S©MR ISSN 1813-3304

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru

Том 16, стр. 144–150 (2022) УДК 512.5
DOI 10.33048/semi.2022.16.xxx MSC 13A99

ON THE NUMBER OF SUITABLE BOOLEAN FUNCTIONS IN CONSTRUCTIONS OF FILTER AND COMBINING MODELS OF STREAM CIPHERS

T.A. BONICH, M.A. PANFEROV, AND N.N. TOKAREVA

Abstract. It is well known that every stream cipher is based on a good pseudorandom generator. For cryptographic purposes we are interested in generating pseudorandom sequences with the maximum possible period. A feedback register is one of the most known cryptographic primitives that is used to construct stream ciphers. In this paper we analyze periodic properties of pseudorandom sequences produced by filter and combiner generators (two known schemes of stream generators based on feedback registers). We analyze functions in these schemes which lead to output sequences of the maximum possible period. We call such functions suitable and count the exact number of them for an arbitrary n.

Keywords: stream cipher, filter generator, combiner generator, Boolean function.

1. Introduction

Symmetric ciphers usually are divided into block and stream ones. Stream ciphers are considered as more fast but not as secure as block ciphers. One of the most known cryptographic primitives that is used to construct stream ciphers is a feedback register. There are many attacks and defenses on such ciphers and counter-measures against them, see for instance [1], [2].

Bonich, T.A., Panferov, M.A., Tokareva, N.N., On the number of suitable Boolean functions in constructions of filter and combining models of stream ciphers.

The work is supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2019-1613 with the Ministry of Science and Higher Education of the Russian Federation and Laboratory of Cryptography JetBrains Research.

Received March, 1, 2022, published March, 1, 2022.

The properties of the pseudorandom sequence (gamma) generated by FSR are well studied in case when f is a linear function. If f is nonlinear (see [3], [4]), there are too many open questions related to pseudorandom sequences that all are connected to analysis of nonlinear recurrent sequences, for example, see [5] for further review. That is why some nonlinear *combinations* of linear FSRs are usually considered, for instance, filter and combining models of stream generators based on LFSR (see [6]).

Let us recall a few definitions. Let $\mathbb{F}_2^n$ be the n-dimensional vector space over $\mathbb{F}_2$. A *Boolean function in n variables* is a function of the form $f : \mathbb{F}_2^n \to \mathbb{F}_2$. A *vector of values* for a given Boolean function f is the vector $(f(x^{(1)}), \ldots, f(x^{(2^n)}))$, where $x^{(1)}, \ldots, x^{(2^n)}$ are binary vectors in $\mathbb{F}_2^n$ that are lexicographically ordered. Any Boolean function f can be represented uniquely in its *algebraic normal form (ANF)*: $f(x_1, \ldots, x_n) = \bigoplus_{I \in \mathcal{P}(N)} a_I \Big(\prod_{i \in I} x_i \Big)$, where $\mathcal{P}(N)$ is a power set of $N = \{1, \ldots, n\}$ and $a_I \in \mathbb{F}_2$. For a Boolean function f the number of variables in the longest item of its ANF is called *the algebraic degree* of a function. If algebraic degree of f is not more than 1 then f is called *affine*. A function is called *linear* if it is affine and $f(0) = 0$. If algebraic degree of a function f is more than 1 then f is called *nonlinear*.

