J. Cryptol. (2012) 25: 464–483
DOI: 10.1007/s00145-011-9100-7

Journal of
CRYPTOLOGY

Perfectly Balanced Boolean Functions and Golić Conjecture

Stanislav V. Smyshlyaev
Computer Science Department, Lomonosov University, Moscow, Russia
smyshsv@gmail.com

Communicated by Willi Meier

Received 22 September 2010
Online publication 2 April 2011

Abstract. In the current paper we consider the following properties of filters: perfect balancedness of a filter function (i.e. preserving pure randomness of the input sequence) and linearity of a filter function in the first or the last essential variable. Previous results on this subject are discussed, including misleading statements in Gouget and Sibert (LNCS, vol. 4876, 2007) about the connection between perfect balancedness and resistance to Anderson conditional correlation attack; the incorrectness of two known results, the sufficient condition of perfect balancedness in Golić (LNCS, vol. 1039, 1996) and the necessary condition of perfect balancedness in Dichtl (LNCS, vol. 1267, 1997), is demonstrated by providing counterexamples.

We present a novel method of constructing large classes of perfectly balanced functions that are nonlinear in the first and the last essential variable and obtain a new lower bound of the number of such functions.

Golić conjecture (LNCS, vol. 1039, 1996) states that the necessary and sufficient condition for a function to be perfectly balanced for any choice of a tapping sequence is linearity of a function in the first or the last essential variable. In the second part of the current paper we prove the Golić conjecture.

Key words. Boolean function, Perfectly balanced function, Keystream generator, Filter, Golić conjecture

1. Introduction

Golić [3] studied cryptographic properties of keystream generators consisting of a shift register and a filter function which is connected to the register according to some tapping sequence. He considered a model of a keystream generator as a filter with a fixed filter function and an arbitrary choice of a tapping sequence. By proposing the inversion attack on the filter he showed a cryptographic weakness of keystream generators in such model in the case of a filter function being linear either in the first or the last variable.

Golić pointed out that even if the input sequence of the filter is purely random and the filter function is balanced, the output sequence is not necessarily such. The property of a Boolean function to preserve pure randomness of the input sequence (when

used as a filter function with consecutive choice of taps; it is not necessary that all variables are essential) is called perfect balancedness. The pure randomness requirement for nonlinear filter generators was introduced independently in [3,11].

Independently in [3,11] it was pointed out that all Boolean functions that are linear in the first or the last variable are perfectly balanced. But keystream generators with such functions are vulnerable to the inversion attack [3] in case of an inappropriate choice of a tapping sequence and thus perfectly balanced functions which are nonlinear in the first and the last essential variable are of primary interest (though the inversion attack in its generalized form [5] is still applicable). Examples of such functions were provided by Sumarokov in [11] and by Dichtl in [2], many examples can be found in a full description of perfectly balanced Boolean functions of four and five variables that was provided in [9]. Some other examples can be obtained using a construction proposed in [7] by Logachev and discussed in [10]. However, two important questions about obtaining lower bounds of the number of such functions and some large classes of such functions were still unanswered.

In the current paper we solve both of these problems by proposing a novel approach, based on the results of Sumarokov and Logachev. We construct large classes of perfectly balanced functions that are nonlinear in the first and the last essential variable and obtain a new lower bound of the number of such functions.

A sufficient condition from [3] and a necessary condition from [2] are shown to be incorrect in the current paper by providing counterexamples together with technical explanations. A criterion of perfect balancedness from [2] stated that a filter preserves pure randomness of an input binary sequence if and only if a vectorial Boolean function corresponding to $M + 1$ consecutive output bits of this filter is balanced. Only the sufficient condition (not the necessary condition) does not hold, because it is not sufficient to consider only $M + 1$ consecutive output bits, despite the finite input memory of M bits, but all of them, which is another characterization also given in [3].

Earlier Anderson [1] proposed a conditional correlation attack and showed a corresponding cryptographic weakness of filters in case of an inappropriate choice of both the tapping sequence and the filter function. The main point of [1] is that the correlations conditioned on the output segments (for linear functions of input bits) are stronger than those conditioned on single output bits only, provided that the tapping sequence consists of consecutive integers (which is the only case analyzed in [1]).

One of important contributions in [3] was that the conditional correlation attack depends on the choice of the tapping sequence and that it is preferable that the tapping sequence corresponds to a full positive difference set (PDS) or to a small-order PDS (if this order is 1, this is a full PDS). In this case the interactions between different phases of the filter function can be minimized. Mathematical arguments provided in [3, Sect. 3] were that a controllable resistance level to the conditional correlation attack can be ensured by choosing the tapping sequence according to a small-order PDS in combination with a correlation-immune filter function of appropriate correlation-immunity order.

In [6] there was a misleading consideration that in the case of pure random input sequence pure randomness of the output sequence of filter (i.e. perfect balancedness of the filter function) implies resistance to Anderson conditional correlation attack. As it was clearly pointed out in [1], as well as in [3,4], Anderson conditional correlation attack is about the correlations conditioned on the output, not about the pure randomness of

the output sequence, despite the fact that in the given examples in [1] the output segments are not balanced. Perfect balancedness of the filter function does not imply that the conditional correlation attack is impossible. For example, for a perfectly balanced filter function $x_1x_2 \oplus x_3$, if two consecutive output bits are $(0, 1)$, then the binary sum of the last two input bits is equal to 1 with probability 1, whereas the maximum probability of linear functions of input bits when conditioned on single output bits is $3/4$. The paper [6] incorrectly states that it is dealing with the conditional correlation attack from [1], while it is dealing with the pure randomness of the output sequence, in both probabilistic and deterministic models. The so-called quasi-immune filter functions are weakened first-order correlation-immune functions, where it is allowed that the function is correlated to at most one input. Proposition 5 from [6] states that if such a function is applied by using a tapping sequence corresponding to a full PDS, then any output segment influenced by any single input bit is balanced and vice versa, in the probabilistic model. This property is weaker than (i.e., does not imply) the sufficient condition from Lemma 1 [3] for perfect balancedness, pointed out to be incorrect in the current paper. Accordingly, this result is relevant neither for perfect balancedness nor for the conditional correlation attack. The authors of [6] do not refer to [3, Sect. 3], although they use the underlying concepts of full PDS and correlation-immune filter functions.

Perfect balancedness of a Boolean function is in fact a property of a filter with this function and consecutive taps to preserve pure randomness of input binary sequences. Sometimes it is more reasonable to consider only essential variables of a filter function and fix a tapping sequence according to positions of ones which are not essential. For example, the filter with a Boolean function x_1x_3 and the consecutive choice of taps (with a tapping sequence $(0, 1, 2)$) can be described as a filter with a Boolean function x_1x_2 and a tapping sequence $(0, 2)$.

An important problem is to describe the set of all Boolean functions such that a filter with any of these functions with an arbitrary choice of tapping sequence preserves pure randomness of an input binary sequence. This set contains all functions which are linear in the first or the last variable, but are there any other functions? This question was considered by Golić in [3]; importance of this problem is implied by the fact that all filters with functions linear in the first or the last variable are vulnerable to Golić inversion attack in case of tapping sequence not being spread all over the LFSR length.

Golić conjectured that in his model (with fixed Boolean function and an arbitrary choice of taps in the filter) a filter with a filter function f preserves pure randomness of an input binary sequence if and only if f is linear either in the first or the last variable. Golić proved an easier part of this conjecture, namely sufficiency, and noted that necessity remained unproven due to a "subtle underlying combinatorial problem remaining to be solved". According to Golić conjecture, the necessary condition for a function to be perfectly balanced (i.e. preserving pure randomness of an input binary sequence when used as a filter function) for any choice of a tapping sequence is linearity of a function in the first or the last essential variable. Golić conjecture implies that in the model being considered (with independent choice of a tapping sequence and a Boolean function) there are no functions both invulnerable to the inversion attack and preserving pure randomness.

To prove Golić conjecture, it suffices to find for an arbitrary Boolean function which is nonlinear in the first and the last essential variable a tapping sequence, such that

the Boolean function which describes input-output behavior of the corresponding filter does not satisfy the conditions of the Sumarokov criterion of perfect balancedness [11]. The trivial case of a function with no linear variables was considered in [8]. In the general case, all linear variables of a function have to be handled in a special way to construct a particular tapping sequence and two binary sequences required by Sumarokov criterion. This in fact solves an underlying combinatorial problem mentioned by Golić.

Organization of the Paper In Sects. 2 and 3 we describe our notation and provide necessary definitions and previous results.

In Sect. 4 we propose a new approach that can be used to construct large classes of perfectly balanced functions which are essential and nonlinear in the first and the last variable; using this approach we demonstrate an error in a necessary condition of perfect balancedness from [2] by providing a counterexample. In the end of Sect. 4 we demonstrate that a sufficient condition of perfect balancedness from [3] is incorrect and provide a necessary counterexample.

In Sect. 5 we provide our proof of Golić conjecture, proofs of technical lemmas are provided in appendices.

2. Definitions

As usual, $\mathbb{F}_2$ denotes the finite field with two elements. For any $n \in \mathbb{N}$ V_n denotes $\mathbb{F}_2^n$, $\mathcal{F}_n$ is the set of all Boolean functions in n variables. A variable x_i is called essential for the function $f(x_1, x_2, \ldots, x_n) \in \mathcal{F}_n$ if there exists $(\alpha_1, \alpha_2, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_n) \in V_{n-1}$ such that $f(\alpha_1, \alpha_2, \ldots, \alpha_{i-1}, 0, \alpha_{i+1}, \ldots, \alpha_n) \neq f(\alpha_1, \alpha_2, \ldots, \alpha_{i-1}, 1, \alpha_{i+1}, \ldots, \alpha_n)$. A variable x_i is called linear for the function $f(x_1, x_2, \ldots, x_n) \in \mathcal{F}_n$ if for any $(\alpha_1, \alpha_2, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_n) \in V_{n-1}$ inequality $f(\alpha_1, \alpha_2, \ldots, \alpha_{i-1}, 0, \alpha_{i+1}, \ldots, \alpha_n) \neq f(\alpha_1, \alpha_2, \ldots, \alpha_{i-1}, 1, \alpha_{i+1}, \ldots, \alpha_n)$ holds. By Φ_n, $\Phi_n \subset \mathcal{F}_n$, we denote the set of all Boolean functions with both the first and the last variables being essential. The subset of $\mathcal{F}_n$ composed by all functions linear in the first (resp. the last) variable is denoted by $\mathcal{L}_n$ (resp. $\mathcal{R}_n$).

