On the Randomness of the Editing Generator

Enjian Bai and Guozhen Xiao

ISN National Key Lab, Xidian University, Xi'an, 710071, China
ejbai@sohu.com; gzxiao@xidian.edu.cn

Abstract. In their paper, G.Gong and S.Q.Jiang construct a new pseudo-random sequence generator by using two ternary linear feedback shift registers (LFSR). The new generator is called an *editing generator* which a combined model of the clock-controlled generator and the shrinking generator. For a special case (Both the base sequence and the control sequence are mm-sequence of degree n), the period, linear complexity, symbol distribution and security analysis are discussed in the same article. In this paper, we expand the randomness results of the edited sequence for general cases, we do not restrict the base sequence and the control sequence has the same length. For four special cases of this generator, the randomness of the edited sequence is discussed in detail. It is shown that for all four cases the editing generator has good properties, such as large periods, high linear complexities, large ratio of linear complexity per symbol, and small un-bias of occurrences of symbol. All these properties make it a suitable crypto-generator for stream cipher applications.

Keywords: Stream ciphers, Pseudo-random sequences, Clock-controlled generator, Shrinking generator

1 Introduction

Pseudo-random sequence generators that produce sequences with large periods, high linear complexities and good statistical properties are very useful for stream cipher applications. Pseudo-random sequence generators based on linear feedback shift registers (LFSR) are most common structures in practice due to their efficient hardware implementation. The structures can be classified into two classes. One class is apply a boolean function in n variables to a set of n LFSRs, such as a combinatorial function generator and a filtering generator [1]. The other class is to use one LFSR to control outputs of another LFSR. There are two different control models: the clock-controlled generators [2][3] and the shrinking generator [4], including self-shrinking generator [5][6].

Recently, a new type of Pseudo-random sequence generator that is called the editing generator (and referred to as EG) is introduced [7]. The EG is a sequence generator composed of two ternary LFSRs, LFSR **A** and LFSR **B**. This is a combined model of the clock-controlled generator and the shrinking generator. The randomness of the editing generator is discussed when LFSR **A** and LFSR **B** has the same length.

In this paper, we will expand the randomness results of this editing generator, i.e., we do not restrict LFSR **A** and LFSR **B** has the same length. For four special cases, (m-sequence, mm-sequence)-EG, (mm-sequence, mm-sequence)-EG, (m-sequence, m-sequence)-EG, and (mm-sequence, m-sequence)-EG, the randomness of the edited sequence is discussed in detail. It is shown that for all cases the editing generator has good properties, such as large periods, high linear complexities, large ratio of linear complexity per symbol, and small un-bias of occurrences of symbol.

This paper is organized as follows. In section 2, we will introduce the construction of the editing generator and determine the period of the edited sequence. In section 3, we will discuss the properties of randomness of the four types editing generator mentioned above. Section 4 is the conclusion.

Next, we introduce some basic definitions and lemmas which will be used in this paper in the remainder of this section.

Definition 1. *Let $S = \{s_i\}$ be a sequence over $GF(3)$, i.e., $s_i \in GF(3)$. If exist $r > 0$ such that $s_i = s_{i+r}$, $i = 0, 1, \cdots$, then S is called periodic and r is a period of S. The smallest positive integer r satisfying the equation is called the least period of S. Then we can write $S = (s_0, s_1, \cdots, s_{r-1})$.*

Definition 2. *Let $f(x) = x^n - c_{n-1}x^{n-1} - \cdots - c_1 x - c_0$, $c_i \in GF(3)$. If the elements of S satisfies the following linear recursive relation*

$$s_{n+k} = c_{n-1}s_{n-1+k} + \cdots + c_1 s_{k+1} + c_0 s_k, k \geq 0,$$

then S is said to be a LFSR sequence generated by $f(x)$ and $f(x)$ is called a characteristic polynomial of S. The characteristic polynomial $f(x)$ with the smallest degree is called the minimal polynomial of S. The degree of the minimal polynomial of S is called linear complexity of S. If $f(x)$ is a primitive polynomial over $GF(3)$, then S is called a m-sequence over $GF(3)$ of degree n.

