.0-like stream cipher, correlation attack, security evaluation, fast Hadamard transform, SNOW 2.0, STRUMOK.

1. Introduction

While evaluating security of block, but also some of stream ciphers, against statistic attacks we need to calculate the values of different substitutions (s-boxes), which are used in these ciphers. The most known examples are the maximal elements of the difference distribution table and the linear approximation table of a substitution (see [1, 2], for example) related to the security of the substitution against differential and linear cryptanalysis respectively. The other parameters of such type can be found in [3, 4] and other papers. Fast calculation algorithms are known for some of these parameters, which is important in case, when we need to select substitutions for the cipher from a wide set of substitutions, check different cases for them, that guarantee security of the cipher against known attacks.

In [5] the method of security evaluation of SNOW 2.0-like ciphers against correlation attacks over a finite fields of characteristic 2 is presented. Let's note, that these attacks are proposed in [6] and currently are the most effective against SNOW 2.0 [7]. Another important example of SNOW 2.0-like cipher is STRUMOK, which is national encryption standard of Ukraine [8]. In [5] relations are obtained, indicating the effectiveness of the correlation attack on an arbitrary SNOW 2.0-like cipher with the numerical parameters of s-blocks. These relations can be considered as peculiar generalizations of classical elements of the linear approximation table of a substitution. In [5] the calculation algorithm of such parameters, which time complexity is $O(2^{3t}t)$ for an t-bit to t-bit s-box, is described.

The main result of our paper is more effective algorithm, which time complexity is $O(2^{2t}t)$.

The rest of the paper has the following structure.

In Section 2 we give preliminary information and specify formulation of the problem. In Section 3 the description of the proposed algorithm, proof of its correctness and assessment of its time complexity is given. In Section 4 experimental results of the distribution research of the parameter for random substitutions for $t=8$ are set out. These results allow us to estimate how large is the part of s-boxes that (with fixed values of the rest of parameters of a SNOW 2.0-like cipher) provides a given level of secutiry against known correlation attacks. Finally, we conclude in Section 5.

2. Preliminaries

For any positive integer t denote by V_t the set of all t-dimensional Boolean vectors. For any $\alpha=(\alpha_1,\dots,\alpha_t)$, $\beta=(\beta_1,\dots,\beta_t)\in V_t$ let's define $\alpha\beta=\alpha_1\beta_1\oplus\cdots\oplus\alpha_t\beta_t$, $\alpha\oplus\beta=(\alpha_1\oplus\beta_1,\dots,\alpha_t\oplus\beta_t)$, where $\oplus$ is the XOR operation. For any $a,b\in\mathbf{Z}$, $a\le b$ denote by $\overline{a,b}$ the set $\{a,a+1,\dots,b\}$. In the sequel, we identify an arbitrary vector

$x = (x_1, \dots, x_t) \in V_t$ with the integer $x = 2^{t-1}x_1 + \dots + 2^0 x_t$.

Let $s : V_t \to V_t$ be a substitution on the set V_t (an s-box). For any $a, b \in V_t$, $u, u' \in \{0,1\}$ let's define

$$A_{a,b}^{(s)}(u,u') = $$

$$= 2^{-2t} \sum_{x,y \in V_t : \mathrm{msb}(x+y+u) = u'} (-1)^{(\overset{t}{x}+\overset{t}{y}+u)a \oplus xa \oplus s(y)b}, \qquad (1)$$

where $\mathrm{msb}(x+y+u)$ is the most significant (i.e., the t-th) bit of the sum of integers corresponds to the mentioned t-dimensional Boolean vectors, and $\overset{t}{x}+\overset{t}{y}+u$ is the sum of these integers modulo 2^t.

