

Perfectly Balanced Functions in Symbolic Dynamics*

Oleg A. Logachev¹, Alexei A. Salnikov², Stanislav V. Smyshlyaev³ and Valery V. Yashchenko⁴

Abstract: In the present paper we study properties of perfectly balanced Boolean functions. Based on the concept of Boolean function barrier, we propose a novel approach to construct large classes of perfectly balanced Boolean functions.

Keywords: Symbolic dynamics, code, filter generator, perfectly balanced function, barrier.

Introduction

A nonfeedback shift register with nonlinear filtering function (finite memory encoder [1, 2]) implements a convolution transformation of binary sequences. A finite memory encoder feeded with linear recurrence sequence turns out to be a filter generator. One of the aims of studying such shift registers is to provide for acceptable statistical properties of output sequences. The best statistical properties of output sequences are reached when using perfectly balanced functions [3, 4].

The paper is organized as follows. In Section 1 we remind some basic concepts of symbolic dynamics. Properties of perfectly balanced Boolean functions in the general case are studied in Section 2. Binary case is treated in Section 3. In Section 4 we propose based on the concept of Boolean function barrier a novel approach to construct large classes of perfectly balanced Boolean functions. In Appendix we present a classification of perfectly balanced Boolean functions of $n = 4$ and $n = 5$ variables.

1 Background

In the present section we give some background on symbolic dynamics. For a more detailed exposition, the reader is referred to [5, 6].

Let I denote the set of all integers. For $i \in I$, let $I_i = \{j \in I \mid j \geq i\}$. The cardinality of a set E will be denoted by $\text{card } E$.

Let $p \in I_2$. In what follows, the set $\mathcal{A}_p = \{0, 1, \dots, p - 1\}$ is treated as the set of symbols. A bisequence over \mathcal{A}_p is a mapping from I to \mathcal{A}_p .

Let $\Sigma_p = \{x = (\dots x_{-1} x_0 x_1 \dots) \mid x_i \in \mathcal{A}_p, i \in I\}$ be the set of all bisequences over \mathcal{A}_p . For any $x \in \Sigma_p$ and any $i \in I$ let $(x)_i = x_i$.

A metrics d on Σ_p is defined as follows. Let $x, y \in \Sigma_p$. Then

$$d(x, y) = \begin{cases} 0 & \text{if } x = y, \\ (1 + k)^{-1} & \text{if } x \neq y, \end{cases}$$

where $k = \max\{m \in I_0 \mid x_i = y_i \text{ for } |i| < m\}$.

It is easily verified that the metric topology induced by d coincides with the product topology induced by discrete topology of \mathcal{A}_p . Therefore Σ_p is a compact, totally disconnected, perfect, metric space and hence is homeomorphic

*The work was partially supported by the Russian Foundation for Basic Research (grant no. 07-01-00154 for the first, second and fourth author and grant no. 09-01-00653 for the first and third author).

¹Information Security Institute, Lomonosov University, Moscow, Russia; E-mail: logol@iisi.msu.ru

²Information Security Institute, Lomonosov University, Moscow, Russia; E-mail: orfeo@rambler.ru

³Computer Science Department, Lomonosov University, Moscow, Russia; E-mail: smyshsv@gmail.com

⁴Information Security Institute, Lomonosov University, Moscow, Russia; E-mail: iisi@iisi.msu.ru

to the Cantor discontinuum. The shift transformation σ acts on Σ_p bijectively. The shift σ takes $x \in \Sigma_p$ to $y \in \Sigma_p$, where for any $i \in I$

$$y_i = (y)_i = [\sigma(x)]_i = x_{i+1}.$$

The shift σ is a homeomorphism of Σ_p .

A symbolic dynamical system is a pair (S, σ) , where S is a closed and shift-invariant (i.e., $\sigma(S) = S$) subset of Σ_p . Such a system is called also subshift. For example, (Σ_p, σ) is a symbolic dynamical system, often called the full shift, as opposed to all other shift, which are called subshifts and define embeddings into Σ_p .

Let S be a closed nonempty σ -invariant subset of Σ_p . Then the restriction of σ to S is a homeomorphism of S onto S . It is clear that S is a compact totally disconnected metric space.

Let (S_i, σ) , $i = 1, 2$, be two subshifts ($S_i \subset \Sigma_{p_i}$, $i = 1, 2$). A continuous mapping $C: S_1 \rightarrow S_2$ is a code if $\sigma \circ C = C \circ \sigma$. An injective code is an embedding; a bijective code defines a topological conjugacy of subshifts is called isomorphism. In the case $S_1 = S_2 = S$ the code C is an endomorphism of the dynamical system (S, σ) .

Let $n \in I_1$. An n -block (word) over \mathcal{A}_p is an ordered set $x_1 x_2 \dots x_n$, where $x_i \in \mathcal{A}_p$ ($i = 1, 2, \dots, n$). Let $\mathcal{B}_{n,p}$ denote the set of all n -blocks over \mathcal{A}_p .

Let $f: \mathcal{B}_{n,p} \rightarrow \mathcal{B}_{1,p} = \mathcal{A}_p$. The set of all such mappings for a given $n \in I_1$ will be denoted by $\mathcal{F}_{n,p}$. Note that $\text{card } \mathcal{B}_{n,p} = p^n$ and $\text{card } \mathcal{F}_{n,p} = p^{p^n}$.

