Description of Binary Sequences Based on the Interval Linear Complexity Profile

Vladimir B. Balakirsky
Electr. Eng. Dept., TUE
P.O.Box 513, 5600 MB Eindhoven,
The Netherlands
v.b.balakirsky@ele.tue.nl

Abstract — **We consider the problem of partial approximation of binary sequences by the outputs of linear feedback shift registers. A generalization of the linear complexity profiles of binary sequences leads to a sequence that is regarded as the profile of interval linear complexity. Some properties of this sequence and possible applications are studied.**

I. Introduction

There is a class of problems, relevant to decoding, cryptology, and data compression, where we are interested in a proper partitioning of a given binary sequence into different fragments. The goal of this partitioning is to represent some fragments as the output sequences of different linear feedback shift registers (LFSRs), and we speak about partial linear approximations of the initial sequence in this sense. The problems of this class can be effectively attacked using a generalized version of the linear complexity profile of the sequence [1], [2], which was introduced in [3] as the interval linear complexity profile.

In the foregoing considerations we assume that the integer $m \geq 1$ is fixed and omit the dependence of all parameters on m. We also denote the given binary sequence by $\mathbf{u} = (u_1, \ldots, u_N)$ and assume that $u_i = 0$ for all $i \leq 0$. Let $\mathbf{u}_j \stackrel{\triangle}{=} (u_j, \ldots, u_{j-m+1})$ for all $j = 1, \ldots, N$. Furthermore, let $\mathbf{0}$ be the vector consisting of m zeroes.

II. The m–Interval Linear Complexity

Definition. *Given integer $m \geq 1$ and a binary sequence $\mathbf{u} = (u_1, \ldots, u_N)$ with $u_1 = 1$, the value of the m–interval linear complexity at position $j \in \{m+1, \ldots, N\}$, denoted by μ_j, is defined by the following rules :*

- *if $\mathbf{u}_j = \mathbf{0}$, then $\mu_j = 0$;*
- *if $\mathbf{u}_j \neq \mathbf{0}$, then μ_j is the smallest integer n such that the vector $\mathbf{u}_j$ can be expressed as a linear combination of the vectors $\mathbf{u}_{j-1}, \ldots, \mathbf{u}_{j-n}$, i.e., there exists a polynomial*

$$g_j(z) = 1 + g_{j,1}z + \ldots + g_{j,n-1}z^{n-1} + z^n$$

with $g_{j,1}, \ldots, g_{j,n-1} \in \{0,1\}$ and

$$\mathbf{u}_j = g_{j,1}\mathbf{u}_{j-1} \oplus \ldots \oplus g_{j,n-1}\mathbf{u}_{j-n+1} \oplus \mathbf{u}_{j-n}.$$

The sequence $(\mu_{m+1}, \ldots, \mu_N)$ is referred to as the m–interval linear complexity profile (m–LCP) of $\mathbf{u}$.

If $\mathbf{u}_j \neq \mathbf{0}$, then we can construct the pair $(\mu_j, g_j(z))$ using a variant of the Berlekamp–Massey algorithm. Some properties of the m–LCP of a binary sequence are summarized below and illustrated in Figure 1.

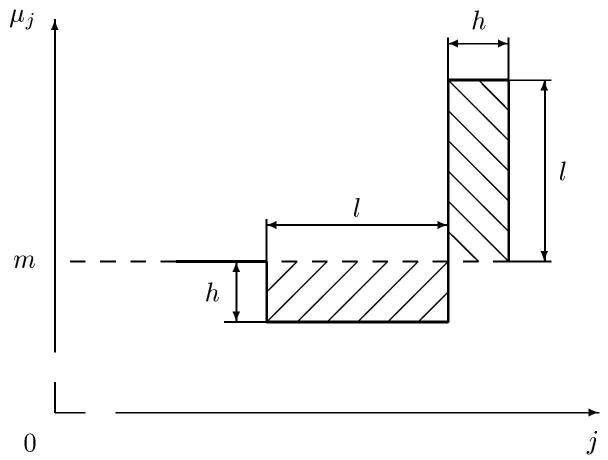

Figure 1: Fragment of the m–LCP of any binary sequence.

Theorem 1. *For all $j \geq m+1$ and $n \in \{0, \ldots, m-1\}$, conditions*

$$\begin{cases} \mu_{j-1} \neq n \\ \mu_j = \ldots = \mu_{j+l-1} = n \\ \mu_{j+l} \neq n \end{cases}$$

imply

$$\begin{cases} \mu_{j-1} \geq m \\ g_j(z) = \ldots = g_{j+l-1}(z) \\ \mu_{j+l} = \ldots = \mu_{j+l+h-1} = m + l > m \\ \mu_{j+l+h} \leq m \end{cases}$$

where $h = m - n$.