Let's define

$$n_{a,b}(s) = \max_{u \in \{0,1\}} \{ | A_{a,b}^{(s)}(u,0) | + | A_{a,b}^{(s)}(u,1) | \}, \qquad (2)$$

$$n_{\max}(s) = \max\{ n_{a,b}(s) : (a,b) \in V_t \times V_t \setminus \{(0,0)\} \}. \qquad (3)$$

Let's note, that in the formal replacement of the sum $x+y+u$ into XOR $x \oplus y \oplus u$ in (1) (so that $\overset{t}{x}+\overset{t}{y}+u = x \oplus y \oplus u$ and $\mathrm{msb}(x+y+u) = 0$) parameters $A_{a,b}^{(s)}(0,1)$ and $A_{a,b}^{(s)}(1,1)$ are equal to zero, and the parameters $A_{a,b}^{(s)}(0,0)$ and $A_{a,b}^{(s)}(1,0)$, up to sign, match the expression $2^{-t} \sum_{y \in V_t} (-1)^{ya \oplus s(y)b}$, which is equal to the (a,b)-th element of the linear approximation table of a substitution s (see [2], for example). Thus, (3) can be considered as a $\mathrm{mod}\, 2^t$ generalization of the classical parameter used for evaluation the security of the substitution s against linear cryptanalysis.

In [5], it is shown that the parameter (3) takes an important role in security evaluation of SNOW 2.0-like ciphers against correlation attacks over finite fields of characteristic 2. Each of such ciphers is defined by a set of s-boxes, a linear feedback shift register (LFSR), and an invertible matrix D over a field of order 2^t. In [5], it is shown that the time and the data complexities of a correlation attack from [6] against such type of ciphers can be evaluated using Algorithm 1 directly by the values (3) of the cipher's s-boxes and the branch number $B(D^T)$ [9] of the matrix D^T transposed to D.

Algorithm 1: *security evaluation of SNOW 2.0-like ciphers against correlation attacks from [ZXM].*

Input:

– integer numbers n, p, t;

– s-boxes $s_i : V_t \to V_t$, $i \in \overline{0, p-1}$;

– an invertible $p \times p$-matrix D over the field $\mathbf{GF}(2^t)$;

– a number $k \geq 2$ that is a power of two;

– a divisor r' of the number $r = pt$.

Step 1. Calculate the values

$$n_{\max} = \max\{ n_{\max}(s_i) : i \in \overline{0, p-1} \},$$

$$\Delta_{r'}(k) = (2^{r'} - 1)(n_{\max})^{2k\left\lfloor \frac{B(D^T)}{2} \right\rfloor},$$

using formulas (1), (2), (3).

Step 2. Put $r'' = pt(r')^{-1}$, $l = nr''$, $\theta = 1 + \log k$.

Step 3. For each $l' = 1, 2, \dots, l-1$ calculate

$$m_{r'}(k) = 2(\Delta_{r'}(k))^{-1} l' r' \ln 2,$$

$$T_{r'}(k,l') = (m_{r'}(k))^{\frac{1}{\theta}} k 2^{\frac{r'(l-l')}{\theta}} +$$

$$+ r'(m_{r'}(k) + r' l' 2^{r'l'}) + 2^{r'(l'+1)}.$$

Step 4. Choose $l^* \in \overline{1, l-1}$ such that

$$T_{r'}(k,l^*) = \min\{ T_{r'}(k,l') : l' \in \overline{1, l-1} \}.$$

Output:

– the number l^* of r'-bit words (of the initial state of the LFSR) that are recovered by the attack;

– the average time complexity of the attack, $T_{r'}(k,l^*)$;

– the data complexity of the attack,

$$N_{r'}(k,l^*) = k 2^{\frac{r'(l-l^*)}{\theta}} (2 l^* r' \ln 2)^{\frac{1}{\theta}} (\Delta_{r'}(k))^{-\frac{1}{\theta}}.$$

It is easy to see that the calculation of $n_{\max}(s_i)$ for each $i \in \overline{0, p-1}$ on the first step of Algorithm 1 is based on the definition of this parameter requires $O(2^{4t})$ operations. In [5] a more efficient algorithm with the time complexity $O(2^{3t} t)$ is presented. In this paper we propose a new algorithm that computes the parameter (3) in time $O(2^{2t} t)$.