Lemma 1. [7] *Let S be a m-sequence of degree n over $GF(3)$, Then S has the following randomness properties:*

1. *The least period of S is $3^n - 1$.*
2. *It satisfies the balance property, i.e., each non-zero elements in $GF(3)$ occurs 3^{n-1} times and zero element occurs $3^{n-1} - 1$ times in one period.*
3. *It satisfies the following run property. In each period,*
   (a) *for $1 \leq k \leq n-2$, the runs of each element in $GF(3)$ of length k occur $4 \cdot 3^{n-k-2}$ times.*
   (b) *the runs of each nonzero element of length $n-1$ occur once, the runs of each zero element of length $n-1$ occur twice.*
   (c) *the run of each nonzero element of length n occurs once.*
4. *Let $d = (3^n - 1)/2$. We write $S = (S_1, S_2)$, where $S_1 = (s_0, s_1, \cdots, s_{d-1})$ and $S_2 = (s_d, s_{d+1}, \cdots, s_{2d-1})$. Then $S_2 = 2S_1$.*

For more results on m-sequence over non-binary fields, the reader is referred to [8, Chapter 5].

Definition 3. *If a sequence is constructed by adding one zero into one of two zero runs of length $n-1$ in a m-sequence of degree n over $GF(3)$, then we call it a modified m-sequence of degree n over $GF(3)$, mm-sequence for short.*

From Lemma 1, we know that a mm-sequence has least period 3^n, each element in $GF(3)$ occurs exactly 3^{n-1} times and n length run of each element occurs exactly once in one period.

2 The Construction and Basic Properties

In this section, we will introduce the construction of the editing generator firstly, and then derive the period of the edited sequence.

2.1 The Construction

Let us recall the construction of EG proposed in [7].

Definition 4. [7] *Let $A = \{a_i\}$ and $B = \{b_j\}$ be two ternary sequences generated by LFSRs $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively. The output sequence $S = \{s_k\}$ is determined by the following rules. At time j, if $a_j = 2$, the generator discards the current output element in LFSR $\boldsymbol{B}$; else if $a_j = 1$, the generator outputs the current output element in LFSR $\boldsymbol{B}$; Otherwise, the generator outputs the previous output element in LFSR $\boldsymbol{B}$. The generator will be called an editing generator, the resulting sequence an edited sequence, denoted as Edit(A, B), the sequence A a control sequence, and the sequence B a base sequence.*

The definition can be written into the following pseudo-code:

Algorithm:(Edit(A, B))[7] Algorithm for Generating Edited Sequences
Input: A and B, two ternary LFSR sequences
Output: an edited sequence S of length l
(1) Set b_{-1} =0 (or 1, 2)
(2) Set $k = j = i$ =0
(3) **while** $(j < l)$ **do**
  a) **if** $a_j = 2$, then set $i = i + 1$
  b) **else if** $a_j = 1$, then set $s_k = b_i$; $k = k + 1$; $i = i + 1$
  c) **else** set $s_k = b_{i-1}$; $k = k + 1$
  d) $j = j + 1$
(4) **return** S.

We will use an example to explain this construction. Let A =(10112022), B =(100220122120200111021121 0), the process of generating the first 7 ele-

ments of the edited sequence S is shown as follows:

$$\begin{array}{l}
a_0 = 1,\ s_0 = b_0 = 1,\ i = 1, k = 1, j = 1;\\
a_1 = 0,\ s_1 = b_0 = 1,\ i = 1, k = 2, j = 2;\\
a_2 = 1,\ s_2 = b_1 = 0,\ i = 2, k = 3, j = 3;\\
a_3 = 1,\ s_3 = b_2 = 0,\ i = 3, k = 4, j = 4;\\
a_4 = 2,\qquad\qquad\ \ i = 4, k = 4, j = 5;\\
a_5 = 0,\ s_4 = b_3 = 2,\ i = 4, k = 5, j = 6;\\
a_6 = 2,\qquad\qquad\ \ i = 5, k = 5, j = 7;\\
a_7 = 2,\qquad\qquad\ \ i = 6, k = 5, j = 8;\\
a_8 = 1,\ s_5 = b_6 = 0,\ i = 7, k = 6, j = 9;\\
a_9 = 0,\ s_6 = b_6 = 0,\ i = 7, k = 7, j = 10.
\end{array}$$