For any $f \in \mathcal{F}_{n,p}$, left $f_\infty: \Sigma_p \rightarrow \Sigma_p$ be a mapping of the form

$$f_\infty(x) = y, \quad y_i = f(x_i x_{i+1} \dots x_{i+n-1}),$$

where $x, y \in \Sigma_p$, $x_i, y_i \in \mathcal{A}_p$, $i \in I$. It is easily verified that a mapping f_∞ is a code for $S_1 = S_2 = \Sigma_p$. It is usually called an n -block code.

In the case $S = \Sigma_p$ denote by \mathcal{E}_p the set of all endomorphisms (codes) of symbolic dynamical system (Σ_p, σ) .

Let \mathcal{F}_p denote the set $\bigcup_{n=1}^{\infty} \mathcal{F}_{n,p}$ and let $\mathcal{F}_p^* = \{\sigma^m f_\infty \mid m \in I, f \in \mathcal{F}_p\}$.

The structure of endomorphism (code) set of full shift (Σ_p, σ) and its subshifts was described by Curtis, Hedlund, and Lyndon and can be summarized in the next theorem.

Theorem 1 ([7]). $\mathcal{E}_p = \mathcal{F}_p^*$.

2 Perfectly Balanced Functions

In what follows, we use the term “function” for elements of the set \mathcal{F}_p only as opposed to all other mappings. This section recalls the general properties of perfectly balanced functions. We also need some notation and definitions.

Let $f \in \mathcal{F}_{n,p}$ and $m \in I_1$. Define a mapping

$$f_m: \mathcal{B}_{m+n-1,p} \rightarrow \mathcal{B}_{m,p}$$

as follows:

$$f_m(b_1 b_2 \dots b_{m+n-1}) = a_1 a_2 \dots a_m,$$

where

$$a_i = f(b_i b_{i+1} \dots b_{i+n-1}), \quad i = 1, 2, \dots, m. \quad (1)$$

Let

$$J(f, m) = \{y \in \mathcal{B}_{m,p} \mid \forall x \in \mathcal{B}_{m+n-1,p} \ f_m(x) \neq y\}.$$

It is evident that $\text{card } J(f, m) = \text{Def}_m(f)$.

Definition 1. A function $f \in \mathcal{F}_{n,p}$ is called perfectly balanced if $\text{card}(f_m^{-1})(y) = p^{n-1}$ for any $m \in I_1$ and for every $y \in \mathcal{B}_{m,p}$.

Definition 2. A function $f \in \mathcal{F}_{n,p}$ is said to be of defect zero if $\text{Def}_m(f) = 0$ for any $m \in I_1$.

Let $b \in \mathcal{B}_{l,p}$ and $E = \{j_1, j_2, \dots, j_k\} \subset \{1, 2, \dots, l\}$, $1 \leq j_1 < \dots < j_k \leq l$, $k \leq l$. We denote by $(b)_E$ a string of the form $(b_{j_1} b_{j_2} \dots b_{j_k})$.

Definition 3. A function $f \in \mathcal{F}_{n,p}$ is called information-lossless if for any $l > 2n$ there do not exist two different blocks $b, b' \in \mathcal{B}_{l,p}$ such that

$$(b)_{\{1,2,\dots,n\}} = (b')_{\{1,2,\dots,n\}}, \quad (b)_{\{l-n+1,l-n+2,\dots,l\}} = (b')_{\{l-n+1,l-n+2,\dots,l\}},$$

and

$$f_{l-n+1}(b) = f_{l-n+1}(b').$$

Denote by \mathcal{E}_p° the set of all surjective mappings in \mathcal{E}_p . The structure of the set \mathcal{E}_p° is closely related to the set of perfectly balanced functions. The next statement is due to Blankenship, Hedlund, Rothaus, and Sumarokov.

Theorem 2 ([7, 8]). *Let $n \in I_1$ and $f \in \mathcal{F}_{n,p}$. Then the following statements are equivalent:*

1. f_∞ is a surjective endomorphism of (Σ_p, σ) , i.e., $f_\infty \in \mathcal{E}_p^\circ$;
2. f is perfectly balanced;
3. f is of defect zero;
4. f is information-lossless.

Let $n \in I_2$ and let $f \in \mathcal{F}_{n,p}$. Let $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n$ ($1 \leq i \leq n$) be a collection of fixed elements of \mathcal{A}_p . Then

$$g(x_i) = f(a_1 \dots a_{i-1} x_i a_{i+1} \dots a_n)$$

defines a function g of \mathcal{A}_p into \mathcal{A}_p , i.e., $g \in \mathcal{F}_{1,p}$. If for any $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n$ the function g is a permutation of \mathcal{A}_p , then we call f permutive in x_i . If $n = 1$, then f maps \mathcal{A}_p into \mathcal{A}_p . A function f is permutive in x_i if its restriction to \mathcal{A}_p is a permutation of \mathcal{A}_p .

Let $f \in \mathcal{F}_{p,n}$. A variable x_i ($1 \leq i \leq n$) is called essential if there exist $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n, b, c \in \mathcal{A}_p$ such that $b \neq c$ and $f(a_1 \dots a_{i-1} ba_{i+1} \dots a_n) \neq f(a_1 \dots a_{i-1} ca_{i+1} \dots a_n)$.