Let $r \in \mathbb{N}$. A Boolean function $g \in \mathcal{F}_N$, $N \in \mathbb{N}$, induces a mapping $g_r \colon V_{r+N-1} \mapsto V_r$ of the form

$$\begin{aligned} &g_r(z_1, z_2, \ldots, z_{r+N-1}) \\ &\quad = \big(g(z_1, \ldots, z_N), g(z_2, \ldots, z_{N+1}), \ldots, g(z_r, \ldots, z_{r+N-1})\big). \end{aligned} \tag{2.1}$$

Let $\gamma = (\gamma_1, \ldots, \gamma_n)$ be a tuple of nonnegative integers such that $\gamma_1 = 0$; $\gamma_{i+1} > \gamma_i$, $i = 1, 2, \ldots, n-1$, and let $N = \gamma_n + 1$. From now on we consider tuples γ of this form. For γ of the above form and arbitrary $f \in \Phi_n$ we denote $f(x_{N-\gamma_n}, x_{N-\gamma_{n-1}}, \ldots, x_{N-\gamma_1})$ by $f^\gamma(x_1, \ldots, x_N)$.

A filter with a tapping sequence γ, and a filter function f is a mapping of the set $\bigcup_{i=\gamma_n+1}^{\infty} V_i$ to $\bigcup_{i=1}^{\infty} V_i$, defined by (2.1) with $m = 1, 2, \ldots$ and $g = f^\gamma$ (see Fig. 1).

Definition 2.1 (Sumarokov [11]). A Boolean function $f \in \mathcal{F}_n$ is said to be perfectly balanced if for any $r \in \mathbb{N}$ and any $\mathbf{y} \in V_r$

$$\left|f_r^{-1}(y)\right| = 2^{n-1},$$

where $|\cdot|$ denotes cardinality.

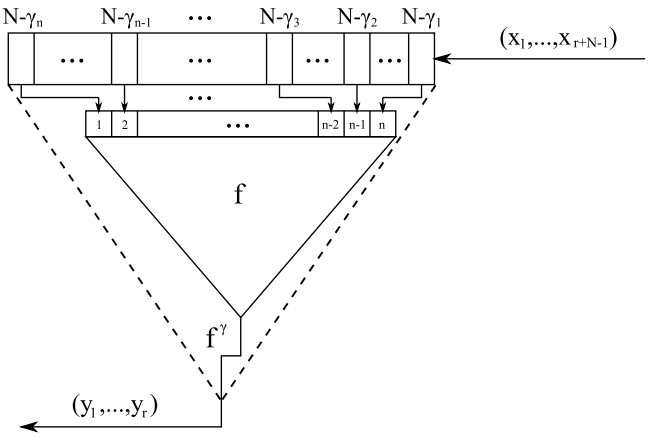

Fig. 1. A filter with a filter function f^{γ}.

Using Definition 2.1 is easy to acquire [3,11] necessary and sufficient condition for an r-tuple in the right-hand side of (2.1) to be distributed uniformly in V_r given the uniform distribution of the vector $X_r = (x_1, \ldots, x_{r+n-1})$.

Statement 2.2. *Let $n \in \mathbb{N}$ and $f \in \mathcal{F}_n$. Let $\{X_r = (x_1, \ldots, x_{r+n-1})\}_{r=1}^{\infty}$ be a sequence of random vectors with distribution*

$$\Pr\{X_r = (a_1, \ldots, a_{r+n-1})\} = 2^{-(r+n-1)}$$

for any $(a_1, \ldots, a_{r+n-1}) \in V_{r+n-1}$. Random vector $Y_r = f_r(X_r)$ is distributed uniformly for each $r \in \mathbb{N}$ iff f is perfectly balanced.

I.e., the output sequence of a filter with function f is purely random given that the input sequence is such if and only if f is perfectly balanced.

3. Preliminaries

We denote the set of all perfectly balanced n-variable functions by $\mathcal{PB}_n$, $\mathcal{PB}_n \subseteq \mathcal{F}_n$. From cryptographic applications point of view the subset $(\Phi_n \cap \mathcal{PB}_n) \setminus (\mathcal{L}_n \cup \mathcal{R}_n)$ is of primary importance.

Theorem 3.1 (Sumarokov [11]). *A Boolean function $f \in \mathcal{F}_n$ is perfectly balanced iff there is no pair of distinct binary sequences*

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_r), \qquad \boldsymbol{z} = (z_1, z_2, \ldots, z_r) \in V_r, \quad r \geq 2n+1, \tag{3.1}$$

such that

$$x_1 = z_1, \quad x_2 = z_2, \quad \ldots, \quad x_n = z_n, \quad x_{r-n+1} = z_{r-n+1}, \quad \ldots, \quad x_r = z_r; \tag{3.2}$$

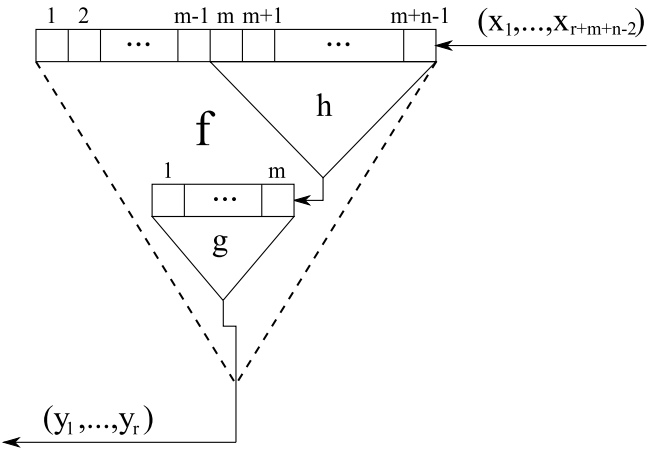

Fig. 2. A filter with a filter function $f = \Xi_{m,n}(g,h) = g[h]$.

$$\boldsymbol{x} \neq z; \tag{3.3}$$

$$f(x_i, x_{i+1}, \ldots, x_{i+n-1}) = f(z_i, z_{i+1}, \ldots, z_{i+n-1}), \quad i = 1, 2, \ldots, r-n+1. \tag{3.4}$$

The proof of Theorem 3.1 can be found in Appendix A.

Example 3.2. Let $n = 3$, $f(x_1, x_2, x_3) = x_1x_3 \oplus x_2$.

Since $f_2^{-1}(0,1) = \{(0,0,1,0), (0,0,1,1), (1,1,1,0)\}$ and $|f_2^{-1}(0,1)| = 3 \neq 2^{n-1} = 4$, $f \notin \mathcal{PB}_3$.

On the other hand, $f_4(0,0,1,0,0,1,0,0) = f_4(0,0,1,1,1,1,0,0)$ and $f \notin \mathcal{PB}_3$ follows from Theorem 3.1.

Remark 3.3. It is easy to see that all functions in $\mathcal{L}_n \cup \mathcal{R}_n$ satisfy the conditions of Theorem 3.1 and thus $\mathcal{L}_n \cup \mathcal{R}_n \subseteq \mathcal{PB}_n$.

For any integers m and n consider a mapping $\Xi_{m,n} \colon \mathcal{F}_m \times \mathcal{F}_n \mapsto \mathcal{F}_{m+n-1}$ of the form

$$\Xi_{m,n}(g,h) = g[h] = f \in \mathcal{F}_{m+n-1}, \quad g \in \mathcal{F}_m,\ h \in \mathcal{F}_n,$$

where (see Fig. 2)

$$f(x_1, x_2, \ldots, x_{m+n-1}) = g[h](x_1, x_2, \ldots, x_{m+n-1}) = g(h_m(x_1, x_2, \ldots, x_{m+n-1}))$$
$$= g(h(x_1, x_2, \ldots, x_n), h(x_2, x_3, \ldots, x_{n+1}), \ldots, h(x_m, x_{m+1}, \ldots, x_{m+n-1})).$$

Example 3.4. Let $g, h \in \mathcal{F}_3$, $h(x_1, x_2, x_3) = x_1x_3 \oplus x_2$, $g(z_1, z_2, z_3) = z_1z_2 \oplus z_3$, $f = \Xi(3,3)(g,h) \in \mathcal{F}_5$. Then $f(x_1, x_2, x_3, x_4, x_5) = g[h](x_1, x_2, x_3, x_4, x_5) = (x_1x_3 \oplus x_2) \cdot (x_2x_4 \oplus x_3) \oplus (x_3x_5 \oplus x_4) = x_1x_2x_3x_4 \oplus x_1x_3 \oplus x_2x_4 \oplus x_2x_3 \oplus x_3x_5 \oplus x_4$.

4. Construction of Perfectly Balanced Functions

Golić and Sumarokov showed independently in [3,11] that all Boolean functions that are linear in the first or the last variable are perfectly balanced, i.e. preserving pure randomness of the input sequence when used as a filter function. On the other hand, Golić demonstrated a cryptographic weakness of filters with such functions if the tapping sequence is not spread all over the LFSR length. As it was shown in [5], in case of inappropriate choice of the tapping sequence the generalized inversion attack is still applicable even if the filter function $f \in \mathcal{F}_n$ is linear in neither the first nor the last input variable. However, it is less effective, since in this case a minimum of fractions $p_f^+ = |\{(x_2, x_3, \ldots, x_n) \in V_{n-1} | f(0, x_2, x_3, \ldots, x_n) = f(1, x_2, x_3, \ldots, x_n)\}|/2^n$ and $p_f^- = |\{(x_1, x_2, \ldots, x_{n-1}) \in V_{n-1} | f(x_1, x_2, \ldots, x_{n-1}, 0) = f(x_1, x_2, \ldots, x_{n-1}, 1)\}|/2^n$ is nonzero. As it was conjectured in [3] and proven in the current paper in Sect. 5, there are no functions that are nonlinear in the first and the last variable and preserving pure randomness of the input sequence when used as a filter function with an arbitrary tapping sequence (i.e. such that $f^\gamma \in (\mathcal{PB}_N \cap \Phi_N) \setminus (\mathcal{L}_N \cup \mathcal{R}_N)$ for any possible γ).

In [2,7,9,11] the examples of perfectly balanced Boolean functions of four and five variables that are nonlinear in the first and the last variable were provided. However, nothing was known about such functions except these examples—nor lower bounds of the number of such functions for arbitrary n, nor even any methods to construct such functions for any $n > 6$. In the current section we propose an approach (based on the construction from [7]) that can be used to construct large classes of such functions and obtain a new lower bound of the cardinality of the set $(\mathcal{PB}_n \cap \Phi_n) \setminus (\mathcal{L}_n \cup \mathcal{R}_n)$ for an arbitrary n.

First, we demonstrate some auxiliary statements.