Theorem 2. *If $j > 2m$ and $u_2, \ldots, u_j$ are i.i.d. random variables chosen from $\{0,1\}$ with probability $1/2$, then*

$$\Pr\{\, \mu_j = n \,\} = \frac{K_n}{2^{m+1-n}} \cdot 2^{-2n+1} = \frac{2^n - 2^{-n}}{3 \cdot 2^m}$$

for all $n = 1, \ldots, m$, where K_n is the number of choices for $a_0, \ldots, a_{n-1}, f_1, \ldots, f_{n-1} \in \{0,1\}$ such that the polynomials $a_0 + a_1 z + \ldots + a_{n-1}z^{n-1}$ and $1 + f_1 z + \ldots + f_{n-1}z^{n-1} + z^n$ are relatively prime. Furthermore, $\Pr\{\, \mu_j = 0 \,\} = 2^{-m}$ and

$$\sum_{n=0}^{m} \sum_{l \geq 1} (m - n + l - 1) \Pr\{\, \mu_j = n,\, l_j = l \,\} = \frac{4}{3} - \frac{1}{3} \cdot 2^{-2m}$$

where $\Pr\{\, \mu_j = n, l_j = l \,\}$ is the probability assigned to the sequence with $g_j(z) = \ldots = g_{j+l-1}(z) \neq g_{j+l}(z)$ and $\deg g_j(z) = n$.

III. Identification of the Positions Where the m–Interval Linear Complexity is Less Than m

Lemma 3. *Given an $i \geq m+1$, let $\mathbf{C}_i$ and $\mathbf{X}_i$ be binary $m \times m$ matrices containing rows $\mathbf{c}_{i,k} = (c_{i,k}^{(1)}, \ldots, c_{i,k}^{(m)})$ and $\mathbf{x}_{i,k} = (x_{i,k}^{(1)}, \ldots, x_{i,k}^{(m)})$, respectively, $k = 1, \ldots, m$. Suppose that the given vector $\mathbf{u}_j \in \{0,1\}^m$ can be represented as*

$$\mathbf{u}_j = f_1 \mathbf{x}_{i,1} \oplus f_2 \mathbf{x}_{i,2} \oplus \ldots \oplus f_m \mathbf{x}_{i,m} \tag{1}$$

where $f_2, \ldots, f_m \in \{0,1\}$ and $f_1 = 1$. Define the rows of the matrices $\mathbf{C}_j$ and $\mathbf{X}_j$ by

$$\mathbf{c}_{j,k} \triangleq \mathrm{L}[\,\mathbf{c}_{i,k}\,] \oplus \begin{cases} \mathbf{0}, & \text{if } c_{i,k}^{(1)} = 0 \\ (f_2, \ldots, f_{m-1}, 1), & \text{if } c_{i,k}^{(1)} = 1 \end{cases}$$

$$\mathbf{x}_{j,k} \triangleq \begin{cases} \mathbf{x}_{i,k+1}, & \text{if } k \in \{1, \ldots, m-1\} \\ \mathbf{u}_j, & \text{if } k = m \end{cases}$$

where $\mathrm{L}[\,\mathbf{c}_{i,k}\,] = (c_{i,k}^{(2)}, \ldots, c_{i,k}^{(m)}, 0)$ and $k = 1, \ldots, m$. Then

$$\mathbf{C}_j \mathbf{X}_j = \mathbf{C}_i \mathbf{X}_i.$$

Lemma 4. *If the rank of the matrix $\mathbf{X}_i$ is equal to m and the matrix $\mathbf{C}_i$ is defined by the equation*

$$\mathbf{C}_i \mathbf{X}_i = \mathbf{I} \tag{2}$$

where $\mathbf{I}$ denotes the $m \times m$ identity matrix, then the coefficients $f_1, \ldots, f_m$, satisfying (1) given $\mathbf{u}_j$, can be expressed as

$$(f_1, \ldots, f_m) = \mathbf{u}_j \mathbf{C}_i. \tag{3}$$

The transformation $(\mathbf{C}_i, \mathbf{X}_i) \longrightarrow (\mathbf{C}_j, \mathbf{X}_j)$ is illustrated in Table 1.

Suppose that $u_1 = 1$. Then $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are linearly independent vectors, and we uniquely define the matrix $\mathbf{C}_m$ by (2), where $\mathbf{x}_{m,k} = \mathbf{u}_k$, $k = 1, \ldots, m$. Let us also set $r_m = 0$ and run the following algorithm for all $j = m+1, \ldots, N$.

1*. Find the largest $i < j$ such that $r_j \neq *$ and compute $(f_1, \ldots, f_m)$ by (3).

2*. If $f_1 = 0$, then set $r_j = *$.

3*. If $f_1 = 1$, then set $r_j = i$.