Property 1. With the notation in Definition 4, let

$$a_t^{'} = \begin{cases} 0, \text{if } a_t = 0 \\ 1, \text{otherwise} \end{cases} \text{ and } a_t^{''} = \begin{cases} 0, \text{if } a_t = 2 \\ 1, \text{otherwise} \end{cases},$$

then the elements of the edited sequence $S = \{s_k\}$ can be represented by

$$s_{k_j} = b_{i_j}, \text{where } i_j = \sum_{t=0}^{j} a_t^{'} - 1 \text{ and } k_j = \sum_{t=0}^{j} a_t^{''} - 1$$

Property 2. let $i^{'} \geq 0$ and suppose that $a_i^{'} \neq 2$. Let $i+1$ be the total number of zeros and ones in $a_0, a_1, \cdots, a_i^{'}$, n_2 the total number of twos in $a_0, a_1, \cdots, a_i^{'}$, and n_0 the total number of zeros in $a_0, a_1, \cdots, a_i^{'}$, then $S = \{s_k\}$ can be represented as well as

$$s(i) = b(k_i), \text{where } k_i = i + n_2 - n_0. \tag{1}$$

Remark 1. If the elements of A and B only take the values 1 and 2, then Edit(A, B) is degenerated into a shrunk sequence; if the elements of A and B only take the value 0 and 1, then Edit(A, B) is degenerated into a clock-controlled sequence. Therefore, the construction of EG is a combined model of the clock-control generator (viewed as insertion) and the shrinking generator (viewed as deletion).

2.2 The Period of Edit(A,B)

Suppose that the ternary LFSRs **A** and **B** have length m and n ($m \leq n$) respectively. Let A and B denote the output sequences of **A** and **B**. Suppose that A and B are periodic of least periods T_A and T_B respectively. Let S be the output sequence of this editing generator. Let ϖ_A be the total number of zeros and ones, ω_A be the total number of ones and twos in a full period of A.

After $T_A \cdot (T_B/\gcd(T_B, \omega_A))$ clock pulses have been applied to LFSR **A**, $\omega_A \cdot (T_B/\gcd(T_B, \omega_A))$ clock pulses have been applied to LFSR **B**. Then both **A** and **B** return to their initial states. Hence the output sequence S is periodic with period dividing $\varpi_A \cdot (T_B/\gcd(T_B, \omega_A))$.

In the following lemma, it is shown that the maximum period of the output sequence S can be attained. The proof of the lemma is similar with the proof of shrinking generator [4].

Lemma 2. *Let A and B form an editing generator S. Then the edited sequence S has the least period $\varpi_A \cdot (T_B/\gcd(T_B, \omega_A))$.*

Proof. For simplicity of the proof we assume $m < d = T_B/\gcd(T_B, \omega_A)$ and recall that $s(i) = b(k_i)$.

Fact 1: Recall that in a full period of A the number of 0's and 1's are ϖ_A, so when considering a full period of A there are ϖ_A outputs s_i and the sequence B advances ω_A elements, so $\forall j \geq 0$,

$$s(i + j\varpi_A) = b(k_i + j\omega_A). \tag{2}$$

Fact 2: let k and $k^{'}$ be any pair of indices. If $\forall j, b(k + j\omega_A) = b(k^{'} + j\omega_A)$, then $d = T_B/\gcd(T_B, \omega_A)$ divides $(k - k^{'})$.

In fact, define a sequence C where $c(t) = b(t\omega_A)$ $\forall t \geq 0$. The sequence C is a decimation of sequence B by ω_A. As B has period T_B, then the sequence C has period $d = T_B/\gcd(T_B, \omega_A)$. Hence the result.

Let T be the period of the sequence S. By the argument given above T must divide $\varpi_A \cdot (T_B/\gcd(T_B, \omega_A))$. We now proceed to show that $\varpi_A \cdot (T_B/\gcd(T_B, \omega_A))$ divides T. By definition $s(i) = s(i + T)$. In particular, $\forall i, j$, $s(i + j\varpi_A) = s(i + T + j\varpi_A)$. Using (2) we get, $\forall i, j$: $b(k_i + j\omega_A) = b(k_{i+T} + j\omega_A)$. Using Fact 2, we have

$$\forall i, d = T_B/\gcd(T_B, \omega_A) \text{ divides } k_{i+T} - k_i. \tag{3}$$