Let M be a subset of $\mathcal{F}_{n,p}$. We denote the subset of all functions in M with essential variables x_1, x_n by \widetilde{M} . The subset of all perfectly balanced functions in $\mathcal{F}_{n,p}$ is denoted by $\mathcal{PBF}_{n,p}$. Furthermore, let $\mathcal{L}_{n,p}$ ($\mathcal{R}_{n,p}$) denote the subset of all functions in $\mathcal{F}_{n,p}$ that are permutive in x_1 (resp., x_n). It is easily verified that

$$\widetilde{\mathcal{L}}_{n,p} \subset \mathcal{PBF}_{n,p}, \quad \widetilde{\mathcal{R}}_{n,p} \subset \mathcal{PBF}_{n,p}. \quad (2)$$

In the remaining sections we study the structure of the set $\mathcal{PBF}_{n,p} \setminus (\widetilde{\mathcal{L}}_{n,p} \cup \widetilde{\mathcal{R}}_{n,p})$ in case $p = 2$.

3 Perfectly Balanced Boolean Functions

The set $\mathcal{PBF}_{n,2}$ is closed under the next operations on the set $\mathcal{F}_{n,2}$ [8]:

1. $\gamma_0: f(x_1 \dots x_n) \rightarrow f(x_1 \dots x_n) \oplus 1$;
2. $\gamma_1: f(x_1 \dots x_n) \rightarrow f((x_1 \oplus 1) \dots (x_n \oplus 1))$;
3. $\gamma_2: f(x_1 \dots x_n) \rightarrow f(x_n \dots x_1)$.

Let ξ and η be two random variables with support in A and B , respectively, with joint probability distribution $p_{\xi,\eta}(x,y)$, marginal distribution $p_\xi(x)$ and $p_\eta(y)$, respectively, and conditional distribution $p_{\eta|\xi}(y|x)$ and $p_{\xi|\eta}(x|y)$ for $x \in A$ and $y \in B$, respectively. The Shannon entropy of ξ is defined by

$$H(\xi) = - \sum_{x \in A} p_\xi(x) \log p_\xi(x).$$

All logarithms are to the base 2. We assume that $0 \log 0 = 0$. The joint Shannon entropy is defined as follows:

$$H(\xi, \eta) = - \sum_{x \in A} \sum_{y \in B} p_{\xi,\eta}(x,y) \log p_{\xi,\eta}(x,y).$$

The conditional Shannon entropy of η given ξ is

$$H(\eta|\xi) = - \sum_{x \in A} \sum_{y \in B} p_{\xi,\eta}(x,y) \log p_{\eta|\xi}(y|x).$$

Finally, the mutual information of ξ and η is

$$I(\eta, \xi) = I(\xi, \eta) = H(\xi) + H(\eta) - H(\xi, \eta) = H(\xi) - H(\xi|\eta) = H(\eta) - H(\eta|\xi).$$

Theorem 3 ([9]). Let $n, m \in I_1$ and $f \in \mathcal{F}_{n,2}$. Let $\{\xi_m\}_{m=1}^\infty$ be a sequence of random variables with probability distributions

$$\Pr[\xi_m = (a_1 \dots a_{m+n-1})] = 2^{-(m+n-1)}$$

for any $(a_1 \dots a_{m+n-1}) \in \mathcal{B}_{m+n-1,2}$ and any $m \in I_1$. A random variable $\eta_m = f_m(\xi_m)$ has a probability distribution

$$\Pr[\eta_m = (b_1 \dots b_m)] = 2^{-m}$$

for any $m \in I_1$ and any $(b_1 \dots b_m) \in \mathcal{B}_{m,2}$ iff $f \in \mathcal{PBF}_{n,2}$.

Theorem 4. Let $f \in \mathcal{F}_{n,2}$. Under conditions of Theorem 3, let $i_m(f) = m^{-1}I(\eta_m, \xi_m)$. Then

1. f is perfectly balanced iff $i_m(f) = 1$ for any $m \in I_1$;
2. for any $g \in \mathcal{F}_{n,2}$ one has $i_m(g) \leq i_m(f)$.

Proof. Let $f \in \mathcal{F}_{n,2}$. Under conditions of Theorem 3 we have

$$\begin{aligned} p_a &= \Pr[\xi_m = a] = 2^{-(m+n-1)} \quad \text{for any } a \in \mathcal{B}_{m+n-1,2}; \\ p_{b/a} &= \Pr[\eta_m = b \mid \xi_m = a] = \begin{cases} 1 & \text{if } a \in f_m^{-1}(b), \\ 0 & \text{if } a \notin f_m^{-1}(b); \end{cases} \\ p_{b,a} &= \Pr[\eta_m = b, \xi_m = a] = p_a p_{b/a} = \begin{cases} 2^{-(m+n-1)} & \text{if } a \in f_m^{-1}(b), \\ 0 & \text{if } a \notin f_m^{-1}(b); \end{cases} \\ q_b &= \Pr[\eta_m = b] = 2^{-(m+n-1)} \operatorname{card} f_m^{-1}(b) \end{aligned}$$