3. New algorithm

The proposed algorithm is based on the application of the fast Hadamard transform (see [10], for example) and allows to significantly reduce the time complexity of the algorithm from [5].

Algorithm 2: *Fast algorithm for calculation the value of (3).*

Input: s-box $s : V_t \to V_t$.

Step 1. For each $u \in \{0,1\}$ calculate the values

$$D^{(s)}_{u,u'}(x,y) = |\{(z_1,z_2) \in V_t \times V_t : \mathrm{msb}(u + z_1 + z_2) = u', (u + \overset{t}{z_1} + \overset{t}{z_2}) \oplus z_1 = x,\ s(z_2) = y\}| \qquad (4)$$

for all $u' \in \{0,1\}$, $x, y \in V_t$;

Step 2. For all $u,u' \in \{0,1\}$ calculate the values

$$A'^{(s)}_{a,b}(u,u') = 2^{-2t} \sum_{x,y \in V_t} D^{(s)}_{u,u'}(x,y)(-1)^{xa \oplus yb},\ a,b \in V_t,$$

using fast Hadamard transform.

Step 3. For of each pair $(a,b) \in V_t \times V_t \setminus \{(0,0)\}$ calculate

$$n'_{a,b}(s) = \max_{u \in \{0,1\}} \{|A'^{(s)}_{a,b}(u,0)| + |A'^{(s)}_{a,b}(u,1)|\}.$$

Output:

$$n'_{\max}(s) = \max\{n'_{a,b}(s) : (a,b) \in V_t \times V_t \setminus \{(0,0)\}\}.$$

Theorem. *Algorithm 2 calculates the value of* $n_{\max}(s)$ *using* $O(t2^{2t})$ *operations.*

Proof. To prove equality $n_{\max}(s) = n'_{\max}(s)$ it is enough to show, that

$$A'^{(s)}_{a,b}(u,u') = A^{(s)}_{a,b}(u,u')$$

For any $a,b \in V_t$, $u,u' \in \{0,1\}$. Really, it follows from (1) that

$$A^{(s)}_{a,b}(u,u') = 2^{-2t} \sum_{z_1,z_2 \in V_t:\ \mathrm{msb}(x+z_1+z_2) = u'} (-1)^{(\overset{t}{z_1} + \overset{t}{z_2} + u)a \oplus z_1 a \oplus s(z_2) b} =$$

$$= 2^{-2t} \sum_{x,y \in V_t} \sum_{\substack{z_1,z_2 \in V_t: \\ \mathrm{msb}(\overset{t}{z_1}+\overset{t}{z_2}+u) = u', \\ (u + \overset{t}{z_1} + \overset{t}{z_2}) \oplus z_1 = x, \\ s(z_2) = y\}|}} (-1)^{(\overset{t}{z_1} + \overset{t}{z_2} + u)a \oplus z_1 a \oplus s(z_2) b} =$$

$$= 2^{-2t} \sum_{x,y \in V_t} \sum_{\substack{z_1,z_2 \in V_t: \\ \mathrm{msb}(\overset{t}{z_1}+\overset{t}{z_2}+u) = u', \\ (u + \overset{t}{z_1} + \overset{t}{z_2}) \oplus z_1 = x, \\ s(z_2) = y\}|}} (-1)^{xa \oplus s(y) b} =$$

$$= 2^{-2t} \sum_{x,y \in V_t} \sum_{\substack{z_1,z_2 \in V_t: \\ \mathrm{msb}(\overset{t}{z_1}+\overset{t}{z_2}+u) = u', \\ (u + \overset{t}{z_1} + \overset{t}{z_2}) \oplus z_1 = x, \\ s(z_2) = y\}}} (-1)^{xa \oplus yb} =$$

$$= 2^{-2t} \sum_{x,y \in V_t} D^{(s)}_{u,u'}(x,y)(-1)^{xa \oplus yb} = A'^{(s)}_{a,b}(u,u').$$

So, Algorithm 2 actually calculates the value $n_{\max}(s)$.