Next step is to show that (3) is possible only if ϖ_A divides T. Rewrite (3) as follows:

$$\forall i, \exists j_i : k_{i+T} = k_i + j_i d. \tag{4}$$

Putting $i + 1$ instead of i in (4) we get

$$k_{i+1+T} = k_{i+1} + j_{i+1} d. \tag{5}$$

Subtracting (4) from (5) we get:

$$\forall i, k_{i+1+T} - k_{i+T} = k_{i+1} - k_i + (j_{i+1} - j_i)d. \tag{6}$$

Notice the definition of k_j, if $j_{i+1} - j_i$ were different than zero, it would imply the existence of at least d consecutive 2's in the A-sequence, which is impossible assuming $m < d$. Therefore, $j_{i+1} - j_i = 0$, and then $\forall i, k_{i+1+T} - k_{i+T} = k_{i+1} - k_i$.

The latter implies that the subsequence of A starting at $a(k_i)$ is identical to the subsequence starting at $a(k_{i+T})$. This means that T_A divides $k_{i+T} - k_i$, or equivalently, that the number of elements in the A-sequence between $a(k_i)$ and $a(k_{i+T})$ is a multiple of its period. But then the number of 0's and 1's in this argument is a multiple of ϖ_A. On the other hand, the number of 0's and 1's is exactly T, thus proving that ϖ_A divides T.

Let h be such that

$$T = h\varpi_A. \tag{7}$$

We have for all j:

$$b(k_0) = s(0) = s(jh\varpi_A) = b(k_0 + jh\omega_A). \tag{8}$$

The last equality follows from (2). So $\forall j : b(k_0) = b(k_0 + jh\omega_A)$. This implies that T_B divides $h\omega_A$, i.e., $T_B/\gcd(T_B, \omega_A)$ divides $h[\omega_A/\gcd(T_B, \omega_A)]$ and since $\gcd(T_B/\gcd(T_B, \omega_A), \omega_A/\gcd(T_B, \omega_A)) = 1$, then $T_B/\gcd(T_B, \omega_A)$ divides h. From (7) $d\varpi_A$ divides T.

Hence, the least period T of S is equal to $\varpi_A \cdot (T_B/\gcd(T_B, \omega_A))$.

In the rest of this paper, we will restrict ourselves into the special case that the control sequence A and base sequence B are ternary m-sequence or ternary mm-sequence respectively. Therefore there are four cases, i.e., (m-sequence, mm-sequence)-EG, (mm-sequence, mm-sequence)-EG, (m-sequence, m-sequence)-EG, and (mm-sequence, m-sequence)-EG. We will analysis the properties of randomness of the edited sequences in the following section.

3 Properties of the Output Sequence S of EG-Special Cases

In this section, we will consider in details some properties of the output sequence S of EG, such as the least period of the edited sequence, the bound for the linear complexity, ratio of linear complexity to period, and occurrences of symbols when we consider the four special cases of this generator mentioned in section 2. We will keep the notation in Section 2.

3.1 Analysis of (m-sequence, mm-sequence)-EG

Suppose that control sequence A and base sequence B are m-sequence and mm-sequence respectively. The period of A is $T_A = 3^m - 1$, the period of B is $T_B = 3^n$ $(m \leq n)$.

The control sequence A is a m-sequence of period $T_A = 3^m - 1$, so the number of ones and twos is $\omega_A = 2 \cdot 3^{m-1}$, the number of zeros and ones is $\varpi_A = 2 \cdot 3^{m-1} - 1 = \omega_A - 1$ in a full period of A. Hence we get the following theorem by lemma 2 directly.

Theorem 1. *Let A and B form an editing generator S, A is a m-sequence and B is a mm-sequence with period $T_A = 3^m - 1$ and $T_B = 3^n$ respectively. Then the edited sequence S has least period $(\omega_A - 1) \cdot 3^{n-m+1}$.*

In the next theorem, we will establish a lower bound for the linear complexity of the edited sequences.