Lemma 4.1 (Logachev [7]). *Let $g \in \mathcal{F}_m$, $h \in \mathcal{F}_n$. The function $f = g[h] \in \mathcal{F}_{m+n-1}$ is perfectly balanced iff both functions g and h are perfectly balanced.*

Proof. *Sufficiency.* Let $g \in \mathcal{PB}_m$, $h \in \mathcal{PB}_n$. For any $l \in \mathbb{N}$ and for any $\mathbf{y} \in V_l$

$$\left|f_l^{-1}(\mathbf{y})\right| = \left|\bigcup_{\mathbf{z} \in g_l^{-1}(\mathbf{y})} h_{l+m-1}^{-1}(\mathbf{z})\right| = \sum_{\mathbf{z} \in g_l^{-1}(\mathbf{y})} \left|h_{l+m-1}^{-1}(\mathbf{z})\right| = 2^{m-1} \cdot 2^{n-1} = 2^{m+n-2},$$

thus f is perfectly balanced.

Necessity. Let $h \notin \mathcal{PB}_n$. As follows from Theorem 3.1, for some $r \geq 2n+1$ there exists a pair of distinct tuples $\mathbf{x}, \mathbf{z} \in V_r$ such that $(x_1, x_2, \ldots, x_n) = (z_1, z_2, \ldots, z_n)$, $(x_{r-n+1}, x_{r-n+2}, \ldots, x_r) = (z_{r-n+1}, z_{r-n+2}, \ldots, z_r)$, $h_{r-n+1}(\mathbf{x}) = h_{r-n+1}(\mathbf{z})$. Thus

$$\begin{aligned} &f_{r+m-n}(\underbrace{0, 0, \ldots, 0}_{m-1}, x_1, x_2, \ldots, x_r, \underbrace{0, 0, \ldots, 0}_{m-1}) \\ &= f_{r+m-n}(\underbrace{0, 0, \ldots, 0}_{m-1}, z_1, z_2, \ldots, z_r, \underbrace{0, 0, \ldots, 0}_{m-1}) \end{aligned}$$

and, as follows from Theorem 3.1, $f \notin \mathcal{PB}_{m+n-1}$.

Let $h \in \mathcal{PB}_n$, $g \notin \mathcal{PB}_m$. In this case there exists $l \in \mathbb{N}$ and $\mathbf{y} \in V_l$ such that $|g_l^{-1}(\mathbf{y})| = A \neq 2^{m-1}$. Hence

$$\left|f_l^{-1}(\mathbf{y})\right| = \sum_{\mathbf{z} \in g_l^{-1}(\mathbf{y})} \left|h_{l+m-1}^{-1}(\mathbf{z})\right| = A \cdot 2^{n-1} \neq 2^{m+n-2}$$

and $f \notin \mathcal{PB}_{m+n-1}$.

Thus if $f = g[h] \in \mathcal{PB}_{m+n-1}$ then $g \in \mathcal{PB}_m$, $h \in \mathcal{PB}_n$. □

Lemma 4.2. *Let $n \in \mathbb{N}$, $h \in \mathcal{PB}_n$. Then for any integer $m \in \mathbb{N}$ and any pair of distinct Boolean functions $g^{(1)}, g^{(2)} \in \mathcal{F}_m$ inequality $g^{(1)}[h] \neq g^{(2)}[h]$ holds.*

Proof. Let $h \in \mathcal{PB}_n$, $m \in \mathbb{N}$, $g^{(1)}, g^{(2)} \in \mathcal{F}_m$, $g^{(1)} \neq g^{(2)}$ and $f^{(1)} = g^{(1)}[h]$, $f^{(2)} = g^{(2)}[h]$. Fix a tuple $\mathbf{z} \in V_m$ such that $g^{(1)}(\mathbf{z}) \neq g^{(2)}(\mathbf{z})$. Function h is perfectly balanced, hence there exists a tuple $\mathbf{x} \in V_{m+n-1}$ such that $h_m(\mathbf{x}) = \mathbf{z}$. Thus $f^{(1)}(\mathbf{x}) = g^{(1)}(h_m(\mathbf{x})) = g^{(1)}(\mathbf{z}) \neq g^{(2)}(\mathbf{z}) = g^{(2)}(h_m(\mathbf{x})) = f^{(2)}(\mathbf{x})$ and $f^{(1)} \neq f^{(2)}$. □

Lemma 4.3. *Let $m \in \mathbb{N}$, $g \in \mathcal{PB}_m$. Then for any integer $n \in \mathbb{N}$ and any pair of distinct Boolean functions $h^{(1)}, h^{(2)} \in \mathcal{F}_n$ such that $h^{(1)}(0, 0, \ldots, 0) = h^{(2)}(0, 0, \ldots, 0)$ inequality $g[h^{(1)}] \neq g[h^{(2)}]$ holds.*

Proof. Let $f^{(1)} = g[h^{(1)}]$, $f^{(2)} = g[h^{(2)}]$. Since $h^{(1)} \neq h^{(2)}$, there exists a tuple $\tilde{\mathbf{x}} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) \in V_n$ such that $h^{(1)}(\tilde{\mathbf{x}}) \neq h^{(2)}(\tilde{\mathbf{x}})$. Consider a tuple $\mathbf{x} \in V_{2m+3n-4}$, $\mathbf{x} = (\underbrace{0, 0, \ldots, 0}_{m+n-2}, \tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n, \underbrace{0, 0, \ldots, 0}_{m+n-2})$. It suffices to show that $f^{(1)}_{m+2n-2}(\mathbf{x}) \neq f^{(2)}_{m+2n-2}(\mathbf{x})$.

By contradiction, let $f^{(1)}_{m+2n-2}(\mathbf{x}) = f^{(2)}_{m+2n-2}(\mathbf{x})$. Then the following system is solvable:

$$\begin{cases}
g(h^{(1)}(0, 0, \ldots, 0), h^{(1)}(0, 0, \ldots, 0), \ldots, h^{(1)}(0, 0, \ldots, 0), h^{(1)}(0, 0, \ldots, 0, \tilde{x}_1)) \\
\quad = g(h^{(2)}(0, 0, \ldots, 0), h^{(2)}(0, 0, \ldots, 0), \ldots, h^{(2)}(0, 0, \ldots, 0), h^{(2)}(0, 0, \ldots, 0, \tilde{x}_1)); \\
g(h^{(1)}(0, 0, \ldots, 0), \ldots, h^{(1)}(0, 0, \ldots, 0, \tilde{x}_1), h^{(1)}(0, 0, \ldots, 0, \tilde{x}_1, \tilde{x}_2)) \\
\quad = g(h^{(2)}(0, 0, \ldots, 0), \ldots, h^{(2)}(0, 0, \ldots, 0, \tilde{x}_1), h^{(2)}(0, 0, \ldots, 0, \tilde{x}_1, \tilde{x}_2)); \\
\ldots \\
g(h^{(1)}(0, 0, \ldots, 0), \ldots, h^{(1)}(0, \tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{n-1}), h^{(1)}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)) \\
\quad = g(h^{(2)}(0, 0, \ldots, 0), \ldots, h^{(2)}(0, \tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{n-1}), h^{(2)}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)); \\
\ldots \\
g(h^{(1)}(\tilde{x}_n, 0, \ldots, 0), h^{(1)}(0, 0, \ldots, 0), \ldots, h^{(1)}(0, 0, \ldots, 0)) \\
\quad = g(h^{(2)}(\tilde{x}_n, 0, \ldots, 0), h^{(2)}(0, 0, \ldots, 0), \ldots, h^{(2)}(0, 0, \ldots, 0)); \\
h^{(1)}(0, 0, \ldots, 0) = h^{(2)}(0, 0, \ldots, 0); \\
h^{(1)}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) \neq h^{(2)}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n).
\end{cases} \tag{4.1}$$

As follows from Theorem 3.1, system (4.1) cannot be solvable in case of perfectly balanced function g. □

Corollary 4.4. *Let* $n \in \mathbb{N}$, $h^{(1)}, h^{(2)}, h^{(3)} \in \mathcal{F}_n$; $h^{(i)} \neq h^{(j)}$, $i \neq j$. *Let* $m \in \mathbb{N}$, $g \in \mathcal{PB}_m$, $f^{(i)} = g[h^{(i)}]$, $i = 1, 2, 3$. *Then at least two of functions* $f^{(1)}, f^{(2)}, f^{(3)}$ *are distinct.*

Remark 4.5. It is important to note that perfect balancedness of Boolean function g and inequality $h^{(1)} \neq h^{(2)}$ do not necessary imply $g[h^{(1)}] \neq g[h^{(2)}]$. To show this, it suffices to consider functions $g(x_1, x_2) = x_1 \oplus x_2 \in \mathcal{PB}_2$ and $h^{(2)} = h^{(1)} \oplus 1$ (for any $h^{(1)}$).

Lemma 4.6. *Let* $m, n \in \mathbb{N}$, $g \in \mathcal{F}_m$, $h \in \mathcal{F}_n$. *If* $g \neq 0$, $g \neq 1$, $h \in \mathcal{R}_n$, *then* $g[h] \neq 0$, $g[h] \neq 1$.

Proof. $h(x_1, x_2, \ldots, x_n) = h'(x_1, x_2, \ldots, x_{n-1}) \oplus x_n$ and $g(z_1, z_2, \ldots, z_m) = z_i \cdot g'(z_1, z_2, \ldots, z_{i-1}) \oplus g''(z_1, z_2, \ldots, z_{i-1})$ for some integer i, $1 \leq i \leq n$, and some $g', g'' \in \mathcal{F}_{i-1}$, $g' \neq 0$.

If $i = 1$ then $g' = 1$ and $g[h] = h$ or $g[h] = h \oplus 1$, hence $g[h] \neq 0$, $g[h] \neq 1$.

If $i \geq 2$ then

$$\begin{aligned} &g[h](x_1, x_2, \ldots, x_{m+n-1}) \\ &\quad = x_{i+n-1} \cdot g'[h](x_1, x_2, \ldots, x_{i+n-2}) \\ &\qquad \oplus h'(x_i, x_{i+1}, \ldots, x_{i+n-2}) \cdot g'[h](x_1, x_2, \ldots, x_{i+n-2}) \\ &\qquad \oplus g''[h](x_1, x_2, \ldots, x_{i+n-2}). \end{aligned}$$

Using induction, it is easy to show that $g'[h] \neq 0$.