Further, on step 1 to calculate all values (4) let's build an array (initialized by zeros), which item addresses are all possible triples (u',x,y), where $u' \in \{0,1\}$, $x, y \in V_t$. For all $u \in \{0,1\}$ calculate value (4) using one cycle for (z_1,z_2), adding 1 to the current value of an array element,

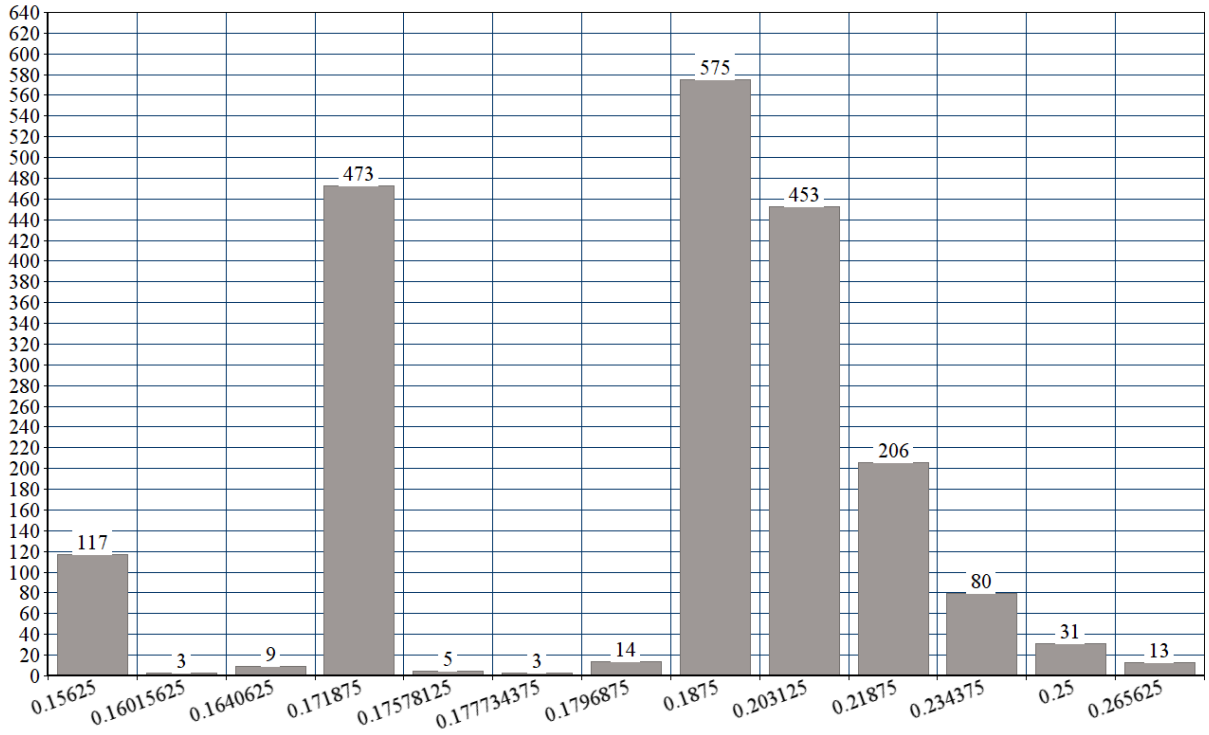

Fig. 1 Histogram built using a sample of 2000 random substitutions

stored at (u', x, y), where

$u' = \mathrm{msb}(u + z_1 + z_2)$, $x = (\overset{t}{u} + \overset{t}{z_1} + z_2) \oplus z_1, y = s(z_2)$.

Obviously, we need $O(2^{2t})$ operations for this.

Finally, on step 2 for all $u,u' \in \{0,1\}$ compute the Hadamard transform of function $D^{(s)}_{u,u'}(x, y)$ where $x, y \in V_t$, which requires $O(t2^{2t})$ operations (see, consequence 5.34 in [10], for example).

Thus, the total complexity of steps 1 and 2, as well as the whole Algorithm 2, will be $O(t2^{2t})$ operations.

The theorem is proved.