Theorem 2. *Under the condition of Theorem 1, the edited sequence S has linear complexity LC, where $LC > (\omega_A - 1) \cdot 3^{n-m}$.*

Proof. From Theorem 1, $f(x) = x^{(\omega_A-1)\cdot 3^{n-m+1}} - 1$ is a characteristic polynomial of S over $GF(3)$. Compare with [9, Sec.4.4], The canonical factorization of $f(x) \in F_3[x]$ is then given by

$$\begin{aligned} f(x) &= x^{(\omega_A-1)\cdot 3^{n-m+1}} - 1 \\ &= (x^{\omega_A-1} - 1)^{3^{n-m+1}} \\ &= \left(\prod_{i=1}^{h} f_i(x)\right)^{3^{n-m+1}}, \end{aligned} \tag{9}$$

where $f_i(x) = \prod_{j\in C_i}(x - \alpha^j), 1 \le i \le h, \alpha$ be a fixed primitive $(\omega_A - 1)$th root of unity in some extension field of $GF(3)$, and $C_1, C_2, \cdots, C_h$ are different cyclotomic cosets modulo $(\omega_A - 1)$. Let $m(x)$ be the minimal polynomial of S. Then $m(x) = \prod_{i=1}^{h}[f_i(x)]^{l_i}$ for some integer $l_i \le 3^{n-m+1}, 1 \le i \le h$. If $l_i \le 3^{n-m}, 1 \le i \le h$. Then $m(x)$ divides $\left(\prod_{i=1}^{h} f_i(x)\right)^{3^{n-m}} = x^{(\omega_A-1)\cdot 3^{n-m}} - 1$. Thus the least period of S is less than or equal to $(\omega_A - 1)\cdot 3^{n-m}$, which is a contradiction with Theorem 1. So LC$> (\omega_A - 1)\cdot 3^{n-m}$.

Corollary 1. *Under the condition of Theorem 1. Let η be the ratio of the linear complexity to the least period of the edited sequence (represents linear complexity per symbol). Then $\eta > 1/3$.*

In the following theorem, we will establish a bound for frequency of each element in $GF(3)$ occurs in the edited sequence.

Theorem 3. *Under the condition of Theorem 1. Each element in $GF(3)$ occurs at least 3^{n-1} times in a full period of the edited sequence S.*

Proof. Let $\tilde{D} = (\tilde{a}_0, \tilde{a}_1, \cdots, \tilde{a}_{3^{m-1}-1})$ be the resulting sequence from the first $(3^m - 1)/2$ elements of A by deleting all 0's, and let $\tilde{A} = (\tilde{a}_0, \tilde{a}_1, \cdots, \tilde{a}_{2\cdot 3^{m-1}-1})$ be the resulting sequence of A by deleting all 0's. Then $\tilde{A}$ has a period of $2\cdot 3^{m-1}$. From Lemma 1 (4), we have $\tilde{A} = (\tilde{D}, 2\tilde{D})$. We write $B = (B_0, B_1, \cdots, B_{3^{n-m+1}-1})$, where$B_i = (b_{i\cdot 3^{m-1}}, b_{i\cdot 3^{m-1}+1}, \cdots, b_{i\cdot 3^{m-1}+3^{m-1}-1})$, $i = 0, 1, \cdots, 3^{n-m+1} - 1$. Let $\tilde{S} = \text{Edit}(\tilde{A}, B)$, an edited sequence from the base sequence B and the control sequence $\tilde{A}$. Then $\tilde{S}$ is the resulting sequence from S by deleting all elements of S which are contributed from the inserting operation. Thus all elements occur in $\tilde{S}$ must occur in S. Note that there is no zeros in $\tilde{A}$, then Edit$(\tilde{A}, B)$ is a shrinking generator. In other words, $\tilde{a}_j$ controls b_j for all $j \ge 0$. Therefore, $(\tilde{D}, 2\tilde{D}, \tilde{D}, 2\tilde{D}, \cdots, \tilde{D}, 2\tilde{D})$ controls $(B_0, B_1, \cdots, B_{3^{n-m+1}-1}, B_0, B_1, \cdots, B_{3^{n-m+1}-1})$ term by term. Consequently, each of $\tilde{D}$ and $2\tilde{D}$ controls $B_i, i = 0, 1, \cdots, 3^{n-m+1} - 1$ once. From Edit$(\tilde{D}, B_i)$ and Edit$(2\tilde{D}, B_i)$, we get all the elements of B_i, $i = 0, 1, \cdots, 3^{n-m+1} - 1$. Thus the set of elements of $\tilde{S}$ is equal to the set of the elements of B. Since each element in $GF(3)$ occurs in B exactly 3^{n-1} times and $\tilde{S}$ is a subsequence of S. Then each element occur in S at least 3^{n-1}.