Thus $g[h] \neq 0$, $g[h] \neq 1$. □

Lemma 4.7. *Let* $m, n \in \mathbb{N}$, $m \geq 3$, $n \geq 3$, $g \in \mathcal{F}_m$, $h \in \mathcal{F}_n$. *If* $g \in \Phi_m \cap (\mathcal{L}_m \setminus \mathcal{R}_m)$ *and* $h \in \Phi_n \cap (\mathcal{R}_n \setminus \mathcal{L}_n)$, *then* $g[h] \in \Phi_{m+n-1} \setminus (\mathcal{L}_{m+n-1} \cup \mathcal{R}_{m+n-1})$.

Proof. $g(z_1, z_2, \ldots, z_m) = z_1 \oplus z_m \cdot g'(z_2, z_3, \ldots, z_{m-1}) \oplus g''(z_2, z_3, \ldots, z_{m-1})$, $h(x_1, x_2, \ldots, x_n) = x_1 \cdot h'(x_2, x_3, \ldots, x_{n-1}) \oplus h''(x_2, x_3, \ldots, x_{n-1}) \oplus x_n$ for some $g', g'' \in \mathcal{F}_{m-2}$ and $h', h'' \in \mathcal{F}_{n-2}$ such that $g' \neq 0$, $g' \neq 1$, $h' \neq 0$, $h' \neq 1$.

Then $g[h](x_1, x_2, \ldots, x_{m+n-1}) = x_1 \cdot h'(x_2, x_3, \ldots, x_{n-1}) \oplus h''(x_2, x_3, \ldots, x_{n-1}) \oplus x_n \oplus x_m \cdot h'(x_{m+1}, x_{m+2}, \ldots, x_{m+n-2}) \cdot g'[h](x_2, x_3, \ldots, x_{m+n-2}) \oplus h''(x_{m+1}, x_{m+2}, \ldots, x_{m+n-2}) \cdot g'[h](x_2, x_3, \ldots, x_{m+n-2}) \oplus x_{m+n-1} \cdot g'[h](x_2, x_3, \ldots, x_{m+n-2}) \oplus g''[h](x_2, x_3, \ldots, x_{m+n-2})$. $h' \neq 0$, $h' \neq 1$ and, as follows from Lemma 4.6, $g'[h] \neq 0$, $g'[h] \neq 1$. Thus $g[h] \in \Phi_{m+n-1} \setminus (\mathcal{L}_{m+n-1} \cup \mathcal{R}_{m+n-1})$. □

The following theorem is a consequence of Lemmas 4.1 and 4.7.

Theorem 4.8. *Let* $m, n \in \mathbb{N}$, $m \geq 3$, $n \geq 3$, $g \in \Phi_m \cap (\mathcal{L}_m \setminus \mathcal{R}_m)$, $h \in \Phi_n \cap (\mathcal{R}_n \setminus \mathcal{L}_n)$, $f = g[h]$. *Then* $f \in (\mathcal{PB}_{m+n-1} \cap \Phi_{m+n-1}) \setminus (\mathcal{L}_{m+n-1} \cup \mathcal{R}_{m+n-1})$.

I.e., the composition of a function linear in the first variable and nonlinear in the last variable and a function linear in the last variable and nonlinear in the first variable is a perfectly balanced function that is linear neither in the first nor the last variable.

Let $n \geq 5$. Consider the sets $Q'_n = \{f = g[h] \mid h(x_1, x_2, x_3) = x_1x_2 \oplus x_3, g \in \Phi_{n-2} \cap (\mathcal{L}_{n-2} \setminus \mathcal{R}_{n-2})\}$, $\widetilde{Q}_n = \{f = g[h] \mid g(z_1, z_2, z_3) = z_1 \oplus z_2z_3, h \in \Phi_{n-2} \cap (\mathcal{R}_{n-2} \setminus \mathcal{L}_{n-2}), h(0, 0, \ldots, 0) = 0\}$, $Q''_n = \widetilde{Q}_n \cup \{f = f' \oplus 1 \mid f' \in \widetilde{Q}_n\}$ and $Q_n = Q'_n \cup Q''_n$.

Lemma 4.9. *Let $n \geq 5$. Then $|Q'_n| = |Q''_n| = 2^{2^{n-4}} \cdot (2^{2^{n-4}} - 2)$, $Q_n \subseteq (\mathcal{PB}_n \cap \Phi_n) \setminus (\mathcal{L}_n \cup \mathcal{R}_n)$.*

Proof. From Theorem 4.8 it follows that $Q'_n \subseteq (\mathcal{PB}_n \cap \Phi_n) \setminus (\mathcal{L}_n \cup \mathcal{R}_n)$, $\widetilde{Q}_n \subseteq (\mathcal{PB}_n \cap \Phi_n) \setminus (\mathcal{L}_n \cup \mathcal{R}_n)$. From Lemma 4.2 it follows that $|Q'_n| = |\Phi_{n-2} \cap (\mathcal{L}_{n-2} \setminus \mathcal{R}_{n-2})| = 2^{2^{n-4}} \cdot (2^{2^{n-4}} - 2)$ and from Lemma 4.3 it follows that $|\widetilde{Q}_n| = |\{h \in \Phi_{n-2} \cap (\mathcal{R}_{n-2} \setminus \mathcal{L}_{n-2}) \mid h(0, 0, \ldots, 0) = 0\}| = \frac{1}{2} \cdot 2^{2^{n-4}} \cdot (2^{2^{n-4}} - 2)$.

It is evident that a Boolean function f is perfectly balanced if and only if the function $f \oplus 1$ is perfectly balanced. Thus $Q''_n \subseteq (\mathcal{PB}_n \cap \Phi_n) \setminus (\mathcal{L}_n \cup \mathcal{R}_n)$ and $Q_n \subseteq (\mathcal{PB}_n \cap \Phi_n) \setminus (\mathcal{L}_n \cup \mathcal{R}_n)$. If $f \in \widetilde{Q}_n$, then $f = g[h]$, $g(z_1, z_2, z_3) = z_1 \oplus z_2z_3$, $h(0, 0, \ldots, 0) = 0$ and $f(0, 0, \ldots, 0) = 0$. Thus $\widetilde{Q}_n \cap \{f = f' \oplus 1 \mid f' \in \widetilde{Q}_n\} = \varnothing$, $|Q''_n| = 2 \cdot |\widetilde{Q}_n| = 2^{2^{n-4}} \cdot (2^{2^{n-4}} - 2)$. □

Lemma 4.10. *Let $n \geq 5$. Then $|Q_n| \geq 2^{2^{n-3}+1} - 5 \cdot 2^{2^{n-4}}$.*

Proof. Let $f \in Q'_n \cap Q''_n$. Then $f(x_1, x_2, \ldots, x_n) = h(x_1, x_2, \ldots, x_{n-2}) \oplus h(x_2, x_3, \ldots, x_{n-1}) \cdot h(x_3, x_4, \ldots, x_n) \oplus b = g(x_1x_2 \oplus x_3, x_2x_3 \oplus x_4, \ldots, x_{n-2}x_{n-1} \oplus x_n)$, where $h \in \Phi_{n-2} \cap (\mathcal{R}_{n-2} \setminus \mathcal{L}_{n-2})$, $h(0, 0, \ldots, 0) = 0$, $g \in \Phi_{n-2} \cap (\mathcal{L}_{n-2} \setminus \mathcal{R}_{n-2})$, $b \in \mathbb{F}_2$.

Let $h(x_1, x_2, \ldots, x_{n-2}) = x_1 \cdot h'(x_2, x_3, \ldots, x_{n-3}) \oplus h''(x_2, x_3, \ldots, x_{n-3}) \oplus x_{n-2}$, $h''(0, 0, \ldots, 0) = 0$, $g(z_1, z_2, \ldots, z_{n-2}) = z_1 \oplus \tilde{g}(z_2, z_3, \ldots, z_{n-2})$. Then

$$
\begin{aligned}
&f(x_1, x_2, \ldots, x_n) \\
&\quad = x_1 \cdot h'(x_2, x_3, \ldots, x_{n-3}) \\
&\qquad \oplus [h''(x_2, x_3, \ldots, x_{n-3}) \oplus x_{n-2} \oplus h(x_2, x_3, \ldots, x_{n-1}) \cdot h(x_3, x_4, \ldots, x_n) \oplus b] \\
&\quad = x_1x_2 \oplus x_3 \oplus \tilde{g}(x_2x_3 \oplus x_4, x_3x_4 \oplus x_5, \ldots, x_{n-2}x_{n-1} \oplus x_n),
\end{aligned}
$$

hence $h'(x_2, x_3, \ldots, x_{n-3}) = x_2$, $h(x_1, x_2, \ldots, x_{n-2}) = x_1x_2 \oplus h''(x_2, \ldots, x_{n-3}) \oplus x_{n-2}$, $h''(0, 0, \ldots, 0) = 0$ and $|Q'_n \cap Q''_n| \leq |\{h'' \in \mathcal{F}_{n-4} \mid h''(0, 0, \ldots, 0) = 0\}| \cdot |\{b \in \mathbb{F}_2\}| = 2^{2^{n-4}-1} \cdot 2 = 2^{2^{n-4}}$.

Thus $|Q_n| = |Q'_n| + |Q''_n| - |Q'_n \cap Q''_n| \geq 2 \cdot (2^{2^{n-4}} \cdot (2^{2^{n-4}} - 2)) - 2^{2^{n-4}} = 2^{2^{n-3}+1} - 5 \cdot 2^{2^{n-4}}$. □

Using Lemmas 4.9 and 4.10 it is easy to obtain the following lower bound of the number of perfectly balanced Boolean functions which are nonlinear in the first and the last variable.

Corollary 4.11. $|(\mathcal{PB}_n \cap \Phi_n) \setminus (\mathcal{L}_n \cup \mathcal{R}_n)| \geq 2^{2^{n-3}+1} - 5 \cdot 2^{2^{n-4}}$.

Using the set Q'_n it is possible to show that the necessary condition of perfect balancedness from [2] is incorrect.

Theorem 1 (Dichtl [2]). *Let $f \in \mathcal{F}_n$ ($n \in \mathbb{N}$) be the filter function of a nonlinear filter generator. If the outputs of the filter function are random independent bits with probability $1/2$ of being 0 and 1 (assuming that input bits are random independent bits with probability $1/2$ of being 0 and 1) (i.e. $f \in \mathcal{PB}_n$), then there exists at most one index j ($1 \leq j \leq n$) such that the function $f^j \in \mathcal{F}_{n-1}$ with $f^j(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) = f(x_1, \ldots, x_{j-1}, 0, x_{j+1}, \ldots, x_n)$, that is, the function for which the jth input bit of f is fixed to 0, is not balanced.*

Example 4.1. Let $g(z_1, z_2, z_3) = z_1 \oplus z_2 z_3$ and $h(x_1, x_2, x_3) = x_1 x_2 \oplus x_3$. Then $f = g[h] \in Q'_5 \subseteq \mathcal{PB}_5$, however Boolean functions $f^3(x_1, x_2, x_4, x_5) = f(x_1, x_2, 0, x_4, x_5)$ and $f^5(x_1, x_2, x_3, x_4) = f(x_1, x_2, x_3, x_4, 0)$ are both of weight 6 and thus are not balanced that contradicts considered theorem.