A *feedback shift register (FSR)* consists of two parts: a binary block $x = (x_1, \ldots, x_n)$ of length n and a feedback function f, where f is a Boolean function in n variables. First, we fill the block x with constants, it is the *initial state* of the register. During the encryption process the register is changing its state using the feedback function. *Gamma* is a pseudorandom sequence generated by FSR. For functioning of the FSR the time is considered to be divided into clock cycles. On each clock cycle, the value $f(x)$ is calculated first, then the state $x = (x_1, \ldots, x_{n-1}, x_n)$ of the register is changed to the state $x' = (x_2, \ldots, x_n, f(x))$ while the bit x_1 will be written as the first bit of the generated *gamma*. A *period* is a length of repeating part of gamma. If f is linear we have *a linear feedback shift register (LFSR)*. Similarly, *nonlinear feedback shift register (NLFSR)* uses nonlinear Boolean function as a feedback function. It is known that LFSR can be also specified by a feedback polynomial. It is a polynomial of degree n defining bits to be summed. If $f(x_1, \ldots, x_n) = a_1x_1 \oplus a_2x_2 \oplus \cdots \oplus a_nx_n$, then the corresponding feedback polynomial is defined as $p(z) = a_1z^n + a_2z^{n-1} + \cdots + a_nz + 1$, where $a_i \in \mathbb{F}_2$, $i = 1, \ldots, n$. If $p(z)$ is a *primitive polynomial*, i.e., the primitive element of the field $GF(2^n)$ is its root, then the period of a pseudorandom sequence generated by LFSR is maximal, i.e. is equal to $2^n - 1$. As a consequence, primitive polynomials mainly are used in LSFRs.

There are many stream ciphers based on LFSR and NLFSR. One of them is Grain, developed in 2004 [7]. It is constructed by combining model, based on two shift registers, one with linear feedback and one with nonlinear feedback, and a nonlinear output function. Both linear and nonlinear shift register sizes are 80 bits. Another one is A5/1 cipher from GSM standard [8]. It has three LSFRs of lengths 19, 22 and 23 bits with irregular clocking. The registers are clocked in a stop/go fashion using a majority rule. The output is the sum of the last bits of the three registers. We could also mention the Gollmann cascade [9]. This cipher is the representative of combining model. It consists of a series of LFSRs that are clock-controlled by the previous LFSR. If all the LFSRs have the same length n, the linear complexity of a system with k LFSRs is equal to $n(2^n - 1)^{k-1}$. Other examples of

ciphers that are based on LFSR and NLFSR are Geffe generator, Jennings generator and Beth-Piper Stop-and-Go generator.

In this paper, we analyze pseudorandom sequences produced by filter and combiner generators. Namely, we study functions in these schemes that lead to pseudorandom sequences of the maximal period. We call such functions *suitable* and count the exact number of them for an arbitrary n.

2. The analysis of gamma for linear feedback shift register generators

2.1. **Filter generators.** The filter generator consists of a single shift register of length n with a linear feedback and uses a primitive polynomial to change states. A Boolean function $h(x_1, \ldots, x_n)$ applied to the current state generates a pseudorandom sequence (gamma). Let us note that the number of all possible functions $h(x_1, \ldots, x_n)$ is equal to 2^{2^n}. The work of the filter generator is shown in [10].

Let gamma be defined as $\gamma = (y_1, y_2, \ldots, y_{2^n-1})$, where $y_1 = h(x_1, \ldots, x_n)$, $y_2 = h(x_2, \ldots, x_n, f(x_1, \ldots, x_n))$, etc., and $f(x_1, \ldots, x_n)$ is the feedback function. Since the number of all nonzero states is equal to $2^n - 1$, the maximum possible value of period of gamma is $2^n - 1$ too. In this paper, we would like to determine all Boolean functions h in n variables that lead to gammas with maximal period. Let us call such functions *suitable*. Functions that lead to gammas with non-maximal period we would call *unsuitable*. Note that the number of such functions does not depend on a linear feedback function. But whether the function is suitable or not for the given generator, depends on the feedback function. When we count the number of suitable functions h, we do not consider a specific set of states. We say that there is a certain number of different states used by the generator (all sets that are generated by primitive polynomials fit this definition). Next, we study which pseudorandom sequences have the maximal length. We analyze the number of unsuitable functions and the number of suitable functions. Thus, our reasonings do not affect the specific order of the states. Therefore, there will be the exact calculated number of suitable functions h for any set of states used by the generator.