3.2 Analysis of (mm-sequence, mm-sequence)-EG

Suppose that both control sequence A and base sequence B are mm-sequence. The period of A is $T_A = 3^m$, the period of B is $T_B = 3^n$ $(m \leq n)$.

The control sequence A is a mm-sequence of period $T_A = 3^m$, so the number of ones and twos is$\omega_A = 2{\cdot}3^{m-1}$, the number of zeros and ones is $\varpi_A = 2{\cdot}3^{m-1} = \omega_A$ in a full period of A. Hence we get the following theorem by lemma 2 directly.

Theorem 4. *Let A and B form an editing generator S, Both A and B are mm-sequence with period $T_A = 3^m$ and $T_B = 3^n$ respectively. Then the edited sequence S has least period $2T_B$.*

Theorem 5. *Under the condition of Theorem 4, the edited sequence S has linear complexity LC, where $LC > 2 \cdot 3^{n-1}$.*

Proof. To show the correctness on the linear complexity of the sequence S it suffices to present a polynomial $P(x)$ of degree d for which the coefficients of P represent a linear relation satisfied by the elements of S [10]. That is

$$P(x) = \sum_{i=0}^{d} p_i x^i \text{ then } \sum_{i=0}^{d} p_i s_{i+t} = 0 \ \forall t \geq 0. \tag{10}$$

Let S^{ω_A} denote the sequence S decimated by ω_A, i.e., the sequence $s(j\omega_A)$, $j \geq 0$. Fact 1 in the proof of Lemma 2 states that this decimation, written in terms of the B-sequence, results in a sequence of the form $b(i + j\omega_A)$. The latter is a subsequence of B with period $T_B/\gcd(T_B, \omega_A) = 3^{n-m+1}$. Let it satisfies a polynomial $Q(x)$. Then $Q(x)$ must be the format $(x-1)^u$, $3^{n-m} < u \leq 3^{n-m+1}$. But then also the decimated sequence S^{ω_A} satisfies the polynomial, i.e., $Q(E)(S^{\omega_A}) = Q(E)(s(j\omega_A)) = 0$, where E is the shift operator. Therefore, we have found a polynomial $P(S) = Q(S^{\omega_A})$ of degree $u \cdot \omega_A$, such that $P(S)$=0, and then the lower bound of the linear complexity of S is $u{\cdot}\omega_A > 3^{n-m}{\cdot}2{\cdot}3^{m-1} = 2 \cdot 3^{n-1}$.

Corollary 2. *Under the condition of Theorem 4. Let η be the ratio of the linear complexity to the least period of the edited sequence (represents linear complexity per symbol). Then $\eta > 1/3$.*

Theorem 6. *Under the condition of Theorem 4. Each element in $GF(3)$ occurs at least 3^{n-1} times in a full period of the edited sequence S.*

Proof. The proof is the same as that of Theorem 3.

Remark 2. When $m = n$, The case has been discussed by G.Gong and S.Q.Jiang in [7]. The result of Theorem 4 $\sim$ Theorem 6 is more tight than theirs. We confirm that the least period of the edited sequence is $2 \cdot 3^n$, and the lower bound of the linear complexity is $2{\cdot}3^{n-1}$. From the results, we can see that in the case of (mm-sequence, mm-sequence)-EG the period and linear complexity is independent of the length of LFSR **A**.

3.3 Analysis of (m-sequence, m-sequence)-EG

Suppose that both control sequence A and base sequence B are m-sequence. The period of A is $T_A = 3^m - 1$, the period of B is $T_B = 3^n - 1$ $(m \leq n)$.

The control sequence A is a m-sequence of period $T_A = 3^m - 1$, so the number of ones and twos is $\omega_A = 2 \cdot 3^{m-1}$, the number of zeros and ones is $\varpi_A = 2 \cdot 3^{m-1} - 1 = \omega_A - 1$ in a full period of A. Hence we get the following theorem by lemma 2 directly.