4. Experimental research

Algorithm 1 allows us to estimate the part of substitutions, which with fixed values of the rest of parameters of the SNOW 2.0-like cipher, assure its security level against the known correlation attacks. This question is practical important when these substitutions are used as an additional (long-term) key parameters of stream cipher.

For all $x \in (0,1)$ let's denote $F(x)$ the relative number (part) of substitutions s (among all possible on the set V_t) such that $n_{\max}(s) < x$. We need to make statistical evaluation of the parameter $F(x)$ with accuracy ε and confidence $1-\delta$, where $\varepsilon, \delta \in (0,1)$.

To solve this task let use Algorithm 2 with Monte Carlo method. According to the Hoeffding bound [11] the value of $F(x)$ is in interval $(F_N(x) - \varepsilon, F_N(x) + \varepsilon)$ with the probability at least $1-\delta$, where $F_N(x)$ is a number of values $i \in \overline{1, N}$ such that $n_{\max}(s) < x$, $N = \left\lceil 1/2 \cdot \varepsilon^{-2} \ln(2\delta^{-1}) \right\rceil$. The obtained results (for $\delta = 0{,}01$, $\varepsilon = 0{,}037$) are presented on Figure 1.

Now, we can use these results to evaluate the security of SNOW 2.0-like ciphers with the same parameters as in STRUMOK [8] using Algorithm 1 with parameters $n = 16$, $p = 8$, $t = 8$, $r' = t$, and $B(D^T) = p + 1 = 9$. The obtained results are presented in Table 1. We have made the calculations for different values of k (2, 4, 8, 16 etc.), the results are presented for $k = 8$, as it leads to the minimal values of $\log N_{r'}(k, l^*)$ and $\log T_{r'}(k, l^*)$. The calculation carried out on macOS Mojave 10.14, 2.2 GHz Intel Core i7, 16 GB 2400 MHz DDR4, Intel UHD Graphics 630 1536 MB.

Table 1

The values of the parameters that determine the security of a SNOW 2.0-like cipher with randomly generated s-boxes.

$N(n_{\max})$	$n_{\max}$	l^*	$\log T_{r'}(k, l^*)$	$\log N_{r'}(k, l^*)$
575	0,1875	29	249,40	249,38
473	0,171875	30	251,61	249,91
453	0,203125	29	247,16	247,07
206	0,21875	29	245,28	244,94
117	0,15625	30	253,08	252,66
80	0,234375	29	243,99	242,95
31	0,25	28	243,08	243,07
14	0,1796875	29	250,62	250,61
13	0,265625	28	241,34	241,32
9	0,1640625	30	252,17	251,25
5	0,17578125	29	251,25	251,25
3	0,177734375	29	250,93	250,93
3	0,1953125	29	248,25	248,21
2	0,16015625	30	252,58	251,94
2	0,18359375	29	250,00	249,99
1	0,140625	30	255,76	255,70
1	0,14453125	30	255,00	254,91
1	0,1611328125	30	252,47	251,77
1	0,166748046875	30	251,94	250,78
1	0,16796875	30	251,85	250,57
1	0,173828125	30	251,52	249,58
1	0,17431640625	29	251,49	251,49
1	0,193359375	29	248,53	248,50
1	0,21484375	29	245,70	245,46
1	0,2421875	29	243,61	242,00
1	0,28125	28	239,73	239,67
1	0,34375	28	235,55	233,88

In a second column of table are the values of $n_{\max}(s_i)$ calculated for $N = \lceil 1/2 \cdot \varepsilon^{-2} \ln(2\delta^{-1}) \rceil = 2000$ random substitution $s_i : V_t \to V_t$, and in first column – number $N(n_{\max})$ of substitution with the defined value of the parameter $n_{\max}$. The results in last three columns of the Table 1 were calculated using Algorithm 1.