Let us provide some examples of suitable and unsuitable functions. Let $n = 4$ be the length of a shift register, $f(x_1, x_2, x_3, x_4) = x_1 \oplus x_2$ be a feedback function and $p(z) = z^4 + z^3 + 1$ be a corresponding primitive polynomial. Let $h_1(x_1, x_2, x_3, x_4) = x_2x_1 \oplus x_3x_1 \oplus x_3x_2 \oplus x_4x_1 \oplus x_1 \oplus x_2 \oplus x_3 \oplus 1$ and $h_2(x_1, x_2, x_3, x_4) = x_4x_2x_1 \oplus x_2x_1 \oplus x_3x_2 \oplus x_3 \oplus 1$ be Boolean functions in n variables. We present generated gamma for these functions on Table 1.

See that h_1 and h_2 generate gamma with period 3 and 15.

Таблица 1. Examples of suitable and unsuitable functions

states	0001	0010	0100	1001	0011	0110	1101	1010
$h_1(x_1, x_2, x_3, x_4)$	1	0	0	1	0	0	1	0
$h_2(x_1, x_2, x_3, x_4)$	1	0	1	1	0	1	1	0
states	0101	1011	0111	1111	1110	1100	1000	0001
$h_1(x_1, x_2, x_3, x_4)$	0	1	0	0	1	0	0	1
$h_2(x_1, x_2, x_3, x_4)$	1	0	1	1	0	0	1	1

Let us prove the main result for filter generators.

Theorem 1. *Let $n \in \mathbb{N}$ and $2^n - 1 = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_s^{\alpha_s}$, where p_i are pairwise distinct prime numbers, $\alpha_i \in \mathbb{N}$, $s \in \mathbb{N}$. Then the number of suitable Boolean functions in n variables for the filter generator with LFSR based on a primitive polynomial of degree n, is equal to*

$$2^{2^n} - 2 \sum_{\beta \in \mathbb{F}_2^s, \beta \neq 0} ((-1)^{\beta_1 + \dots + \beta_s + 1} 2^{p_1^{\alpha_1 - \beta_1} \dots p_s^{\alpha_s - \beta_s}}), \text{ where } \beta = (\beta_1, \dots, \beta_s).$$

Proof. Consider sequences of length $2^n - 1$ with non-maximal period (*unsuitable* sequences). Let A_i be a set of sequences that can be divided on p_i identical subsequences, where $i = 1, \dots, s$. Then $A_i \cap A_j$ is a set of sequences that can be divided on $p_{i \cdot j}$ identical subsequences where $i \neq j$ and $i, j = 1, \dots, s$. Then $A_i \cup A_j$ is a set of sequences that can be divided on p_i or p_j identical subsequences where $i \neq j$ and $i, j = 1, \dots, s$. Hence, all unsuitable sequences belong to the set $\cup_{i=1}^s A_i$ and the number of these sequences is equal to $|\cup_{i=1}^s A_i|$. Dividing the sequence into p_i identical subsequences, the length of the subsequence is equal to $p_1^{\alpha_1} p_2^{\alpha_2} \dots p_i^{\alpha_i - 1} \dots p_s^{\alpha_s}$. Since elements of subsequences are equal to 0 or 1 then

$$|A_i| = 2^{p_1^{\alpha_1} p_2^{\alpha_2} \dots p_{(i-1)}^{\alpha_{(i-1)}} p_i^{\alpha_i - 1} p_{(i+1)}^{\alpha_{(i+1)}} \dots p_s^{\alpha_s}},$$

$$|A_i \cap A_j| = 2^{p_1^{\alpha_1} p_2^{\alpha_2} \dots p_{(i-1)}^{\alpha_{(i-1)}} p_i^{\alpha_i - 1} p_{(i+1)}^{\alpha_{(i+1)}} \dots p_{(j-1)}^{\alpha_{(j-1)}} p_j^{\alpha_j - 1} p_{(j+1)}^{\alpha_{(j+1)}} \dots p_s^{\alpha_s}},$$

$$\dots$$

$$|\cap_{i=1}^s A_i| = 2^{p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} \dots p_s^{\alpha_s - 1}}.$$