Theorem 5. *Let $(\mu_{m+1}, \ldots, \mu_N)$ be the m–LCP of $\mathbf{u}$. If $r_{m+1}, \ldots, r_N$ are constructed by the $1^* - 3^*$ algorithm, then $r_j \neq *$ implies $\mu_j = m$ and $r_j = *$ implies $\mu_j < m$, for all $j = m+1, \ldots, N$.*

IV. Identification of the Positions Where the m–Fragments of the Sequence Generated by the Known LFSR are Corrupted

Suppose that the binary sequence $\mathbf{v} = (v_1, \ldots, v_{N'})$ is generated by an LFSR in such a way that

$$v_i = g_1 v_{i-1} \oplus \ldots \oplus g_{m-1} v_{i-m+1} \oplus v_{i-m} \tag{4}$$

for all $i = m+1, \ldots, N'$, where $g_1, \ldots, g_{m-1} \in \{0,1\}$. Then the sequence $\mathbf{v}$ is defined by $2m-1$ bits $v_1, \ldots, v_m, g_1, \ldots, g_{m-1}$, and specification of these bits is the shortest possible description of $\mathbf{v}$ in most of the cases. Suppose that the sequence $\mathbf{v}$ is distorted in such a way that some components are deleted, some 0's and 1's are inserted, and some components are inverted. We assume that the distorting process is implemented in accordance with the following algorithm that will be referred to as *the DIS–procedure.*

Let $(\alpha_1, \ldots, \alpha_{N'})$ be a given vector whose components belong to the set $\{+, S, D, I_0, I_1\}$.

Set $i = j = 1$ and until $i \leq N'$ do

Table 1: Example of the transformation of the basis for $i = 10, j = 16, \mathbf{u}_j = (10110)$, $(f_1, \ldots, f_5) = \mathbf{u}_j \mathbf{C}_i = (11001)$.

$\mathbf{C}_i$	$\mathbf{X}_i$	$\mathbf{e}_i$	$\mathbf{C}_j$	$\mathbf{X}_j$	$\mathbf{e}_j$
10011	10001	4	10101	11000	9
11011	11000	3	00101	11100	8
01100	11100	2	11000	11110	7
00110	11110	1	01100	11111	6
00011	11111	0	00110	10110	0

- if $\alpha_i = +$, then set $u_j = v_i$, increase i and j by 1;
- if $\alpha_i = S$, then set $u_j = v_i \oplus 1$, increase i and j by 1;
- if $\alpha_i = D$, then increase i by 1;
- if $\alpha_i = I_b$, where $b \in \{0,1\}$, then set $u_j = v_i$, $u_{j+1} = b$, increase i by 1 and j by 2.

Lemma 6. *Suppose that the components of the sequence $\mathbf{v}$ satisfy (4) and that the coefficients $g_1, \ldots, g_{m-1}$ are known. Let the sequence $\mathbf{u}$ be constructed from $\mathbf{v}$ by the DIS–procedure. Given $j \in \{m+1, \ldots, N\}$, let there exist $n \in \{1, \ldots, m\}$, $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \{\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}\}$, and integers $e_1 > \ldots > e_n > 0$ such that*

- *$\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly independent vectors;*
- *$\mathbf{u}_j = \mathbf{x}_1 \oplus \ldots \oplus \mathbf{x}_n$.*

If $\mathbf{u}_j = \mathbf{v}_{j'}$ and $\mathbf{x}_1 = \mathbf{v}_{j'-e_1}, \ldots, \mathbf{x}_n = \mathbf{v}_{j'-e_n}$ for some $j' \in \{1, \ldots, N'\}$, then

$$g(z) \;\Big|\; 1 + z^{e_1} + \ldots + z^{e_n}$$

where $g(z) = 1 + g_1 z + \ldots + g_{m-1} z^{m-1} + z^m$.

Let us consider data in Table 1 corresponding to the sequence $\mathbf{v}$ generated by the polynomial $g(z) = 1 + z^2 + z^5$. In this case, $\mathbf{u}_j = \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \mathbf{x}_3$, where $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = (\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \mathbf{x}_{i,5})$. Furthermore, $j' = j$, $(e_1, e_2, e_3) = (16 - 10 + 4, 16 - 10 + 3, 16 - 10 + 0)$ and $g(z) \,|\, 1 + z^6 + z^9 + z^{10}$. We can meet these data when $\alpha_{11} = S$ and all other components of the vector $(\alpha_1, \ldots, \alpha_{16})$ are equal to $+$. If we do not allow deletions and insertions, then j' is always equal to j, and the divisibility constraint is not satisfied for $j = 11, \ldots, 15$.

In the complete version of the paper, we discuss different approaches that are based on the use of Lemma 6 in a more general case, when the divisibility constraint is considered as a necessary condition for $u_j = v_{j'}$ under the assumption that m linearly independent m–fragments of the sequence $\mathbf{v}$ are already found.

References

[1] R. A. Rueppel, "Stream ciphers," in : *Contemporary Cryptology,* G. J. Simmons, Ed., New York : IEEE Press, pp. 65–134, 1991.

[2] J. L. Massey, "Shift register synthesis and BCH decoding," *IEEE Trans. Inform. Theory,* vol. 15, pp. 122–127, Jan. 1969.

[3] V. B. Balakirsky, "On interval linear complexity of binary sequences," SFB–343 *"Diskrete Strukturen in der Mathematik",* Preprint 97–014, Bielefeld, 1997.
Proc. 1995 IEEE International Symposium on Information Theory, Whistler, British Columbia, Canada, p. 490, 1995.