for any $b \in \mathcal{B}_{m,2}$. Then one has

$$\begin{aligned} i_m(f) &= m^{-1}I(\eta_m, \xi_m) = m^{-1}[H(\eta_m) - H(\eta_m|\xi_m)] \\ &= m^{-1} \sum_{b \in \mathcal{B}_{m,2}} \sum_{a \in f_m^{-1}(b)} p_{b,a} [\log p_{b/a} - \log q_b] \\ &= m^{-1} \sum_{b \in \mathcal{B}_{m,2}} \sum_{a \in f_m^{-1}(b)} 2^{-(m+n-1)} [(m+n-1) - \log(\operatorname{card} f_m^{-1}(b))] \\ &= m^{-1} \sum_{b \in \mathcal{B}_{m,2}} \operatorname{card} f_m^{-1}(b) 2^{-(m+n-1)} [(m+n-1) - \log(\operatorname{card} f_m^{-1}(b))] \\ &= m^{-1} \left[(m+n-1) \right. \\ &\quad \left. - 2^{-(m+n-1)} \sum_{b \in \mathcal{B}_{m,2}} \operatorname{card} f_m^{-1}(b) \log(\operatorname{card} f_m^{-1}(b)) \right]. \end{aligned} \tag{3}$$

1. Let $f \in \mathcal{PBF}_{n,2}$. Then $\operatorname{card} f_m^{-1}(b) = 2^{n-1}$ for any $b \in \mathcal{B}_{m,2}$ and any $m \in I_1$. From (3) one has $i_m(f) = 1$ for any $m \in I_1$.

To prove the opposed implication, let $f \in \mathcal{F}_{n,2}$ and $i_m(f) = 1$ for any $m \in I_1$. Then (3) implies

$$2^{-(m+n-1)} \sum_{b \in \mathcal{B}_{m,2}} \operatorname{card} f_m^{-1}(b) \log(\operatorname{card} f_m^{-1}(b)) = n-1.$$

Hence,

$$-\sum_{b \in \mathcal{B}_{m,2}} \frac{\text{card } f_m^{-1}(b)}{2^{m+n-1}} \log \left(\frac{\text{card } f_m^{-1}(b)}{2^{m+n-1}} \right) = m. \quad (4)$$

Moreover,

$$\sum_{b \in \mathcal{B}_{m,2}} \frac{\text{card } f_m^{-1}(b)}{2^{m+n-1}} = 1 \quad (5)$$

Thus (4) holds if and only if $\text{card } f_m^{-1}(b)/2^{m+n-1} = 2^{-m}$ for any $b \in \mathcal{B}_{m,2}$ [10]. For an arbitrary positive integer m one has $\text{card } f_m^{-1}(b) = 2^{n-1}$ for any $b \in \mathcal{B}_{m,2}$ and any $m \in I_1$. Thus $f \in \mathcal{PBF}_{n,2}$.

2. Let $g \in \mathcal{F}_{n,2}$ and $f \in \mathcal{PBF}_{n,2}$. Then

$$\begin{aligned} i_m(f) - i_m(g) &= m^{-1} [2^{-(m+n-1)} \sum_{b \in \mathcal{B}_{m,2}} \text{card } g_m^{-1}(b) \log(\text{card } g_m^{-1}(b)) - (n-1)] \\ &= m^{-1} \left[\sum_{b \in \mathcal{B}_{m,2}} \frac{\text{card } g_m^{-1}(b)}{2^{m+n-1}} \log(\text{card } g_m^{-1}(b)) \right. \\ &\quad \left. - \sum_{b \in \mathcal{B}_{m,2}} \frac{\text{card } g_m^{-1}(b)}{2^{m+n-1}} \log(2^{n-1}) \right] \\ &= m^{-1} \left[\sum_{b \in \mathcal{B}_{m,2}} \frac{\text{card } g_m^{-1}(b)}{2^{m+n-1}} \log \left(\frac{\text{card } g_m^{-1}(b)}{2^{m+n-1}} / \frac{2^{n-1}}{2^{m+n-1}} \right) \right]. \end{aligned} \quad (6)$$

The next inequality is well-known (see [10]). Let $\alpha_j, \beta_j, j = 1, \dots, s$ be nonnegative real numbers such that $\sum_{j=1}^s \alpha_j = \sum_{j=1}^s \beta_j = 1$. Then

$$\sum_{j=1}^s \alpha_j \log \alpha_j \geq \sum_{j=1}^s \alpha_j \log \beta_j \quad (7)$$

Combining (5), (6) for Boolean function g , and (7), one gets $i_m(f) - i_m(g) \geq 0$. \square

For a pair of positive integers m, k , a mapping $\Xi_{m,k}: \mathcal{F}_{m,2} \times \mathcal{F}_{k,2} \rightarrow \mathcal{F}_{m+k-1,2}$ is defined as follows:

$$\Xi_{m,k}(g, h) = g[h] = f \in \mathcal{F}_{m+k-1,2}, \quad g \in \mathcal{F}_{m,2}, h \in \mathcal{F}_{k,2}, \quad (8)$$

where

$$\begin{aligned} f(x_1 \dots x_{m+k-1}) &= g[h](x_1 \dots x_{m+k-1}) \\ &= g(h(x_1 \dots x_k)h(x_2 \dots x_{k+1}) \dots h(x_m \dots x_{m+k-1})). \end{aligned}$$

Theorem 5 ([11]). *Let $g \in \mathcal{F}_{m,2}$, $h \in \mathcal{F}_{k,2}$. The function $f = g[h] \in \mathcal{F}_{m+k-1,2}$ is perfectly balanced iff g and h are perfectly balanced.*

Theorem 5 allows one to generate functions in $\mathcal{PBF}_{n,2} \setminus (\mathcal{L}_{n,2} \cup \mathcal{R}_{n,2})$.