Remark 4.2. The problem with the proof of considered theorem presented in [2] is that for a perfectly balanced function it is not necessary that output bits y_j and y_k are still independent (as they are when an input sequence is purely random, with no fixed bits) in case when one input bit b is fixed to zero.

Another incorrect result about perfectly balanced Boolean functions is the sufficient condition presented in [3].

Lemma 1 (Golić [3]). *For a nonlinear filter generator with function f and input memory size M, the output sequence is purely random given that the input sequence is such (i.e. $f \in \mathcal{PB}_{M+1}$) if and only if f_{M+1} is balanced.*

Example 4.3. Let $f(x_1, x_2, x_3, x_4) = x_1 \oplus x_4 \oplus x_1 x_3 \oplus x_1 x_4 \oplus x_1 x_2 x_4$, $M = 3$. Then $f_{M+1} = f_4$ is balanced but f_5 is not and thus $f \notin \mathcal{PB}_4$, which contradicts the considered lemma.

Example 4.4. Let $f(x_1, x_2, x_3, x_4, x_5) = x_2 \oplus x_3 \oplus x_5 \oplus x_1 x_4 \oplus x_1 x_5 \oplus x_2 x_4 \oplus x_2 x_5 \oplus x_3 x_5 \oplus x_1 x_4 x_5 \oplus x_2 x_4 x_5$, $M = 4$. Then $f_{M+1} = f_5$ is balanced and even $f_{M+2} = f_6$ is balanced but f_7 is not and thus $f \notin \mathcal{PB}_5$ that contradicts considered lemma and even its weaker version (with f_{M+2} instead of f_{M+1}).

Remark 4.5. The problem with the proof of considered lemma presented in [3] is that (for $f \in \mathcal{F}_n, M = n - 1$) balancedness of output bit $y_t = f(x_t, x_{t+1}, \ldots, x_{t+M})$ for any fixed value of preceding $M = n - 1$ output bits $(y_{t-M}, y_{t-M+1}, \ldots, y_{t-1}) = f_M(x_{t-M}, x_{t-M+1}, \ldots, x_{t+M-1})$ does not necessary imply balancedness of y_t for any fixed value of preceding $t - 1$ output bits $y_1, y_2, \ldots, y_{t-1}$.

5. Proof of Golić Conjecture

Theorem 5.1 (Golić [3]). *For a filter with a filter function f for any choice of a tapping sequence γ the output sequence is purely random given that the input sequence is such if (and only if [not proven]) $f(z_1, \ldots, z_n)$ is balanced for each value of $(z_2, z_3, \ldots, z_n)$ (i.e. f is linear in the first variable) or $f(z_1, \ldots, z_n)$ is balanced for each value of $(z_1, z_2, \ldots, z_{n-1})$ (i.e. f is linear in the last variable).*

For completeness, first we provide the proof of sufficient condition.[1]

Proof of sufficiency. Let $f \in \mathcal{L}_n \cup \mathcal{R}_n$. It is evident that for any $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)$, $N = \gamma_n + 1$ it follows that $f^\gamma \in \mathcal{L}_N \cup \mathcal{R}_N$ and thus (see Remark 3.3) $f^\gamma \in \mathcal{PB}_N$. Sufficiency follows from Statement 2.2. □

According to Dichtl [2], unproven necessary condition in Theorem 5.1 is referred to as Golić conjecture. Statement 2.2 implies that Golić conjecture can be stated in the following form.

Conjecture 5.2. *If f^γ is perfectly balanced for every possible choice of γ, then f is linear in the first or the last variable.*

Example 5.3. Boolean function $f(x_1, x_2, x_3, x_4) = x_2 \oplus x_1x_3 \oplus x_1x_3x_4 \in \mathcal{F}_4$, considered in [2], is perfectly balanced (this can be verified by applying Theorem 3.1, since the system $f_3(x_1, x_2, 0, x_4, x_5, x_6) = f_3(z_1, z_2, 1, x_4, x_5, x_6)$ is not solvable) and it is not linear neither in the first nor the last variable. According to Conjecture 5.2, there must exist tapping sequence γ, such that f^γ is not perfectly balanced.

Consider $\gamma = (0, 2, 3, 4)$, $N = 5$, $f^\gamma(x_1, x_2, x_3, x_4, x_5) = x_2 \oplus x_1x_3 \oplus x_1x_3x_5$. To prove that $f^\gamma \notin \mathcal{PB}_5$ it suffices to consider $\mathbf{x} = (0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0)$, $\mathbf{z} = (0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0)$ and to verify that all conditions of Theorem 3.1 are satisfied.

To prove Golić conjecture it suffices to construct for arbitrary $f \in \Phi_n \setminus (\mathcal{L}_n \cup \mathcal{R}_n)$ a particular tapping sequence making function f^γ not perfectly balanced. The key idea is to force γ_i increase exponentially in i. After choosing appropriate γ we construct two different binary sequences of the special form required by Sumarokov criterion (Theorem 3.1) to prove that f^γ is not perfectly balanced.

Theorem 5.4. *For any $f \in \Phi_n \setminus (\mathcal{L}_n \cup \mathcal{R}_n)$ there exists a tuple γ such that $f^\gamma \notin \mathcal{PB}_N$.*

Proof. Let $f \in \Phi_n \setminus (\mathcal{L}_n \cup \mathcal{R}_n)$. Suppose that f depends on each variable essentially (this is w.l.o.g. since we are free to choose any tuple γ) and exactly l variables are linear for f. Thus f is of the following form:

$$f(x_1, x_2, \ldots, x_n) = \bigoplus_{j=1,2,\ldots,l} x_{i_j^0} \oplus g(x_{i_1^1}, x_{i_2^1}, \ldots, x_{i_{n-l}^1}),$$

[1] Although the proof given in [3] formally invokes Lemma 1, the proof technique is essentially correct.

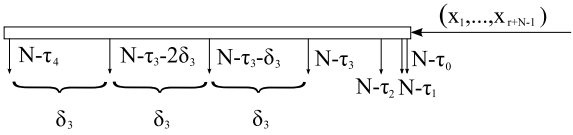

Fig. 3. Example of γ.

where $g \in \mathcal{F}_{n-l}$ depends on each variable essentially and nonlinearly, $\{i_1^0, i_2^0, \dots, i_l^0\} \cup \{i_1^1, i_2^1, \dots, i_{n-l}^1\} = \{1, 2, \dots, n\}$, $i_1^0 < i_2^0 < \dots < i_l^0$, $1 = i_1^1 < i_2^1 < \dots < i_{n-l}^1 = n$.

Let $m_0 = i_{n-l}^1 - i_{n-l-1}^1, m_1 = i_{n-l-1}^1 - i_{n-l-2}^1, \dots, m_{n-l-2} = i_2^1 - i_1^1, m = \max_{k=0,\dots,n-l-2} m_k$. Then

$$f(x_1, x_2, \dots, x_n) = g(x_1, x_{m_{n-l-2}+1}, \dots, x_{m_{n-l-2}+m_{n-l-3}+\dots+m_0+1}) \oplus \bigoplus_{j=2,3,\dots,m_{n-l-2}} x_j \oplus \bigoplus_{j=m_{n-l-2}+2,\dots,m_{n-l-2}+m_{n-l-3}} x_j \oplus \dots \oplus \bigoplus_{j=m_{n-l-2}+m_{n-l-3}+\dots+m_1+2,\dots,m_{n-l-2}+m_{n-l-3}+\dots+m_0} x_j.$$

Let $\tau_0 = 0, \tau_1 = m_0, \tau_{k+1} > (4m^2+1)\tau_k, k = 1, \dots, n-l-2$ and $\tau_{k+1} - \tau_k$ be a multiple of m_k; let $\delta_k = \frac{\tau_{k+1}-\tau_k}{m_k}$, $k = 1, 2, \dots, n-l-2$. Choose γ as follows:

$$\gamma = \big(\tau_0, \tau_0 + \delta_0, \dots, \tau_0 + (m_0-1)\delta_0, \tau_1, \dots, \tau_k, \tau_k + \delta_k, \dots, \tau_k + (m_k-1)\delta_k, \tau_{k+1}, \dots, \tau_{n-l-1}\big). \qquad \square$$

Example 5.5. Let $f(x_1, x_2, x_3, x_4, x_5, x_6, x_7) = x_1x_4x_5x_6x_7 \oplus x_2 \oplus x_3$. One has $n = 7$, $l = 2$, $m_0 = m_1 = m_2 = 1$, $m_3 = 3$, $m = 3$ and $\tau_0 = 0$, $\tau_1 = 1$, $\tau_2 = 38$, $\tau_3 = 1407$, $\tau_4 = 52062$, $\delta_3 = 16885$. Thus $N = 52063$ and $f^\gamma(x_1, x_2, \dots, x_{52063}) = f(x_1, x_{16886}, x_{33771}, x_{50656}, x_{52025}, x_{52062}, x_{52063})$ (see Fig. 3).

Consider two binary sequences $\mathbf{x} = (x_0, x_1, \dots, x_M), \mathbf{z} = (z_0, z_1, \dots, z_M)$, $M = 2N + \sum_{j=1}^{l'} \delta_{k_j}$, where k_j are indices such that $m_{k_j} > 1$ (l' denotes the total number of these indices). Fix certain bits of these sequences as follows: $x_{N+\sum_{j=1}^{l'} a_j\delta_{k_j}} = 0, z_{N+\sum_{j=1}^{l'} a_j\delta_{k_j}} = 1, \forall a_j \in \{0,1\}, j = 1, \dots, l'$.

Indices of the form $N + \sum_{j=1}^{l'} a_j\delta_{k_j}$ are referred to as B-indices and all the others as A-indices. It is easy to conclude, using Theorem 3.1, that to prove the theorem it suffices to show that one can set all yet unfixed bits of $\mathbf{x}$ so that ${f^\gamma}_{M-N+2}(\mathbf{y}) = {f^\gamma}_{M-N+2}(\mathbf{z})$ and $x_j = z_j$ holds for any A-index j. Thereby we have distinct binary sequences $\mathbf{x}$, $\mathbf{z}$ with coinciding leading as well as tailing N-bit subsequences and such that ${f^\gamma}_{M-N+2}(\mathbf{x}) = {f^\gamma}_{M-N+2}(\mathbf{z})$. Then, using Theorem 3.1, one concludes that γ is required tapping sequence, $f^\gamma \notin \mathcal{PB}_N$ and the theorem follows.