As we can see from Figure 1 and Table 1, when $x = 0{,}19$ the condition $n_{\max}(s) < x$ works for $575 + 14 + 3 + 5 + 473 + 9 + 3 + 117 = 1199$ from total $N = 2000$ randomly generated substitutions S. Together with this the have average complexity of correlation attack against SNOW 2.0-like cipher with such substitutions is at least $2^{251{,}25}$. Therefore, $F_N(x) = \frac{1199}{2000} = 0{,}5995$ and with the confidence at least $1 - \delta$ (that is 99%) the relative number of all substitutions, that guaranty cipher security against correlation attacks at the level at least $2^{251{,}25}$, is within range from $0{,}5995 - \varepsilon$ to $0{,}5995 + \varepsilon$. In other words, almost 60% of randomly generated substitutions guaranty security of SNOW 2.0-like cipher at the level of $2^{251{,}25}$.

When $x = 0{,}266$ we have $F_N(x) = 1$; together with this, the smallest value of $\log T_{r'}(k, l^*)$ in the fourth column of Table 1 equals 235,55, so the average time complexity of the attack is at least $2^{235{,}55}$. Thus, with the confidence of 99 % the part of substitutions, that guarantee specified security of SNOW 2.0-like ciphers, is at least $1 - \varepsilon$.

5. Conclusion

The paper proposes fast algorithm for computation the parameters of s-boxes that determine the security of SNOW 2.0-like stream ciphers against correlation attacks over extension fields. New algorithm requires $O(2^{2t} t)$ operations in comparison with the previous one, which time complexity is $O(2^{3t} t)$.

The experimental research shows that with the confidence at least 99 % the relative number of all substitutions, which guarantee security of the cipher against correlation attacks at the level at least $2^{251{,}25}$ are in the range from $0{,}5625$ to $0{,}6365$. Together with this, with the confidence 99 %, the part of substitutions which guarantee security at the level at least $2^{235{,}55}$ is $0{,}963$.

References

[1] Zhang B., Xu C., Hideki I. Relating differential distribution tables to other properties of of substitution boxes // Des. Codes Cryptogr., 19(1):45–63, 2000.

[2] Nyberg K. // Cryptology ePrint Archive, Report 2019/1381. http://eprint.iacr.org/2019/1381.

[3] Alekseychuk A.N., Shevtsov A.S. Upper estimates of imbalance of bilinear approximations for round functions of block ciphers // Cybernetics and Systems Analysis – 2010. – № 3. – P. 42–51.

[4] Alekseychuk A.N., Kovalchuk L.V. Towards a theory of security evaluation for GOST-like ciphers against differential and linear cryptanalysis // Cryptology ePrint Archive, Report 2011/489. http://eprint.iacr.org/2011/489

[5] Alekseychuk A.N., Koniushok S.M., Poremskyi M.V. A method for security evaluation of snow 2.0-like ciphers against correlation attacks over finite extensions of the field of two elements // Cybernetics and Systems Analysis – 2020. – № 1. – P. 49–63.

[6] Zhang B., Xu C., Meier W. Fast correlation attacks over extension fields, large-unit linear approximation and cryptanalysis of SNOW 2.0 // Cryptology ePrint Archive, Report 2016/311. http://eprint.iacr.org/2016/311.

[7] Ekdahl P., Johansson T. A new version of the stream cipher SNOW // Selected Areas in Cryptography – SAC 2002. – LNCS 2295. – Springer-Verlag, 2002. – P. 47 – 61.

[8] Gorbenko I., Kuznetsov A., Gorbenko Yu., Alekseychuk A., Timchenko V. Strumok Keystream Generator // The 9th IEEE International Conference on Dependable Systems, Services and Technologies, DESSERT'2018, 24 – 27 May, 2018, Kyiv, Ukraine. – P. 292 – 299

[9] Daemen J. Cipher and hash function design strategies based on linear and differential cryptanalysis. – Doctoral Dissertation, 1995.

[10] Logachev O.A. Salnikov A.A, Yashchenko B.B. Boolean functions in codding theory and cryptography. M.: CCNE, 2004. –P. 470.

[11] Hoeffding W. Probability inequalities for sums of bounded random variables // J. Amer. Statist. Assoc. – 1963. – Vol. 58. – № 301. – P. 13 – 30.