We consider integer sequences $\gamma = (\gamma_1 \dots \gamma_n)$ such that

$$\gamma_1 = 0, \quad \forall i \in \{1, \dots, n-1\} \quad \gamma_{i+1} > \gamma_i, \quad N = \gamma_n + 1.$$

These sequences correspond to tapping sequences of filter generators. For any $f \in \tilde{\mathcal{F}}_{n,2}$, let $f_\gamma(x_1 \dots x_N) \equiv f(x_{N-\gamma_n} x_{N-\gamma_{n-1}} \dots x_{N-\gamma_1})$.

Golić [3] stated the next theorem:

Theorem 6 ([3]). *For a nonlinear filter generator with the filter function f_γ and independent of the tapping sequence γ , the output sequence is purely random given that the input sequence is such if (and only if [not proven]) $f \in (\mathcal{L}_{N,2} \cup \mathcal{R}_{N,2})$.*

The “if” part of Theorem 6 was proven in [3]. The next statement can help to prove the “only if” part.

Theorem 7. For any $f \in \tilde{\mathcal{F}}_{n,2}$ which has no linear arguments there exists a sequence γ such that $f_\gamma \notin \mathcal{PBF}_{N,2}$.

Proof. Consider a sequence γ such that $\gamma_1 = 0$ and $\gamma_{i+1} > 2\gamma_i$ for each $i \in \{1, \dots, n-1\}$. The system of equations

$$\left\{ \begin{array}{l} f(x_{N-\gamma_n-1}x_{N-\gamma_{n-1}-1} \dots x_{N-\gamma_2-1}x_{N-\gamma_1-1}) \\ \quad = f(z_{N-\gamma_n-1}z_{N-\gamma_{n-1}-1} \dots z_{N-\gamma_2-1}z_{N-\gamma_1-1}), \\ f(x_{N-\gamma_n}x_{N-\gamma_{n-1}} \dots x_{N-\gamma_2}x_{N-\gamma_1}) \\ \quad = f(z_{N-\gamma_n}z_{N-\gamma_{n-1}} \dots z_{N-\gamma_2}z_{N-\gamma_1}), \\ f(x_{N-\gamma_n+1}x_{N-\gamma_{n-1}+1} \dots x_{N-\gamma_2+1}x_{N-\gamma_1+1}) \\ \quad = f(z_{N-\gamma_n+1}z_{N-\gamma_{n-1}+1} \dots z_{N-\gamma_2+1}z_{N-\gamma_1+1}), \\ \dots \\ f(x_{N-\gamma_1}x_{N-\gamma_1+\gamma_n-\gamma_{n-1}} \dots x_{N-2\gamma_1+\gamma_n}) \\ \quad = f(z_{N-\gamma_1}z_{N-\gamma_1+\gamma_n-\gamma_{n-1}} \dots z_{N-2\gamma_1+\gamma_n}), \\ f(x_{N-\gamma_1+1}x_{N-\gamma_1+\gamma_n-\gamma_{n-1}+1} \dots x_{N-2\gamma_1+\gamma_n+1}) \\ \quad = f(z_{N-\gamma_1+1}z_{N-\gamma_1+\gamma_n-\gamma_{n-1}+1} \dots z_{N-2\gamma_1+\gamma_n+1}), \\ x_{N-\gamma_n-1} = z_{N-\gamma_n-1}, \\ x_{N-\gamma_n} = z_{N-\gamma_n}, \\ \dots \\ x_{N-\gamma_1-1} = z_{N-\gamma_1-1}, \\ x_{N-\gamma_1} = 0, z_{N-\gamma_1} = 1, \\ x_{N-\gamma_1+1} = z_{N-\gamma_1+1}, \\ x_{N-\gamma_1+2} = z_{N-\gamma_1+2}, \\ \dots \\ x_{N-2\gamma_1+\gamma_n+1} = z_{N-2\gamma_1+\gamma_n+1}. \end{array} \right. \quad (9)$$

is solvable. Indeed, all equations that do not depend on $x_{N-\gamma_1}$ and $z_{N-\gamma_1}$ evidently hold. Therefore one has to solve the system

$$\left\{ \begin{array}{l} f(x_{N-\gamma_n+\gamma_1}x_{N-\gamma_{n-1}+\gamma_1} \dots x_{N-\gamma_2+\gamma_1}0) \\ \quad = f(x_{N-\gamma_n+\gamma_1}x_{N-\gamma_{n-1}+\gamma_1} \dots x_{N-\gamma_2+\gamma_1}1), \\ f(x_{N-\gamma_n+\gamma_2}x_{N-\gamma_{n-1}+\gamma_2} \dots x_{N-\gamma_3+\gamma_2}0x_{N-\gamma_1+\gamma_2}) \\ \quad = f(x_{N-\gamma_n+\gamma_2}x_{N-\gamma_{n-1}+\gamma_2} \dots x_{N-\gamma_3+\gamma_2}1x_{N-\gamma_1+\gamma_2}), \\ \dots \\ f(0x_{N-\gamma_{n-1}+\gamma_n} \dots x_{N-\gamma_1+\gamma_n}) \\ \quad = f(1x_{N-\gamma_{n-1}+\gamma_n} \dots x_{N-\gamma_1+\gamma_n}). \end{array} \right. \quad (10)$$

