On Interval Linear Complexity of Binary Sequences

V.B.Balakirsky
Data Security Association "Confident", 193060 St.-Petersburg, Russia

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 are examined.**

I. Introduction

A widely used criterion of the linear complexity of binary sequences was introduced by R.Rueppel [1]. In accordance with this criterion, we construct an integer-valued sequence, called the 'linear complexity profile' (LCP), whose j-th component, L_j, is the shortest length of a linear feedback shift register (LFSR) generating the first j bits of our binary sequence. The component L_j can be found using Berlekamp-Massey algorithm [2], and pseudo-random sequences possess an LCP with $L_j \approx j/2$ for all j. However, there are many examples when the deviations of the LCP from the sequence $j/2$, $j = 1, 2, ...$ do not characterize the 'randomness' of binary sequences. For example, let us suppose that we are given a sequence $\mathbf{u}^* 01^\infty$, where $\mathbf{u}^*$ is a sequence of length n having a 'good' LCP. Then $\mathbf{u}^*$ is generated by an LFSR of length $\approx n/2$. Nevertheless, the final result of Berlekamp-Massey algorithm, applied to the whole sequence, is the LFSR of length $n + 2$, and the LCP is as good as before up to the $\approx 2n$-th component. It is easy to see that if we construct the LCP of the sequence starting at the $n + 1$-st bit then the conclusion would be different. Thus, to extend the Rueppel's approach we need to construct the LCPs for all subsequences of the input sequence and select the worst one, whose deviations from the line $j/2$ should be used as a measure of complexity. Such a procedure seems to be rather complicated.

In this paper, we introduce a new measure of complexity, called an interval linear complexity. For all $L < m$, where m is a given positive integer, we find all the fragments of the binary sequence that have length $L + m$ and can be generated by an LFSR of length L. The number of such fragments and their lengths contain information on LCPs for all starting positions, and the results of analysis can be useful for different methods based on linear approximations.

II. Some properties of the m-interval linear complexity

Let $\mathbf{u} = u_1, u_2, ...$ be a binary sequence. We assume that $u_1 = 1$ and $u_i = 0$ for $i = 0, -1, ...$ For all $k > 0$, we set $\mathbf{u}_j^{(k)} = (u_{j-k+1}, ..., u_j)$ and write $\mathbf{u}_j^{(k)} \prec \mathcal{F}(L)$ iff there exists a binary vector $(a_1, ..., a_L)$ such that $u_t = a_1 \cdot u_{t-1} + ... + a_L \cdot u_{t-L}$ for all $t \in \{j - k + 1, ..., j\}$. Furthermore, we write $\mathbf{u}_j^{(k)} \not\prec \mathcal{F}(L')$ iff $u_t \neq b_1 \cdot u_{t-1} + ... + b_{L'} \cdot u_{t-L'}$ for at least one $t \in \{j - k + 1, ..., j\}$, where $(b_1, ..., b_{L'})$ is any binary vector.

Let us fix $j > m$ and define $L_j^{(m)}$ as the shortest length of an LFSR, generating the fragment $\mathbf{u}_j^{(m)}$ provided that the subsequence $\mathbf{u}_{j-m}^{(L)}$, where $L = L_j^{(m)}$, forms the initial content of the shift register, i.e., (a) $\mathbf{u}_j^{(m)} \prec \mathcal{F}(L_j^{(m)})$; (b) if $L' < L_j^{(m)}$, then $\mathbf{u}_j^{(m)} \not\prec \mathcal{F}(L')$. Using conventional notations [2], we claim that $L_j^{(m)} = L$ iff L is the shortest length of an LFSR generating the subsequence $u_{j-m-L+1}, ..., u_j$, and all the subsequences $u_{j-m-L'+1}, ..., u_j$, where $L' < L$, cannot be generated by an LFSR of length L'. The parameter $L_j^{(m)}$ will be referred to as the *m-Interval Linear Complexity* (m-ILC) of $\mathbf{u}$ at position j, and the sequence $L_{m+1}^{(m)}, L_{m+2}^{(m)}, ...$ will be regarded as the profile of the m-ILC of $\mathbf{u}$. Some properties of the m-ILC are detailed below.

Theorem.

1. Let L_{ij} be the shortest length of an LFSR generating $u_i, ..., u_j$. Then
$$L_j^{(m)} = \min_{i:\ i+L_{ij} \leq j-m+1} L_{ij}.$$
2. If $L_j^{(m)} = L \leq m$, then there is exactly one LFSR of length L generating $u_{j-m-L+1}, ..., u_j$.
3. If
$$\begin{cases} L_{j-1}^{(m)} \neq L_j^{(m)}, \\ L_j^{(m)} = ... = L_{j+l-1}^{(m)} = L < m, \\ L_{j+l}^{(m)} \neq L_{j+l-1}^{(m)} \end{cases}$$
then
$$\begin{cases} L_{j-1}^{(m)} \geq m, \\ L_{j+l+\Delta l}^{(m)} = m + l \ \text{ for all } \ \Delta l = 0, ..., m - 1 - L, \\ L_{j+l+m-L}^{(m)} \leq m. \end{cases}$$

The theorem claims that the profiles of the m-ILC have very regular structure. If the current element of the profile, $L_{j-1}^{(m)}$, is greater than m then the next element, $L_j^{(m)}$, can be less than m, i.e., the profile 'falls into the pit'. In this case, the profile can stay in the pit for l times or jumps at the level $m+l$ and stays at this level for $m - L_j^{(m)}$ times. The parameters l and $m - L_j^{(m)}$ can be interpreted as the 'length' and the 'depth' of the pit, and the duality between them takes place.

Such a behaviour gives an opportunity to realize an interval atack on the stream cipher when the cipher is constructed using some complex scheme, but an eavesdropper approximates its fragments by LFSRs of length $< m$. Suppose that the eavesdropper has some set of the key words and assumes that they are written in the plain text. If he is right and the position of one of these words corresponds to a pit in the profile of the m-ILC, then he reconstructs the LFSR and reads all the other words of the plain text while the corresponding elements of the profile belong to this pit.

References

[1] R. A. Rueppel, "New approaches to stream ciphers," Ph.D.dissertation, Swiss Federal Institute of Technology, 1984.

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