First, we demonstrate some simple relations.

1. $\delta_k = \frac{\tau_{k+1}-\tau_k}{m_k} > \frac{(1+4m^2)\tau_k - \tau_k}{m_k} \geq 4m\tau_k$.

2. If $m_{k-1} > 1$, then $\tau_k = \tau_{k-1} + m_{k-1}\delta_{k-1} \geq 2\delta_{k-1}$.
3. $\delta_k > \delta_{k-1}$. From 1 and 2 it follows that if $m_{k-1} > 1$, then $\delta_k > 8m\delta_{k-1}$.
4. From 3 it follows that $\sum_{j=j'}^{l'} \delta_{k_j} < \sum_{j=j'}^{l'} \delta_{k_{l'}} \frac{1}{(8m)^{l'-j}} < \sum_{i=0}^{\infty} \delta_{k_{l'}} \frac{1}{(8m)^i} = \frac{\delta_{k_{l'}}}{1-\frac{1}{8m}} = \delta_{k_{l'}} \frac{8m}{8m-1}$.
5. $\delta_k = \frac{\tau_{k+1}-\tau_k}{m_k} < \frac{\tau_{k+1}}{m_k} \leq \tau_{k+1}$ if $k \geq 1$; $\delta_0 \leq \tau_1$.

According to Theorem 3.1, to prove the theorem it suffices to prove solvability of the following system of equations:

$$\begin{cases} \begin{cases} f^\gamma(x_0, \ldots, x_{N-1}) = f^\gamma(z_0, \ldots, z_{N-1}); \\ \ldots \\ f^\gamma(x_{M-N+1}, \ldots, x_M) = f^\gamma(z_{M-N+1}, \ldots, z_M); \end{cases} \\ \begin{cases} x_{N+\sum_{j=1}^{l'} a_j \delta_{k_j}} = 0, \quad \forall a_j \in \{0,1\}, j = 1, \ldots, l'; \\ z_{N+\sum_{j=1}^{l'} a_j \delta_{k_j}} = 1, \quad \forall a_j \in \{0,1\}, j = 1, \ldots, l'; \\ x_t = z_t, \qquad t \neq N + \sum_{j=1}^{l'} a_j \delta_{k_j}, \quad \forall a_j \in \{0,1\}, j = 1, \ldots, l'. \end{cases} \end{cases} \tag{5.1}$$

Now we fix variables involved in the second subsystem of (5.1) and consider ith equation ($i = 0, \ldots, M - N + 1$) of the first subsystem. Three cases are possible.

Case 1. Each B-index variable, which is essential for $f^\gamma_{(i)} \equiv f^\gamma(x_i, \ldots, x_{i+N-1})$, is linear for $f^\gamma_{(i)}$.

In Lemma 5.6 (the proof can be found in Appendix B) we prove that in this case $f^\gamma_{(i)}$ depends on exactly two such variables. Then, from definition of linear dependence and from our fixation of B-index variables, we see that the ith equation of the first subsystem turns out to be identity.

Lemma 5.6. *Let the set of B-index variables that are essential for $f^\gamma_{(i)}$ be nonempty, and let $f^\gamma_{(i)}$ be linear in any B-index variable. Then $f^\gamma_{(i)}$ is linear in exactly two B-index variables.*

Case 2. $f^\gamma_{(i)}$ depends essentially on no B-index variable.

In this case we have a trivial equality, since variables with equal A-indices are equal, i.e. $x_j = z_j$, $j \neq N + \sum_{j=1}^{l'} a_j \delta_{k_j}$.

Case 3. $f^\gamma_{(i)}$ depends essentially and nonlinearly on some B-index variable $x_{j^{i-}}$.

Lemma 5.7 (the proof is in Appendix C) states that in this case $f^\gamma_{(i)}$ is nonlinear in exactly one essential B-index variable. In other words, if any other B-index variable is essential for $f^\gamma_{(i)}$, then the latter is linear essential variable of $f^\gamma_{(i)}$.

Lemma 5.7. *If $f^\gamma_{(i)}$ depends essentially and nonlinearly on some B-index variable, then there is exactly one such (nonlinear, essential, B-index) variable.*

Therefore, the ith equation of the system could be written in the form

$$\begin{aligned}&\phi(x_{j_1^i}, x_{j_2^i}, \dots, x_{\bar{j}^i-1}, 0, x_{\bar{j}^i+1}, \dots, x_{j_{n-l}^i})\\ &\quad = \phi(x_{j_1^i}, x_{j_2^i}, \dots, x_{\bar{j}^i-1}, 1, x_{\bar{j}^i+1}, \dots, x_{j_{n-l}^i}) \oplus \zeta_i,\end{aligned}$$

where ϕ is the function constructed from f by setting all linear essential variables to zero; $(x_{j_1^i}, x_{j_2^i}, \dots, x_{\bar{j}^i-1}, x_{\bar{j}^i+1}, \dots, x_{j_{n-l}^i})$ are yet unfixed variables and ζ_i is a constant. The variable $x_{\bar{j}^i}$ is essential and nonlinear for ϕ, thus there exists at least one setting of variables $(x_{j_1^i}, x_{j_2^i}, \dots, x_{\bar{j}^i-1}, x_{\bar{j}^i+1}, \dots, x_{j_{n-l}^i})$, which turns ith equation of the first subsystem of system (5.1) to identity. The Theorem is proven if one shows that no indices j_m^i appear in any other equation whose function satisfies conditions of the Case 3. In other words, each index of a nonlinear essential variable of $f_{(i)}^{\gamma}$ appears in at most one equation with function $f_{(i)}^{\gamma}$ depending essentially and nonlinearly on a B-index variable. This fact is proven in Lemma 5.8 (the proof is in Appendix D).

Lemma 5.8. *There is no index of a nonlinear essential variable of $f_{(j)}^{\gamma}$ (for any j) that occurs in at least two equations with functions $f_{(i)}^{\gamma}$ satisfying conditions of the Case* 3.

Also, each variable that is present in the equations corresponding to Case 3 is present only in one such equation. Equations which correspond to Case 1 and Case 2 turn into trivial equalities, and each equation corresponding to Case 3 is solvable. So, we can conclude that the whole system is solvable, and that fact directly implies statement of the theorem.

Remark 5.9. Proof of Theorem 5.4 is much easier in the case of f without linear essential variables. In this case one has $l = 0$, $m_k = 1$, $k = 0, \dots, n-2$; $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)$, where $\gamma_1 = 0$ and $\gamma_i = 6^{i-2}$, $i = 1, 2, \dots, n$; $N = 6^{n-2} + 1$. The pair of sequences $(\mathbf{x}, \mathbf{z})$ is the solution of the following system of equations:

$$\begin{cases} f_{2N+1}^{\gamma}(\mathbf{x}) = f_{2N+1}^{\gamma}(\mathbf{z}); \\ x_{N+1} = 0, \qquad z_{N+1} = 1; \\ x_i = z_i, \quad i = 1, 2, \dots, N; \\ x_i = z_i, \quad i = N+2, N+3, \dots, 2N+1. \end{cases}$$

6. Conclusion and Open Questions

The property of preserving pure randomness of an input binary sequence (perfect balancedness) is one of the most important for filter functions. However, there was a number of incorrect results in the previous works on this subject [2,3]. We demonstrate errors in the incorrect proofs in [2,3] and provide necessary counterexamples.

In Sect. 4 we present a new approach that can be used to construct large classes of perfectly balanced functions which are essential and nonlinear in the first and the last variable. Using this approach we obtain a new lower bound $(2^{2^{n-3}+1} - 5 \cdot 2^{2^{n-4}})$ of the

cardinality of the set $(\mathcal{PB}_n \cap \Phi_n) \setminus (\mathcal{L}_n \cup \mathcal{R}_n)$. Interesting open problems are obtaining upper bounds (tighter than trivial ones such as $\binom{2^n}{2^{n-1}} - 2\binom{2^{n-1}}{2^{n-2}} + \binom{2^{n-2}}{2^{n-3}} - 2^{2^{n-1}+1} + 3 \cdot 2^{2^{n-2}}$) and tighter lower bounds of the cardinality of this set.

Theorem 5.4 implies the negative answer to the question of existence (in Golić model) of keystream generators without undesirable properties mentioned in introduction. But our proof is based on a register whose size exponentially grows with the number of taps. Thus, though theoretically the question with Golić conjecture is now closed, there remains the following open question: whether it is possible to prove a similar statement without forcing a sequence γ increase exponentially (e.g. in the model where the size of a register is bounded by some polynomial).

Acknowledgements

The work was supported by the Russian Foundation for Basic Research (Grant No. 09-01-00653).

The author is very grateful to Oleg A. Logachev and Nick P. Varnovsky for valuable discussions and suggestions concerning this work and to anonymous reviewers for their careful reading and useful comments.

Appendix A

Proof of Theorem 3.1 (Sumarokov [11]). Denote by $\gamma(f, l)$ the maximum possible (over all $(y_1, y_2, \ldots, y_l) \in V_l$) number of solutions to the system

$$\begin{cases} f(x_s, x_{s+1}, \ldots, x_{s+n-1}) = y_s; \\ s = 1, 2, \ldots, l. \end{cases} \tag{A.1}$$

It is evident that if for some f there are no sequences $\mathbf{x}$, $\mathbf{z}$ such that (3.1)–(3.4) hold, then the output sequence $(y_1, y_2, \ldots, y_{r-n+1}) = f_{r-n+1}(\mathbf{x})$ and $x_1, x_2, \ldots, x_n$; $x_{r-n+1}, x_{r-n+2}, \ldots, x_r$ determine the whole input sequence $\mathbf{x} = (x_1, x_2, \ldots, x_n, x_{n+1}, \ldots, x_r)$, and thus for any integer l inequality $\gamma(f, l) \le 2^{2n-2}$ holds. In the opposite case (i.e. there exists a pair of distinct sequences $\tilde{\mathbf{x}} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{\tilde{r}})$, $\tilde{\mathbf{z}}(\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_{\tilde{r}})$ such that $f_{\tilde{r}-n+1}(\tilde{\mathbf{x}}) = f_{\tilde{r}-n+1}(\tilde{\mathbf{z}})$ and $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) = (\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_n)$, $(\tilde{x}_{\tilde{r}-n+1}, \tilde{x}_{\tilde{r}-n+2}, \ldots, \tilde{x}_{\tilde{r}}) = (\tilde{z}_{\tilde{r}-n+1}, \tilde{z}_{\tilde{r}-n+2}, \ldots, \tilde{z}_{\tilde{r}})$) there are at least 2^t solutions to the system

$$\begin{cases} f(x_s, x_{s+1}, \ldots, x_{s+n-1}) = y_s; \\ s = 1, 2, \ldots, t \cdot \tilde{r} - n + 1; \end{cases}$$

where $y_s = f(\tilde{x}_{s \bmod \tilde{r}}, \tilde{x}_{(s+1) \bmod \tilde{r}}, \ldots, \tilde{x}_{(s+n-1) \bmod \tilde{r}})$, $s = 1, 2, \ldots, t \cdot \tilde{r} - n + 1$, and thus $\gamma(f, l)$ is unbounded as a function of l with $l \to \infty$.