Therefore, we can compute $|\cup_{i=1}^s A_i|$ using the inclusion-exclusion principle:

$$|\cup_{i=1}^s A_i| = \sum_{i=1}^s |A_i| - \sum_{1 \leq i < j \leq s} |A_i \cap A_j| + \sum_{1 \leq i < j < k \leq s} |A_i \cap A_j \cap A_k| - \dots$$

$$+ (-1)^{s-1} |A_1 \cap A_2 \cap \dots \cap A_s| = \sum_{i=1}^s 2^{p_1^{\alpha_1} p_2^{\alpha_2} \dots p_{(i-1)}^{\alpha_{(i-1)}} p_i^{\alpha_i - 1} p_{(i+1)}^{\alpha_{(i+1)}} \dots p_s^{\alpha_s}} -$$

$$- \sum_{1 \leq i < j \leq s} 2^{p_1^{\alpha_1} p_2^{\alpha_2} \dots p_{(i-1)}^{\alpha_{(i-1)}} p_i^{\alpha_i - 1} p_{(i+1)}^{\alpha_{(i+1)}} \dots p_{(j-1)}^{\alpha_{(j-1)}} p_j^{\alpha_j - 1} p_{(j+1)}^{\alpha_{(j+1)}} \dots p_s^{\alpha_s}} +$$

$$\dots + (-1)^{s-1} 2^{p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} \dots p_s^{\alpha_s - 1}} = \sum_{\beta \in \mathbb{F}_2^s, \beta \neq 0} ((-1)^{\beta_1 + \dots + \beta_s + 1} 2^{p_1^{\alpha_1 - \beta_1} \dots p_s^{\alpha_s - \beta_s}}),$$

where $\beta = (\beta_1, \dots, \beta_s)$.

We can write all states of our register one by one and from one state we get the second one as the next state. Consider the vector of values of a Boolean function h that generates our gamma. Since there is no zero state in our set of states (it generates the cycle of length 1), function h can take any value (0 or 1) on zero vector. That is why there are exactly two Boolean functions that generate the same sequence.

Hence, the number of unsuitable functions is equal to

$$2|\cup_{i=1}^s A_i| = 2 \sum_{\beta \in \mathbb{F}_2^s, \beta \neq 0} ((-1)^{\beta_1 + \dots + \beta_s + 1} 2^{p_1^{\alpha_1 - \beta_1} \dots p_s^{\alpha_s - \beta_s}}).$$

Then, the number of suitable functions is
$2^{2^n} - 2 \sum_{\beta \in \mathbb{F}_2^s, \beta \neq 0} ((-1)^{\beta_1 + \dots + \beta_s + 1} 2^{p_1^{\alpha_1 - \beta_1} \dots p_s^{\alpha_s - \beta_s}})$, where $\beta = (\beta_1, \dots, \beta_s)$. □

2.2. **Combining model.** Combiner generators use several linear feedback shift registers. Each register has its own length n_i and uses its own primitive polynomial for changing states. A Boolean function $h(X_1, \dots, X_m)$ generates a pseudorandom sequence gamma, where X_i is a register bit string i. The work of the combiner generator is shown in [10].

Since we do not use zero state in combiner generator, the total number of states does not exceed $(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)$. In this case, the maximum is reached when $gcd(n_i, n_j) = 1$, where $i, j = 1, \dots, m$, $i \neq j$, and if all LFSRs have primitive feedback polynomials. Then a Boolean function can generate a gamma with a period ranging from 1 to $(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)$. Boolean functions h in n variables leading to gammas of maximal period in this case are called *suitable*. Similarly, Boolean functions h in n variables leading to gammas of non-maximal period are called *unsuitable*. Notice that $gcd(2^{n_i}-1, 2^{n_j}-1) = 1$, where $i, j = 1, \dots, m$, $i \neq j$. It means that each number $(2^{n_i}-1)$ can be presented as $p_{k_1}^{\alpha_{k_1}} p_{k_2}^{\alpha_{k_2}} \dots p_{k_s}^{\alpha_{k_s}}$, where $k_1, k_2, \dots, k_s$ are integers depend on i.

