The Alternating Step(*r, s*) Generator

A. A. Kanso[1]

1: King Fahd University of Petroleum and Minerals
P. O. Box 2440, Hail, Saudi Arabia
akanso@hotmail.com

Abstract

A new construction of a pseudo-random generator based on a simple combination of three feedback shift registers (FSRs) is introduced. The main characteristic of its structure is that the output of one of the three FSRs controls the clocking of the other two FSRs. This construction allows users to generate a large family of sequences using the same initial states and the same feedback functions of the three combined FSRs. The construction is related to the Alternating Step Generator that is a special case of this construction. The period, and the lower and upper bound of the linear complexity of the output sequences of the construction whose control FSR generates a de Bruijn sequence and the other two FSRs generate m-sequences are established. Furthermore, it is established that the distribution of short patterns in these output sequences occur equally likely and that they are secure against correlation attacks. All these properties make it a suitable crypto-generator for stream cipher applications.

Keywords. Feedback Shift Registers, Stream Ciphers, Clock-Controlled Registers, Alternating Step Generator.

1. Introduction

A k-stage *feedback shift register* (FSR) is a device that generates binary sequences.

An FSR is made up of two parts: a shift register S, and a feedback function Q. The shift register S consists of k stages $\mathbf{S}_0, \mathbf{S}_1, ..., \mathbf{S}_{k-1}$ which contains one bit 0 or 1. The contents of these stages at a given time t is known as the state of the register S and is denoted by: $\underline{S}_t = S_0(t), S_1(t), ..., S_{k-1}(t)$. (Where at time $t = 0$ the state $\underline{S}_0 = S_0(0), S_1(0), ..., S_{k-1}(0)$ is called the initial state of S).

The feedback function Q is a function that maps the state of the register S to the bit 0 or 1. At time t, $Q(S_0(t), ..., S_{k-1}(t)) = 0$ or 1.

The shift register S is clocked at a time interval, when this happens the contents of S are shifted one bit to the left (i.e., the content of $\mathbf{S}_i$ is transferred into $\mathbf{S}_{i-1}(i = 1, 2, .., k-1)$) and the new content of $\mathbf{S}_{k-1}$ is computed by applying the feedback function Q to the old contents of S.

The above can be expressed as follows:

$$\begin{cases} S_i(t+1) &= S_{i+1}(t) \qquad \text{for } i = 0, 1, ..., k-2 \\ S_{k-1}(t+1) &= Q(S_0(t), ..., S_{k-1}(t)) \end{cases}$$

The binary sequence (S_t) generated by this device is the sequence of contents of the 0^{th} stage $\mathbf{S}_0$ of S for all t (i.e., the binary sequence $(S_t) = S_0, S_1, S_2, ...$ where $S_t = S_0(t) \in GF(2)$ for $t = 0, 1, 2, ...$).

The state sequence of this device is given by the sequence of states of the register S:

$$(\underline{S}_t) = \underline{S}_0, \underline{S}_1, ...$$

where $\underline{S}_t = S_0(t), ..., S_{k-1}(t)$ for $t = 0, 1, 2, ...$

Since the output sequence of a feedback shift register is the content of the 0^{th} stage of the register then clearly each of the output sequence (S_t) and the state sequence $(\underline{S}_t)$ determine the other.

If the feedback function Q of a feedback shift register can be written in the form: $Q(S_0(t), ..., S_{k-1}(t)) = (C_0 S_0(t) \oplus ... \oplus C_{k-1} S_{k-1}(t))$ for a given time t, for some binary constants $C_0, C_1, ..., C_{k-1}$ called the feedback coefficients, then the shift register is called *linear*, where $\oplus$ denotes addition modulo 2.

The feedback coefficients $C_0, ..., C_{k-1}$ determine a polynomial $C_0 \oplus C_1 x^1 \oplus ... \oplus C_{k-1} x^{k-1} \oplus x^k$ of degree k associated with the feedback function Q. We write $\mu(x)$ to denote this polynomial and call it the *characteristic feedback polynomial* of the linear feedback shift register.

Any k-stage linear feedback shift register can be uniquely described by a characteristic feedback polynomial $\mu(x)$ over the finite field of order 2 of the form: $\mu(x) = C_0 \oplus C_1 x^1 \oplus ... \oplus C_{k-1} x^{k-1} \oplus x^k$.

1.1. Construction

Keystream sequence generators that produce sequences with large periods, high linear complexities and good statistical properties are very useful as building blocks for stream cipher applications. The use of clock-controlled generators in keystream generators appears to be a good way of achieving sequences with these properties [1].