In the remaining part of the proof it is shown that $\gamma(f, l)$ is bounded (with $l \to \infty$) iff f is perfectly balanced.

By definition, for any $f \in \mathcal{PB}_n$ and any natural l $\gamma(f,l) = 2^{n-1}$, i.e. $\gamma(f,l)$ is not unbounded with $l \to \infty$. Let $f \notin \mathcal{PB}_n$. Then there is an integer l and a tuple $\widetilde{\mathbf{y}} = (\widetilde{y}_1, \widetilde{y}_2, \dots, \widetilde{y}_l) \in V_l$, such that there exist $2^{n-1} + \alpha$ solutions to (A.1), where $\alpha \geq 1$.

For the tuple $\widetilde{\mathbf{y}} \in V_l$ construct the set of all possible sequences of length $(k+1)l + k(n-1)$ of the following form:

$$\widetilde{y}_1, \dots, \widetilde{y}_l, y_{l+1}, \dots, y_{l+n-1}, \widetilde{y}_1, \dots, \widetilde{y}_l, y_{2l+n-1}, \dots, y_{2l+2(n-1)}, \dots,$$
$$y_{kl+(k-1)(n-1)+1}, \dots, y_{kl+k(n-1)}, \widetilde{y}_1, \dots, \widetilde{y}_l, \tag{A.2}$$

$k = 1, 2, \dots$, where $y_i \in \mathbb{F}_2$, $i = l+1, l+2, \dots, l+n-1; 2l+n-1, 2l+n-2, \dots$. Let μ_k denote the average number of inputs of $f_{(k+1)l+k(n-1)}$ that correspond to one output of the form (A.2) (including unreachable outputs, if they exist). In this case,

$$\mu_k = 2^{n-1}\left(1 + \frac{\alpha}{2^{n-1}}\right)^{k+1},$$

so $\mu_k \to \infty$ with $k \to \infty$. That is, for any integer M there is an integer $k = k(M)$ such that $\mu_{k(M)} > M$, i.e. preimage of one of the sequences (A.2) of length $t(M) = (k(M)+1)l + k(M)(n-1)$ is of cardinality greater than M. This means that for arbitrary M there exists $t(M)$ such that $\gamma(f, t(M)) > M$ and thus $\gamma(f,l)$ is unbounded as a function of l. □

Appendix B

Proof of Lemma 5.6. Consider the set of all essential B-index variables of $f^{\gamma}_{(i)}$ and let the variable in this set with the maximal B-index correspond to the $(N - \tau_k - r\delta_k)$th variable of f^{γ}, $1 \leq r \leq m_k - 1$. It is evident that in this case there is another B-index variable corresponding to $(N - \tau_k - (r+1)\delta_k)$th variable of f^{γ}. According to conditions of Case 1, this variable is linear as well. Therefore $1 \leq r \leq m_k - 2, m_k \geq 3$. Next one has to prove that no other B-index variable is essential for $f^{\gamma}_{(i)}$.

It suffices to show that variables of f^{γ} with indices $N - \tau_k - r\delta_k - \sum_{j=1}^{l'} b_j \delta_{k_j}$, $b_j \in \{-1, 0, 1\}$, $j = 1, \dots, l'$ are not essential for f^{γ} except for two trivial cases ($\sum_{j=1}^{l'} b_j \delta_{k_j} = \delta_k$ and $\sum_{j=1}^{l'} b_j \delta_{k_j} = 0$).

Let $k_{j^*} = k$. Two cases are possible.

(1) $\exists j^\circ > j^* : b_{j^\circ} \neq 0$ and let j° be the maximal index j such that $b_j \neq 0$. Evidently, it suffices to consider the case of $b_{j^\circ} = 1$. Then $N - \tau_k - r\delta_k - \sum_{j=1}^{l'} b_j \delta_{k_j} \leq N - \tau_k - r\delta_k - \delta_{k_{j^\circ}} + \sum_{j=1}^{j^\circ - 1} \delta_{k_j} < N - \tau_k - r\delta_k - 4m(1 - \frac{1}{8m-1})\tau_{k_{j^\circ}} < N - \tau_{k_{j^\circ}}$.

Also, the following inequality holds. $N - \tau_k - r\delta_k - \sum_{j=1}^{l'} b_j \delta_{k_j} = N - (\tau_{k+1} - (m_k - r)\delta_k) - \sum_{j=1}^{l'} b_j \delta_{k_j} \geq N - (\tau_{k+1} - 2\delta_k) - \sum_{j=1}^{l'} b_j \delta_{k_j} \geq N - (\tau_{k+1} - 2\delta_k) - \sum_{j=1}^{j^\circ} \delta_{k_j} > N - (\tau_{k+1} - 2\delta_k) - \delta_{k_{j^\circ - 1}} \frac{8m}{8m-1} - \delta_{k_{j^\circ}}$. If $k_{j^\circ} - 1 = k$, then $k_{j^\circ - 1} = k$ and one can estimate the last expression as follows: $N - (\tau_{k+1} - 2\delta_k) - \delta_{k_{j^\circ - 1}} \frac{8m}{8m-1} - \delta_{k_{j^\circ}} = N - \tau_{k+1} + \delta_k - \frac{\delta_k}{8m-1} - \delta_{k+1} > N - \tau_{k+1} - \delta_{k+1} = N - \tau_{k_{j^\circ}} - \delta_{k_{j^\circ}}$. Else

$N-(\tau_{k+1}-2\delta_k)-\delta_{k_{j^\circ-1}}\frac{8m}{8m-1}-\delta_{k_{j^\circ}}>N-\tau_{k+1}-\delta_{k_{j^\circ-1}}\frac{8m}{8m-1}-\delta_{k_{j^\circ}}\geq N-\tau_{k+1}-\delta_{k_{j^\circ-1}}\frac{8m}{8m-1}-\delta_{k_{j^\circ}}\geq N-\tau_{k_{j^\circ-1}}-\delta_{k_{j^\circ-1}}\frac{8m}{8m-1}-\delta_{k_{j^\circ}}>N-\delta_{k_{j^\circ-1}}(\frac{1}{4m}+\frac{8m}{8m-1})-\delta_{k_{j^\circ}}=N-\frac{\tau_{k_{j^\circ}}-\tau_{k_{j^\circ-1}}}{m_{k_{j^\circ}}}(\frac{1}{4m}+\frac{8m}{8m-1})-\delta_{k_{j^\circ}}>N-\frac{\tau_{k_{j^\circ}}}{m_{k_{j^\circ}}}(\frac{1}{4m}+1+\frac{1}{8m-1})-\delta_{k_{j^\circ}}\geq N-\frac{\tau_{k_{j^\circ}}}{m_{k_{j^\circ}}}(\frac{1}{12}+1+\frac{1}{23})-\delta_{k_{j^\circ}}\geq N-\frac{\tau_{k_{j^\circ}}}{2}(\frac{1}{12}+1+\frac{1}{23})-\delta_{k_{j^\circ}}>N-\tau_{k_{j^\circ}}-\delta_{k_{j^\circ}}$. This implies that all the variables with indices $j^\circ>j^*, b_{j^\circ}=1$ occur in the interval $(N-\tau_{k_{j^\circ}}-\delta_{k_{j^\circ}}, N-\tau_{k_{j^\circ}})$ and thus could not be essential for f^γ.

(2) $\forall j>j^*\Rightarrow b_j=0; \exists j^\circ<j^*: b_{j^\circ}\neq 0$ (if there are multiple such j°, we choose the largest one).

If $b_{j^*}=1$, then $N-\tau_k-r\delta_k-\sum_{j=1}^{l'}b_j\delta_{k_j}>N-\tau_k-r\delta_k-\frac{8m}{8m-1}\delta_k>N-\tau_k-(r+2)\delta_k$; $N-\tau_k-r\delta_k-\sum_{j=1}^{l'}b_j\delta_{k_j}<N-\tau_k-r\delta_k, N-\tau_k-r\delta_k-\sum_{j=1}^{l'}b_j\delta_{k_j}\neq N-\tau_k-(r+1)\delta_k$.

If $b_{j^*}=0$, then $N-\tau_k-r\delta_k-\sum_{j=1}^{l'}b_j\delta_{k_j}>N-\tau_k-r\delta_k-\frac{1}{8m-1}\delta_k>N-\tau_k-(r+1)\delta_k$; $N-\tau_k-r\delta_k-\sum_{j=1}^{l'}b_j\delta_{k_j}<N-\tau_k-(r-1)\delta_k, N-\tau_k-r\delta_k-\sum_{j=1}^{l'}b_j\delta_{k_j}\neq N-\tau_k-r\delta_k$.

Thus, in this case variables are not essential too. □

Appendix C

Proof of Lemma 5.7. By contradiction, let for some $f^\gamma_{(i)}$ two B-index variables correspond to the $(N-\tau_k)$th and $(N-\tau_p)$th variables of f^γ, $p>k$. Then

$$\tau_p-\tau_k=\sum_{j=1}^{l'}b_j\delta_{k_j},\quad b_j\in\{-1,0,1\}. \tag{C.1}$$

(1) Let the set $K=\{j|k_j\geq p, b_j=1\}$ be nonempty and let j° be the maximum element of this set. Then $\sum_{j=1}^{l'}b_j\delta_{k_j}\geq\delta_{k_{j^\circ}}-\sum_{j=1}^{j^\circ-1}\delta_{k_j}>\delta_{k_{j^\circ}}-\frac{8m}{8m-1}\delta_{k_{j^\circ-1}}>\delta_{k_{j^\circ}}(1-\frac{1}{8m-1})>4m(1-\frac{1}{8m-1})\tau_p>\tau_p-\tau_k$.