Theorem 7. *Let A and B form an editing generator S, both A and B are m-sequence with period $T_A = 3^m - 1$ and $T_B = 3^n - 1$ respectively. Then the edited sequence S has least period $(\omega_A - 1) \cdot (T_B/2)$.*

Theorem 8. *Under the condition of Theorem 7, the edited sequence S has linear complexity LC, where $LC = n \cdot (\omega_A - 1)$.*

Proof. Let $S^{(\omega_A - 1)}$ denote the sequence S decimated by $(\omega_A - 1)$, i.e., the sequence $s(j(\omega_A - 1))$, $j \geq 0$. Fact 1 in the proof of Lemma 2 states that this decimation, written in terms of the B-sequence, results in a sequence of the form $b(i + j\omega_A)$. The latter is a subsequence of B with period $T_B / \gcd(T_B, \omega_A) = (3^n - 1)/2$. Let $Q(x)$ be its minimal polynomial.

Letβ be a root of the minimal polynomial $f(x)$ of B, then its order is $T_B = 3^n - 1$. While β^{ω_A} is a root of the minimal polynomial $Q(x)$ of $b(i + j\omega_A)$, its order is $(3^n - 1)/2$. Suppose that v is the order of 3 modulo $(3^n - 1)/2$, i.e., $3^v \equiv 1 \bmod (3^n - 1)/2$. Obviously, $v = n$ is the smallest integer such that the equation, therefore the degree of the minimal polynomial of $b(i + j\omega_A)$ is n. But then also the decimated sequence $S^{(\omega_A - 1)}$ satisfies the polynomial, i.e., $Q(E)(S^{(\omega_A - 1)}) = Q(E)(s(j(\omega_A - 1))) = 0$, where E is the shift operator . Therefore, we have found a polynomial $P(S) = Q(S^{(\omega_A - 1)})$ of degree $n \cdot (\omega_A - 1)$, such that $P(S)$=0.

Let $m(x)$ be the minimal polynomial of S, then $m(x)$ divides $P(x)$. While $P(x)$ is an irreducible polynomial, therefore $m(x) = P(x)$. Hence the linear complexity of the edited sequence is $n \cdot (\omega_A - 1)$.

Corollary 3. *Under the condition of Theorem 7. Let η be the ratio of the linear complexity to the least period of the edited sequence (represents linear complexity per symbol). Then $\eta > n/3^n$.*

About the occurrences of symbols of edited sequence in the case of (m-sequence, m-sequence)-EG we only have the following experiment result.

Conjecture 1. In the case of (m-sequence, m-sequence)-EG, the 0's element occurs at least $2 \cdot 3^{m-2} \cdot (3^{n-1} - 1)$ times, the 1's and 2's element occur at least $2 \cdot 3^{m-2} \cdot 3^{n-1}$ times in a full period of the edited sequence S respectively.

3.4 Analysis of (mm-sequence, m-sequence)-EG

Suppose that control sequence A and base sequence B are mm-sequence and m-sequence respectively. The period of A is $T_A = 3^m$, the period of B is $T_B = 3^n - 1$ $(m \leq n)$.

The control sequence A is a mm-sequence of period $T_A = 3^m$, so the number of ones and twos is $\omega_A = 2 \cdot 3^{m-1}$, the number of zeros and ones is $\varpi_A = 2 \cdot 3^{m-1} = \omega_A$ in a full period of A. Hence we get the following theorem by lemma 2 directly.

Theorem 9. *Let A and B form an editing generator S, A is a mm-sequence and B is a m-sequence with period $T_A = 3^m$ and $T_B = 3^n - 1$ respectively. Then the edited sequence S has least period $3^{m-1} \cdot T_B$.*

Theorem 10. *Under the condition of Theorem 9, the edited sequence S has linear complexity LC, where $2n \cdot 3^{m-2} < LC \leq 2n \cdot 3^{m-1}$.*

Proof. Upper bound on the linear complexity: The proof is similar with the proof of Theorem 8. We can find a polynomial $Q(x)$ of degree n such that $P(S) = Q(S^{\omega_A})$=0, and then the linear complexity of the edited sequence is at most $n \cdot \omega_A$.