We consider a more general model of a combiner generator. This generalized combining model is applied in ciphers such as Grain[7]. Note that the classical combining model does not allow to describe a number of modern stream ciphers based on the more complicated operating with bits from different registers.

Theorem 2. *Let $n, m, n_1, \dots, n_m \in \mathbb{N}$, $\sum_{i=1}^{m} n_i = n$. And $(2^{n_1}-1)\dots(2^{n_m}-1) = p_1^{\alpha_1}\dots p_s^{\alpha_s}$, where p_i are different prime numbers, $\alpha_i, s \in \mathbb{N}$. Then the number of suitable Boolean functions in n variables for the combiner generator with LFSRs of lengths $n_1, \dots, n_m$ all based on primitive polynomials is equal to*

$$2^{2^n} - 2^{2^{n_1+\dots+n_m}-(2^{n_1}-1)\dots(2^{n_m}-1)} \sum_{\beta\in\mathbb{F}_2^s, \beta\neq 0} ((-1)^{\beta_1+\dots+\beta_s+1} 2^{p_1^{\alpha_1-\beta_1}\dots p_s^{\alpha_s-\beta_s}}),$$

where $\beta = (\beta_1, \dots, \beta_s)$.

Proof. Number of unsuitable sequences for the combiner generators is equal to $\sum_{\beta\in\mathbb{F}_2^s, \beta\neq 0}((-1)^{\beta_1+\dots+\beta_s+1} 2^{p_1^{\alpha_1-\beta_1}\dots p_s^{\alpha_s-\beta_s}})$. Proof of this is similar to the proof for the number of unsuitable sequences for the filter generators in Theorem 1. Since we use only $(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)$ states and the total number of states is equal to $2^{n_1}\cdot 2^{n_2}\dots 2^{n_m} = 2^{n_1+n_2+\dots+n_m}$, then we have $2^{n_1+n_2+\dots+n_m} - (2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)$ states, where our function can be equal to 0 or 1. Therefore, for one of these states we have two functions. Thus, the number of unsuitable Boolean functions in n variables for the combiner generators equals

$$2^{2^{n_1+n_2+\dots+n_m}-(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)} \cdot \sum_{\beta\in\mathbb{F}_2^s, \beta\neq 0} ((-1)^{\beta_1+\dots+\beta_s+1} 2^{p_1^{\alpha_1-\beta_1}\dots p_s^{\alpha_s-\beta_s}}),$$

where $\beta = (\beta_1, \dots, \beta_s)$.

Then, the number of suitable functions is equal to

$$2^{2^n} - 2^{2^{n_1+\dots+n_m}-(2^{n_1}-1)\dots(2^{n_m}-1)} \sum_{\beta\in\mathbb{F}_2^s, \beta\neq 0} ((-1)^{\beta_1+\dots+\beta_s+1} 2^{p_1^{\alpha_1-\beta_1}\dots p_s^{\alpha_s-\beta_s}}),$$

where $\beta = (\beta_1, \dots, \beta_s)$ □

3. Functions for models with nonlinear registers

A *nonlinear feedback shift register (NFSR)* consists of two parts: a binary vector $x = (x_1, \dots, x_n)$ of length n and a nonlinear state function $f : (x_1, \dots, x_n) \to \{0, 1\}$ in n variables.

Similarly to the linear case, let us consider the filter generator. We assume that NFSR passes over all 2^n states, i.e., it has the maximum possible period.