Every equation in this system is solvable because f does not have linear arguments. Next, we show that j th and l th equations are independent. This amounts to show:

$$\begin{aligned} \forall i, j, k, l \in \{1, \dots, n\} \quad & i \neq j, k \neq l, i \neq k, j \neq l \\ \implies N - \gamma_i + \gamma_j & \neq N - \gamma_k + \gamma_l. \end{aligned} \quad (11)$$

It can be easily verified that it suffices to show for each $i \in \{1, \dots, n-1\}$, for all $j \leq i, k \leq i, l$ that $\gamma_{i+1} - \gamma_j > \gamma_k - \gamma_l$ holds. We have chosen a sequence γ such that for each $i \in \{1, \dots, n-1\}$ inequality $\gamma_{i+1} > 2\gamma_i$ holds. Therefore $\gamma_{i+1} - \gamma_j > \gamma_i + (\gamma_i - \gamma_j) \geq \gamma_i \geq \gamma_i - \gamma_l \geq \gamma_k - \gamma_l$.

System (11) has n independent equations each of which is solvable. Thus systems (10) and (9) are solvable. By Theorem 2 one has $f_\gamma \notin \mathcal{PBF}_{N,2}$. \square

To prove the “only if” part (the Golić conjecture), one has to handle linear arguments of f and solve “an underlying combinatorial problem” mentioned by Golić.

4 Barriers of Boolean Functions

A barrier of a Boolean function is closely related to the property of perfect balancedness.

Definition 4 ([12]). A Boolean function $f \in \mathcal{F}_{n,2}$ has a right barrier of length b if the system

$$\begin{cases} f(y_1y_2 \dots y_n) = f(z_1z_2 \dots z_n), \\ f(y_2y_3 \dots y_{n+1}) = f(z_2z_3 \dots z_{n+1}), \\ \dots \\ f(y_{b-1}y_b \dots y_{b+n-2}) = f(z_{b-1}z_b \dots z_{b+n-2}), \\ y_1 = z_1 = x_1, \dots, y_{n-1} = z_{n-1} = x_{n-1}, y_n = 0, z_n = 1. \end{cases} \quad (12)$$

is solvable, but the system

$$\begin{cases} f(y_1y_2 \dots y_n) = f(z_1z_2 \dots z_n), \\ f(y_2y_3 \dots y_{n+1}) = f(z_2z_3 \dots z_{n+1}), \\ \dots \\ f(y_{b-1}y_b \dots y_{b+n-2}) = f(z_{b-1}z_b \dots z_{b+n-2}), \\ f(y_b y_{b+1} \dots y_{b+n-1}) = f(z_b z_{b+1} \dots z_{b+n-1}), \\ y_1 = z_1 = x_1, \dots, y_{n-1} = z_{n-1} = x_{n-1}, y_n = 0, z_n = 1. \end{cases} \quad (13)$$

is not solvable.

A Boolean function $f \in \mathcal{F}_{n,2}$ has a left barrier of length b if $f^{\gamma_2}(x_1 \dots x_n) \equiv f(x_n \dots x_1)$ has a right barrier of length b .

A Boolean function $f \in \mathcal{F}_{n,2}$ has a barrier if it has a right barrier, a left barrier, or both. The length of a Boolean function barrier is defined to be the minimum of the length of its right and left barrier.

Theorem 8. *If a Boolean function has a barrier, then it is perfectly balanced. An opposed implication does not hold.*

Proof. If a function has left or right barrier, then it is easily verified using the definition of barrier and Theorem 6 that the function is perfectly balanced.

Now we give an example of a perfectly balanced function without barrier. Let $f(x_1x_2x_3x_4x_5) = x_1 \oplus x_2 \oplus x_5 \oplus x_1x_2 \oplus x_1x_3 \oplus x_1x_4 \oplus x_2x_3 \oplus x_2x_4 \oplus x_3x_4 \oplus x_4x_5 \oplus x_1x_3x_4 \oplus x_2x_3x_4 = g^{\gamma_2}[g]$, where $g(x_1x_2x_3) = x_1 \oplus x_2 \oplus x_3 \oplus x_2x_3$, $g^{\gamma_2}(x_1x_2x_3) \equiv g(x_3x_2x_1)$. The function g has a left barrier of length $k = 1$, hence $g, g^{\gamma_2} \in \mathcal{PBF}_{3,2}$. Using Theorem 5 one concludes that $f \in \mathcal{PBF}_5$. To show that f does not have a barrier, we use next two pairs of binary sequences:

$$\left[\begin{array}{c} 0001(010100)(010100)(010100)\dots \\ 0001(100010)(100010)(100010)\dots \end{array} \right], \quad (14)$$

$$\left[\begin{array}{c} \dots(00111110)(00111110)(00111110)000100 \\ \dots(11100011)(11100011)(11100011)010100 \end{array} \right]. \quad (15)$$

For any $k > 0$ one can make pair (14) to have length more than $k + 3$ by repeating periodical parts of the sequences sufficiently many times and then truncating the sequences obtained to exact length. Then one substitutes the sequences in (12) and (13). Eqs. (12) and (13) hold, hence no k can satisfy Definition 4 and f has no right barrier. Similarly one can show that f has no left barrier. \square

The next theorem can be used for classification of Boolean functions with a right (left) barrier.