In this paper, a new clock-controlled generator that is called the *Alternating Step(r, s) Generator* (and referred to as ASG(r, s)) is introduced. The ASG(r, s) is a sequence generator composed of three FSRs **A**, **B** and **C** [2] which are interconnected such that **B** is clocked by the constant integer r and **C** is not clocked if the content of the 0^{th} stage of **A** is 1, otherwise, **B** is not clocked and **C** is clocked by the constant integer s. FSR **A** is called the *control register* and FSRs **B** and **C** are called the *generating registers*. The output bits of the ASG(r, s) are produced by adding modulo 2 the output bits of FSRs **B** and**C** under the control of FSR **A**.

Suppose that the control register FSR **A** has k stages and feedback function R. Similarly, suppose that the generating registers FSRs **B** and **C** have m and n stages respectively and feedback functions S and T respectively. Let $\underline{A}_0 = A_0(0), A_1(0), ..., A_{k-1}(0)$, $\underline{B}_0 = B_0(0), B_1(0), ..., B_{m-1}(0)$ and $\underline{C}_0 = C_0(0), C_1(0), ..., C_{n-1}(0)$ be the initial states of **A**, **B** and **C** respectively.

The initial state of the ASG(r, s) at time $t = 0$ is given by: $\underline{S}_0 = (\underline{A}_0, \underline{B}_0, \underline{C}_0)$.

Define a function F that acts on the state of FSR **A** at a given time t to determine the number of times FSR **B** or FSR **C** is clocked such that: at any time t, $F(\underline{A}_t) = rA_0(t) + s(A_0(t) \oplus 1)$.

Define two cumulative functions of FSR **A**, G_A and Q_A: $\{0, 1, 2, ...\} \rightarrow \{0, 1, 2, ...\}$ such that:

$$G_A(t) = \sum_{i=0}^{t-1} A_0(i) F(\underline{A}_i) = r \sum_{i=0}^{t-1} A_0(i), \text{ for } t > 0, \text{ and } G_A(0) = 0,$$

and

$$Q_A(t) = \sum_{i=0}^{t-1} (A_0(i) \oplus 1) F(\underline{A}_i) = s \sum_{i=0}^{t-1} (A_0(i) \oplus 1), \text{ for } t > 0, \text{ and } Q_A(0) = 0.$$

Thus, with initial state $\underline{S}_0 = (\underline{A}_0, \underline{B}_0, \underline{C}_0)$, at time t the state of the ASG(r, s) is given by: $\underline{S}_t = (\underline{A}_t, \underline{B}_{G_A(t)}, \underline{C}_{Q_A(t)})$.

At any time t, the output of the ASG(r, s) is the content of the 0^{th} stage of **B** added modulo 2 to the content of the 0^{th} stage of **C**, i.e., $B_0(G_A(t)) \oplus C_0(Q_A(t))$.

The ASG(r, s) may also be described in terms of the three output sequences $(A_t), (B_t)$ and (C_t) of the feedback shift registers **A**, **B** and **C** respectively.

Acting on their own, suppose that FSR **A**, FSR **B** and FSR **C** produce output sequences $(A_t) = A_0, A_1, ...,$ $(B_t) = B_0, B_1, ...,$ and $(C_t) = C_0, C_1, ...$ respectively. The sequence (A_t) is called the *control sequence*, and the sequences (B_t) and (C_t) are called the *generating sequences* of the ASG(r, s) respectively and referred to as component sequences.

The output sequence (Z_t) of the ASG(r, s) whose control and generating sequences are (A_t), (B_t) and (C_t) respectively is given by: $Z_t = B_{G_A(t)} \oplus C_{Q_A(t)}$ where:

$$G_A(t) = r\sum_{i=0}^{t-1} A_i \text{ and } Q_A(t) = s\sum_{i=0}^{t-1}(A_i \oplus 1),\ \text{for } t > 0, \text{and } G_A(0) = Q_A(0) = 0.$$

2. Properties of the Output Sequence (Z_t) of the ASG(r, s)

Suppose that **A** is an FSR with initial state $\underline{A}_0$ and feedback function R such that the output sequence (A_t) of **A** is a de Bruijn sequence of span κ and it has period $K = 2^\kappa$ [2]. Suppose that the feedback shift registers **B** and **C** are primitive linear feedback shift registers (LFSRs) with non-zero initial states $\underline{B}_0$ and $\underline{C}_0$ respectively, and primitive characteristic feedback polynomials $g(x)$ of degree m and $h(x)$ of degree n respectively (where $g(x)$ and $h(x)$ are associated with the feedback functions S and T respectively) [2]. Let (B_t) and (C_t) denote the output sequences of LFSRs **B** and **C** respectively. Then (B_t) and (C_t) are m-sequences of periods $M = (2^m - 1)$ and $N = (2^n - 1)$ respectively [2]. Let (Z_t) be the output sequence of the ASG(r, s) whose component sequences are (A_t), (B_t) and (C_t).