Theorem 3. *Let $n \in \mathbb{N}$. Then the number of suitable Boolean functions in n variables for the filter generator with NFSR of the maximum possible period is equal to $2^{2^n} - 2^{2^{n-1}}$.*

Proof. Maximum possible period is equal to $2^n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_s^{\alpha_s}$, where $s = 1$, $p_1 = 2$, $\alpha_1 = n$. Number of unsuitable sequences is equal to

$$\sum_{\beta \in \mathbb{F}_2^s, \beta \neq 0} ((-1)^{\beta_1 + \dots + \beta_s + 1} 2^{p_1^{\alpha_1 - \beta_1} \dots p_s^{\alpha_s - \beta_s}}),$$

where $\beta = (\beta_1, \dots, \beta_s)$. Proof of this is similar to the part of the proof of Theorem 1, in which the number of unsuitable sequences for the filter generators is calculated. Then the number of unsuitable sequences for the filter generator with NFSR is equal to $2^{2^{n-1}}$. Since we use all the states then the number of unsuitable sequences is equal to the number of unsuitable Boolean functions. Hence, the number of unsuitable Boolean functions in n variables for the filter generator with NFSR is equal to $2^{2^{n-1}}$. Therefore, the number of suitable functions is $2^{2^n} - 2^{2^{n-1}}$. □

There is another question related to NFSRs: how to determine for which nonlinear feedback functions NFSR of length n generates gamma with the maximum possible period 2^n? This question is still open.

References

[1] J.D. Golić, *On the security of nonlinear filter generators*, Fast Software Encryption, Lecture notes in Computer Science, Springer, Berlin, Heidelberg, 1996, 173–188.

[2] N.T. Courtois, W. Meier, *Algebraic attacks on stream ciphers with linear feedback*, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2003, 345–359.

[3] E. Key, *An analysis of the structure and complexity of nonlinear binary sequence generators*, IEEE Trans. Inf. Theory, Institute of Electrical and Electronics Engineers (IEEE), **22**:6 (1976), 732–736.

[4] A.A. Gorodilova, *From cryptanalysis to cryptographic property of a boolean function*, Prikl. Diskretn. Mat.,**33**:3 (2016), 16–44.

[5] M.M. Gluhov, V.P. Elizarov, A.A. Nechaev, *Algebra*, Lan, 2015, 327–333.

[6] A.J. Menezes, P.C. Oorschot, S.A. Vanstone, *Stream Ciphers*, Discrete Mathematics and Its Applications, CRC Press, 1996, 191–222.

[7] M. Hell, T. Johansson, W. Meier, *Grain: a stream cipher for constrained environments*, Int. j. wirel. mob. comput., Inderscience Publishers, **2**:1 (2007), 86.

[8] A. Canteaut, *A5/1*, Encyclopedia of Cryptography and Security, Springer US, Boston, MA, 2011, 1–2.

[9] D. Gollmann, *Kaskadenschaltungen taktgesteuerter Schieberegister als Pseudozufallszahlengeneratoren*, Verband d. wiss. Ges. Österreichs, 1986.

[10] C. Carlet, Y. Crama, P.L. Hammer, *Boolean Functions for Cryptography and Error-Correcting Codes*, Boolean Models and Methods in Mathematics, Comp. Sci., and Engin., Cambridge Univ. Press, 2010, 257-–397.

Tatiana Andreevna Bonich
Novosibirsk State University,
Pirogova str., 1,
630090, Novosibirsk, Russia
Email address: t.bonich@g.nsu.ru

Matvey Andreevich Panferov
Sobolev Institute of Mathematics,
pr. Koptyuga, 4,
630090, Novosibirsk, Russia
Email address: m.panferov@g.nsu.ru

Natalia Nikolaevna Tokareva
Sobolev Institute of Mathematics,
pr. Koptyuga, 4,
630090, Novosibirsk, Russia
Email address: tokareva@math.nsc.ru