Theorem 9 ([12]). *Let $f \in \mathcal{F}_{n,2}$ be a function with a right (left) barrier of length $b < n$ and $g \in \mathcal{F}_{n-b,2}$ be arbitrary. Then $h(x_1 \dots x_n) = f(x_1 \dots x_n) \oplus g(x_1 \dots x_{n-b})$ (resp., $h(x_1 \dots x_n) = f(x_1 \dots x_n) \oplus g(x_{b+1} \dots x_n)$) has a right (resp., left) barrier of length b .*

The proof of this theorem is easy and follows from (12) and (13).

Barrier length is a value that characterizes, in a particular way, invertibility of mapping f_m . In the proof of Theorem 8 we show that composition of functions $g^{\gamma_2}[g]$ is a function without barrier. One can show that g does not have a right barrier and g^{γ_2} does not have a left barrier. The next theorem answers the question of existence of right and left barrier of a composition function $g[h]$. We assume that length of a right (left) barrier is $+\infty$ if a function does not have right (left) barrier.

Theorem 10. *Let $h \in \mathcal{F}_{k,2}$, $g \in \mathcal{F}_{m,2}$, $f = g[h]$ and length of right (left) barrier of functions h, g, f are b, c, d , respectively. Then $\max\{b, c\} \leq d \leq b + c - 1$*

Now we outline ideas behind the proof. Inequalities $d \geq b$ and $d \leq b + c - 1$ (both in the case of finite b, c, d and in the case when some of b, c, d is infinite) are proved by analysing systems produced from (12), (13) by substituting values of h taken on the corresponding vectors into g . In this way we get inequalities $d \geq b$ and $b + c - 1 \geq d$. To prove inequality $d \geq c$ we use Lemma 1 whose proof is based on some special properties of perfectly balanced Boolean functions.

Lemma 1. *Let $f \in \mathcal{PBF}_{n,2}$. Then for any positive integer u and for any binary sequences $z_0, z_1 \in \mathcal{B}_{u,2}$ there exist a positive integer r and binary sequences $x, y \in \mathcal{B}_{r+n-1,2}, z \in \mathcal{B}_{r-u,2}$ such that*

$$x_1 = y_1, \dots, x_{n-1} = y_{n-1}, f^*(x) = (z|z_0), f^*(y) = (z|z_1).$$

Finally, we show an example of method of a class of perfectly balanced functions without barriers.

Lemma 2. Functions

$$\begin{aligned} f = & x_1 \oplus x_{m_1^\circ} x_{m_1^\circ+1} \dots x_{m_1} h_1(x_{m_1+1} x_{m_1+2} \dots x_{m_k}) \\ & \oplus x_{m_2^\circ} x_{m_2^\circ+1} \dots x_{m_2} h_2(x_{m_2+1} x_{m_2+2} \dots x_{m_k}) \oplus \dots \\ & \oplus x_{m_k^\circ} x_{m_k^\circ+1} \dots x_{m_k}, \quad 1 < m_1^\circ < m_1 < m_2^\circ < m_2 < \dots < m_k^\circ < m_k, \end{aligned} \quad (16)$$

where h_i are product terms and k is odd are perfectly balanced functions without right barrier.

Proof. All such functions are in $\mathcal{L}_{n,2}$ and therefore in $\mathcal{PBF}_{n,2}$. To prove that such a function has no right barrier consider for any positive integer b the following pair of binary sequences:

$$\begin{array}{ccccccc} 00\dots0 & 101\dots0 & 000\dots0 & 010\dots0 & & \dots & \\ \underbrace{00\dots0}_{m_k-1} & \underbrace{010\dots1}_{m_k^\circ-m_{k-1}^\circ} & \underbrace{000\dots0}_{m_{k-1}^\circ-m_{k-2}^\circ} & \underbrace{101\dots1}_{m_{k-2}^\circ-m_{k-3}^\circ} & & & \\ \dots & \dots & \dots & \dots & & & \\ & \underbrace{101\dots1}_{m_3^\circ-m_2^\circ} & \underbrace{000\dots0}_{m_2^\circ-m_1^\circ} & \underbrace{010\dots1}_{m_1^\circ-1} & \underbrace{11\dots1}_{b-m_k-m_k^\circ+2} & & . \end{array}$$

Substituting this pair of sequences into systems of Definition 4 one can see f has no right barrier of length at most b . Since this holds for any b , f doesn't have right barrier. \square

Now using Lemma 2 and Theorem 10 we can describe a large class of perfectly balanced functions without barriers.

Theorem 11. *Let f_1, f_2 be of the form (16) or be produced by applying to such functions operations $\gamma_1: f(x_1 \dots x_n) \rightarrow f((x_1 \oplus 1) \dots (x_n \oplus 1))$. Then $f_1[f_2^{\gamma_2}]$ and $f_1^{\gamma_2}[f_2]$ are perfectly balanced functions without barriers.*

References

- [1] F.P. Preparata, Convolutional transformations of binary sequences: Boolean functions and their resynchronizing properties, *IEEE Trans. Electron. Comput.* **15**(6) (1966), 898–909.
- [2] D. A. Huffman, Canonical forms for information-lossless finite-state logical machines, *IRE Trans. Circuit Theory* **5** (1959), spec. suppl., 41–59.
- [3] J.Dj. Golić, On the security of nonlinear filter generators, *Proc. of Fast Software Encryption 1996, Lecture Notes in Computer Science* **1039** (1996), Springer-Verlag, 173–188.
- [4] A. Gouget and H. Sibert, Revisiting correlation-immunity in filter generators, *Proc. of SAC 2007, Lecture Notes in Computer Science* **4876** (2007), Springer-Verlag, 378–395.
- [5] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, Cambridge, 1995.
- [6] B.R. Kitchens, *Symbolic Dynamics, One-sided, Two-sided, Countable State Markov Shifts*, Springer-Verlag, Berlin, 1998.