(2) Let K be empty, i.e. $b_{j^\circ}\leq 0, k_{j^\circ}\geq p$. Then $\sum_{j=1}^{l'}b_j\delta_{k_j}\leq\sum_{j=1}^{j_p}\delta_{k_j}<\frac{8m}{8m-1}\delta_{k_{j_p}}$, where $k_{j_p}\leq p-1$. $\frac{8m}{8m-1}\delta_{k_{j_p}}=\frac{8m}{8m-1}\frac{\tau_{k_{j_p}+1}-\tau_{k_{j_p}}}{m_{k_{j_p}}}\leq\frac{8}{7}\frac{\tau_{k_{j_p}+1}-\tau_{k_{j_p}}}{2}\leq\frac{8}{7}\frac{\tau_p-\tau_{p-1}}{2}<\tau_p-\tau_{p-1}\leq\tau_p-\tau_k$.

Hence, (C.1) is impossible and this concludes the proof of the lemma. □

Appendix D

Proof of Lemma 5.8. We have to prove that the equality

$$\tau_a-\tau_b+\sum_{j=1}^{l'}a'_j\delta_{k_j}=\tau_c-\tau_d+\sum_{j=1}^{l'}a''_j\delta_{k_j} \tag{D.1}$$

does not hold if conditions

$$\begin{cases} \tau_a = \tau_c; \\ \tau_b = \tau_d; \\ a'_j = a''_j, \quad j = 1, \dots, l' \end{cases} \tag{D.2}$$

are not satisfied.

First, we prove that equality $\sum_{j=1}^{l'} a'_j \delta_{k_j} = \sum_{j=1}^{l'} a''_j \delta_{k_j}$ holds only if $a'_j = a''_j$, $j = 1, \dots, l'$. Let $a'_{j^\circ} \neq a''_{j^\circ}$, $a'_{j^\circ} = 1, a''_{j^\circ} = 0$ and let j° be the largest index such that $a'_j \neq a''_j$. Then

$$\sum_{j=1}^{l'} a'_j \delta_{k_j} - \sum_{j=1}^{l'} a''_j \delta_{k_j} = \delta_{k_{j^\circ}} + \sum_{j=1}^{j^\circ-1} a'_j \delta_{k_j} - \sum_{j=1}^{j^\circ-1} a''_j \delta_{k_j}$$

$$\geq \delta_{k_{j^\circ}} - \sum_{j=1}^{j^\circ-1} a''_j \delta_{k_j} \geq \delta_{k_{j^\circ}} - \sum_{j=1}^{j^\circ-1} \delta_{k_j} > \delta_{k_{j^\circ}} - \frac{1}{8m-1} \delta_{k_{j^\circ}} > 0.$$

Consider indices $a, b, c, d, e+1$, $e = k_{j^\circ}$, where j° is the largest index such that $a'_j \neq a''_j$. One can transform (D.1) as follows: $\tau_a - \tau_b = \tau_c - \tau_d + \sum_{j=1}^{j^\circ} b_j \delta_{k_j}$, $b_j = a''_j - a'_j$, $j = 1, \dots, j^\circ$.

Let $q = \max\{a, b, c, d, e+1\}$. We have (up to equivalence) five possibilities.

(1) $q = a, q > b, q > c, q > d, q > e+1$. Then $\tau_a = \tau_q > (4m^2+1)\tau_{q-1} \geq 5\tau_{q-1} \geq \tau_b + (\tau_c - \tau_d) + 3\tau_{q-1} \geq \tau_b + \tau_c - \tau_d + 3\delta_{q-2} > \tau_b + \tau_c - \tau_d + \delta_{q-2}\frac{8m}{8m-1} > \tau_b + \tau_c - \tau_d + \sum_{j=1}^{j^\circ} \delta_{k_j} \geq \tau_b + \tau_c - \tau_d + \sum_{j=1}^{j^\circ} b_j \delta_{k_j}$, hence equality (D.1) does not hold.

(2) $q = e+1, a \leq e, b \leq e, c \leq e, d \leq e$. Let $b_{j^\circ} = 1$ (the case of $b_{j^\circ} = -1$ is treated along the same lines). $\delta_e > 4m\tau_e > 2\tau_e + 2\delta_{e-1} > (\tau_a - \tau_b + \tau_d - \tau_c) + \delta_{e-1} + \frac{1}{8m-1}\delta_{e-1} > \tau_a - \tau_b + \tau_d - \tau_c + \sum_{j=1}^{j^\circ-1} \delta_{k_j} \geq \tau_a - \tau_b + \tau_d - \tau_c + \sum_{j=1}^{j^\circ-1} b_j \delta_{k_j}$, thus equality (D.1) does not hold.

(3) $q = a = c$. Then (D.1) can be transformed into $\tau_d = \tau_b + \sum_{j=1}^{j^\circ} b_j \delta_{k_j}$. If $b = d$, then (D.1) turns into $\sum_{j=1}^{j^\circ} b_j \delta_{k_j} = 0$, which holds only if $b_j = 0$, $j = 1, \dots, j^\circ$.

If $d > b$ (or $d < b$, that can be treated similarly), we denote $q' = \max\{b, d, e+1\}$ and consider three subcases.

- $d = q' > e+1$. Then $\tau_d > (4m^2+1)\tau_{q'-1} > \tau_b + 4m^2\tau_{q'-1} > \tau_b + 4m^2\delta_e > \tau_b + \frac{8m}{8m-1}\delta_e > \tau_b + \sum_{j=1}^{j^\circ} \delta_{k_j} \geq \tau_b + \sum_{j=1}^{j^\circ} b_j \delta_{k_j}$.
- $q' = e+1 > d$. Then $\sum_{j=1}^{j^\circ} b_j \delta_{k_j} > 4m\tau_{q'-1} - \sum_{j=1}^{j^\circ-1} \delta_{k_j} \geq \tau_d + \tau_b + 2\tau_{q'-1} - \sum_{j=1}^{j^\circ-1} \delta_{k_j} > \tau_d + \tau_b + 2\delta_{q'-2} - \sum_{j=1}^{j^\circ-1} \delta_{k_j} = \tau_d + \tau_b + 2\delta_{e-1} - \sum_{j=1}^{j^\circ-1} \delta_{k_j} > \tau_d + \tau_b + \frac{8m}{8m-1}\delta_{k_{j^\circ-1}} - \sum_{j=1}^{j^\circ-1} \delta_{k_j} > \tau_d + \tau_b$.
- $q' = e+1 = d$. Then $\tau_d > \frac{3}{4}\tau_d + m^2\tau_{q'-1} \geq \frac{3\tau_{e+1}}{2m_{j^\circ}} + \tau_b > \frac{8m}{8m-1}\delta_e + \tau_b > \tau_b + \sum_{j=1}^{j^\circ} \delta_{k_j} \geq \tau_b + \sum_{j=1}^{j^\circ} b_{k_j} \delta_{k_j}$.

In fact, other subcases are possible but each of them is equivalent to one of the above.

(4) $q = a = d, b < q, c < q$. Then $e + 1 \le q$ and hence $\tau_a + \tau_d > (4m^2 + 1)\tau_{q-1} + \tau_q > \tau_c + \tau_b + \tau_q \ge \tau_c + \tau_b + 2\delta_e > \tau_c + \tau_b + \frac{8m}{8m-1}\delta_e > \tau_c + \tau_b + \sum_{j=1}^{j^\circ} b_j \delta_{k_j}$, thus (D.1) does not hold in this case either.

(5) $q = a = e + 1, b < q, c < q, d < q$. Then $\tau_a = \frac{3\tau_a}{4} + \frac{\tau_a}{4} > \frac{3\tau_{e+1}}{4} + m^2\tau_{q-1}$. $e = k_{j^\circ}$, so $m_e \ge 2, m \ge 2$. Then $\frac{3\tau_{e+1}}{4} + m^2\tau_{q-1} > \frac{3\tau_{e+1}}{2m_e} + 3\tau_{q-1} > \frac{3}{2}\delta_e + 3\tau_{q-1} > (1 + \frac{1}{8m-1})\delta_e + \tau_b + (\tau_c - \tau_d) > \sum_{j=1}^{j^\circ} \delta_{k_j} + \tau_b + (\tau_c - \tau_d) \ge \sum_{j=1}^{j^\circ} b_j \delta_{k_j} + \tau_b + \tau_c - \tau_d$. This implies that (D.1) does not hold in this case either. □

References

[1] R.J. Anderson, Searching for the optimum correlation attack, in *Fast Software Encryption*, ed. by B. Preneel. LNCS, vol. 1008 (Springer, Heidelberg, 1995), pp. 137–143

[2] M. Dichtl, On nonlinear filter generators, in *FSE 1997*, ed. by E. Biham. LNCS, vol. 1267 (Springer, Heidelberg, 1997), pp. 103–106

[3] J.Dj. Golić, On the security of nonlinear filter generators, in *Proceedings of Fast Software Encryption 1996*, ed. by D. Gollmann. LNCS, vol. 1039 (Springer, Heidelberg, 1996), pp. 173–188

[4] J.Dj. Golić, Conditional correlation attack on combiners with memory. *Electron. Lett.* **32**(24), 2193–2195 (1996)

[5] J.Dj. Golić, A. Clark, E. Dawson, Generalized inversion attack on nonlinear filter generators. *IEEE Trans. Comput.* **C-49**, 1100–1109 (2000)

[6] A. Gouget, H. Sibert, Revisiting correlation-immunity in filter generators, in *Proceedings of SAC 2007*, ed. by C. Adams, A. Miri, M. Wiener. LNCS, vol. 4876 (Springer, Heidelberg, 2007), pp. 378–395

[7] O.A. Logachev, On perfectly balanced Boolean functions. Cryptology ePrint Archive, Report 2007/022. http://eprint.iacr.org/

[8] O.A. Logachev, A.A. Salnikov, S.V. Smyshlyaev, V.V. Yashchenko, Perfectly balanced functions in symbolic dynamics, in *Proceedings of the NATO Advanced Research Workshop on Enhancing Cryptographic Primitives with Techniques from Error Correcting Codes*, Veliko Tarnovo, Bulgaria, 6–9 October 2008 (IOS Press, Amsterdam, 2009), pp. 222–233

[9] O.A. Logachev, S.V. Smyshlyaev, V.V. Yashchenko, New methods of investigation of perfectly balanced Boolean functions. *Discrete Math. Appl.* **19**(3), 237–262 (2009)

[10] S.V. Smyshlyaev, Barriers of perfectly balanced Boolean functions. *Discrete Math. Appl.* **20**(3), 321–336 (2010)

[11] S.N. Sumarokov, Functions of defect zero and invertibility of one class of finite-memory encoders. *Obozr. Prom. Prikl. Mat.* **1**(1), 33–55 (1994) (in Russian)