Lower bound on the linear complexity: The proof is similar with the proof of Theorem 2. Let $m(x)$ denote the minimal polynomial of S. Since the sequence is satisfies $Q(S^{\omega_A})$=0, we have that $m(x)$ must divide $Q(x^{\omega_A})$. Since $\omega_A = 2 \cdot 3^{m-1}$, we have $Q(x^{\omega_A}) = (Q(x^2))^{3^{m-1}}$, and then $m(x)$ must be of the format $(Q(x^2))^u, u \leq 3^{m-1}$. Assume$u \leq 3^{m-2}$. Then $m(x)$ divides $(Q(x^2))^{3^{m-2}}$. While $Q(x^2)$ divides $x^{T_B} - 1$, therefore, $m(x)$ divides $(x^{T_B} - 1)^{3^{m-2}} = x^{3^{m-2}T_B} - 1$, but then the period of S is at most $3^{m-2} \cdot T_B$ contradicting Theorem 9. Therefore, $u > 3^{m-2}$ and the lower bound follows.

Corollary 4. *Under the condition of Theorem 9. Let η be the ratio of the linear complexity to the least period of the edited sequence (represents linear complexity per symbol). Then $2n/(3^{n+1}) < \eta \leq 2n/(3^n - 1)$.*

We have the same conjecture about the occurrences of symbols of edited sequence in the case of (mm-sequence, m-sequence)-EG as Conjecture 1.

Remark 3. When $n \geq m > 1$, $(\omega_A - 1) \cdot 3^{n-m+1} < 2 \cdot 3^n < (\omega_A - 1) \cdot (3^n - 1) < 3^{m-1} \cdot (3^n - 1)$. Therefore, the editing generator with a m-sequence as the base sequence generate sequences with large periods and high linear complexities than the editing generator with a mm-sequence as the base sequence. The (mm-sequence, m-sequence)-EG has the maximum least period. Furthermore, all these four types generate edited sequence with good symbol distribution.

Remark 4. In [7], the author proposed three methods (Brute-force like attack, Unconstraint embedding attack, and Entropy attack) to attack this new generator. The analysis shows that the edited sequences resist to all of these three attacks in the case of (mm-sequence, mm-sequence)-EG and $m = n$. From the

process of analysis, we find the three attacks can be applied to a general case because the analysis is independent of the length of LFSR **A** and LFSR **B**. Therefore, the four cases we discussed in this paper also resist these three attacks, and then the edited sequence can be used in the stream cipher applications.

4 Conclusion

In this paper, we expand the randomness properties of the editing generator. A general result of least period is given and four special cases are discussed in detail. We have shown that the edited sequences generated by the four types editing generator have good properties of randomness, such as large period, high linear complexity, large rate of linear complexity per symbol, and low un-bias of occurrences of symbols. Furthermore, the edited sequence can resist three attacks: the brute-force like attack, the unconstraint embedding attack, and the entropy attack. All these properties enhance its use as a suitable crypto-generator for stream cipher applications.

References

1. R.A.Rueppel, Stream ciphers, Contemporary Cryptology, the Science of Information, IEEE Press, pp.65-134, 1992.
2. Dieter Gollman and W.G.Chambers, "Clock-controlled shift registers: a review, "IEEE Journal on Selected Areas in Communications. vol.7, no.4, pp.525-533, 1989.
3. A.Kanso, Clock-controlled generators, PhD thesis, University of London, 1999.
4. D.Coppersmith, H.Krawczys, and Y.Mansour, "The shrinking generator," Advances in Cryptology-Crypt'93, LNCS 773, pp.22-39, Springer-Verlag, 1994.
5. W.Meier and O.Staffelbach, "The self-shrinking generator," Advances in Cryptology, Eurocrypt'94, LNCS 950, pp.05-214, Springer-Verlag, 1994.
6. Yupu HU and Guozhen XIAO, "The generalized self-shrinking generator," IEEE Trans. on Info. Theory, to appear.
7. G.Gong and J.S.Quan, "The editing generator and its cryptanalysis," `http://www.cacr.math.uwaterloo.ca/`.
8. M.K.Simon, J.K.Omura, R.A.Scholtz, and B.K.Levitt, Spread Spectrum Communications Handbook, McGrawHill, Inc., Revised version, 1994.
9. V.Pless, Introduction to the Theory of Error-Correcting Codes, 2^{nd} ed. New York: Wiley, 1989.
10. R.Lidl and H.Niederreiter, Introduction to Finite Fields and Their Applications, UK: Combridge University Press, 1986.