- [7] G.A. Hedlund, Endomorphisms and automorphisms of the shift dynamical system, *Math. Systems Th.* **3** (1969) 320–375.
 - [8] S.N. Sumarokov, Functions of defect zero and invertability of some class of finite-memory encoders, *Obozrenie prom. i prikl. mat.* **1**(1) (1994) 33–55 (in Russian).
 - [9] O.A. Logachev, A.A. Salnikov, and V.V. Yashchenko, *Boolean Functions in Coding Theory and Cryptology*, MCCME, Moscow, 2004 (in Russian).
 - [10] R. Ash, *Information Theory*, John Wiley and Sons Inc., New York–London–Sydney, 1967.
 - [11] O.A. Logachev, On perfectly balanced Boolean functions, Cryptology ePrint Archive, Report 2007/022, <http://eprint.iacr.org/>.
 - [12] O.A. Logachev, S.V. Smyshlyaev, and V.V. Yashchenko, New approach to perfectly balanced functions design, *Discrete Mathematics* **21**(2) (2009), 51–74 (in Russian).

5 Appendix

Sets of $\widetilde{\mathcal{PBF}}_{4,2} \setminus (\mathcal{L}_{4,2} \cup \mathcal{R}_{4,2})$ and $\widetilde{\mathcal{PBF}}_{5,2} \setminus (\mathcal{L}_{5,2} \cup \mathcal{R}_{5,2})$ are divided into classes of equivalence by operations $\gamma_0 - \gamma_2$. These classes are subdivided into subsets by barrier length, which are later classified by operations noted in Theorem 9. Then we choose one element from each class such that length of barrier is equal to length of right barrier.

$\widetilde{\mathcal{PBF}}_{4,2} \setminus (\mathcal{L}_{4,2} \cup \mathcal{R}_{4,2})$ with barrier of length $k = 3$:

$$\begin{aligned} 1 \ f(x_1, x_2, x_3, x_4) &= x_3 \oplus x_4x_2 \oplus x_4x_2x_1 \\ 2 \ f(x_1, x_2, x_3, x_4) &= x_3 \oplus x_4x_1 \oplus x_4x_2x_1 \\ 3 \ f(x_1, x_2, x_3, x_4) &= x_3 \oplus x_2 \oplus x_4x_2 \oplus x_3x_2 \oplus x_4x_2x_1 \oplus x_3x_2x_1 \\ 4 \ f(x_1, x_2, x_3, x_4) &= x_3 \oplus x_4x_2 \oplus x_3x_2 \oplus x_2x_1 \oplus x_4x_2x_1 \oplus x_3x_2x_1 \end{aligned}$$

Functions in remaining part of Appendix are presented with ANF coefficients vector.

$\widetilde{\mathcal{PBF}}_{5,2} \setminus (\mathcal{L}_{5,2} \cup \mathcal{R}_{5,2})$ with barrier of length $k = 3$:

$\widetilde{\mathcal{PBF}}_{5,2} \setminus (\mathcal{L}_{5,2} \cup \mathcal{R}_{5,2})$ with barrier of length $k = 4$:

$\widetilde{\mathcal{PBF}}_{5,2} \setminus (\mathcal{L}_{5,2} \cup \mathcal{R}_{5,2})$ with barrier of length $k = 5$:

$\widetilde{\mathcal{PBF}}_{5,2} \setminus (\mathcal{L}_{5,2} \cup \mathcal{R}_{5,2})$ without barrier:

$f_1 : 00001000100100100000000010000000; f_2 : 0101100001010000000000010000000; f_3 : 01011000101000000000000010000000;$
 $f_4 : 01011000000000000000001000101000000; f_5 : 0001101000001100000000010000000; f_6 : 01011000000001100000000010000000;$
 $f_7 : 0101001001000100100000010000000; f_8 : 01011100100001000000000010000000; f_9 : 01011100010010000000000010000000;$
 $f_{10} : 01111000001000000000000011010000; f_{11} : 0001011010000000000000000011110; f_{12} : 00111010000011100000000010000000;$
 $f_{13} : 0001101010000010100000101000000; f_{14} : 0011100010001000000010101000000; f_{15} : 0010010111011000100010000000;$
 $f_{16} : 0010101100110010100000101000000; f_{17} : 0001100110110010100000101000000; f_{18} : 01011110001010000000000010101000;$
 $f_{19} : 01011000010100000000000011010110; f_{20} : 01111010100000000010001011010000; f_{21} : 01010010010001001000000010011110;$
 $f_{22} : 01011100110101000000000011010000; f_{23} : 010100100010101010010010000100; f_{24} : 00001000001110000011001010110010;$
 $f_{25} : 00101011001100011011000110110000; f_{26} : 00101000000011011101110111010000; f_{27} : 0101111001111110000000000011111110.$