Note that a de Bruijn sequence of span κ can be easily obtained from an m-sequence generated by a κ-stage primitive LFSR by simply adding a 0 to the end of each subsequence of $(\kappa - 1)$ 0s occurring in the m-sequence.

In a full period $K = 2^\kappa$ of (A_t) the number of ones and zeroes is $K_1 = K_0 = 2^{\kappa-1}$ [2]. Thus, after clocking FSR **A** K times, LFSR **B** is clocked $G_A(K) = r2^{\kappa-1}$ times and LFSR **C** is clocked $Q_A(K) = s2^{\kappa-1}$ times.

In this section, some properties of the output sequences such as period and linear complexity are established. It is shown that, when m and n are positive integers greater than 1 satisfying $gcd(m, n) = 1$, and r and s satisfy $gcd(r, 2^m - 1) = gcd(s, 2^n - 1) = 1$, then the period of the output sequences is exponential in κ, m and n, and that the linear complexity is exponential in κ. Finally, it is established that the distribution of short patterns in the output sequences of this ASG(r, s) turns out to be ideal.

2.1. Period and Linear Complexity of (Z_t)

The output sequence (Z_t) can be seen as two sequences added modulo 2, $(Z_t) = (B_{G_A(t)}) \oplus (C_{Q_A(t)})$, where $(B_{G_A(t)})$ and $(C_{Q_A(t)})$ are generated by the sub-generators whose component sequences are (A_t), (B_t) and (A_t), (C_t) respectively.

In order to establish the period and the linear complexity of (Z_t) one needs to first consider the periods and the linear complexities of the two sequences $(B_{G_A(t)})$ and $(C_{Q_A(t)})$.

In the following two lemmas, the periods of the sequences $(B_{G_A(t)})$ and $(C_{Q_A(t)})$ are considered. Tretter [3] has considered this proof for the output sequences of the stop and go generator [4]. His proof is also valid for the sequences $(B_{G_A(t)})$ and $(C_{Q_A(t)})$.

Lemma 1 *If $gcd(r, 2^m - 1) = 1$, then the period P_G of the sequence $(B_{G_A(t)})$ is $2^\kappa(2^m - 1)$.*

Proof. The sequence $(B_{G_A(t)})$ will repeat whenever the states of the shift registers **A** and **B** return to their initial states $\underline{A}_0$ and $\underline{B}_0$ respectively. The register **A** returns to its initial state once every $K = 2^\kappa$ clock pulses. Thus, for λ cycles of register **A**, register **B** is clocked $\lambda G_A(K)$ times.

Therefore, if for some integers U and λ, $\lambda G_A(K) = UM$, then the feedback shift registers **A** and **B** will simultaneously be in their initial states. The period of the sequence $(B_{G_A(t)})$ corresponds to the smallest integer value that the integer U can take.

Now $U = \lambda G_A(K)/M$. Therefore, if $gcd(G_A(K), M) = 1$ i.e., $gcd(r2^{\kappa-1}, 2^m - 1) = 1$, then the smallest value that U can take is when $\lambda = M$. Clearly $gcd(2^{\kappa-1}, 2^m - 1) = 1$, hence, if $gcd(r, 2^m - 1) = 1$ then $gcd(G_A(K), M) = 1$.

Thus, in M cycles of register **A**, register **B** cycles $G_A(K)$ times and the period of $(B_{G_A(t)})$ is $KM = 2^\kappa(2^m - 1)$. ■

Lemma 2 *If* $gcd(s, 2^n - 1) = 1$, *then the period* P_Q *of the sequence* $(C_{Q_A(t)})$ *is* $2^\kappa(2^n - 1)$.

Proof. Similar to the proof of the above lemma. ■

Definition 3 *The* linear complexity *of a purely periodic sequence is equal to the degree of its minimal polynomial. The minimal polynomial is the characteristic feedback polynomial of the shortest LFSR that can produce the given sequence.*

In the following two lemmas, the minimal polynomials of $(B_{G_A(t)})$ and $(C_{Q_A(t)})$ are considered.

Lemma 4 *If* $gcd(r, 2^m - 1) = 1$, *then the minimal polynomial of the sequence* $(B_{G_A(t)})$ *is of the form* $I(x)^\alpha$ *where* $2^{\kappa-1} < \alpha \leq 2^\kappa$ *and* $I(x)$ *is an irreducible polynomial of degree* m. *In particular, the linear complexity of* $(B_{G_A(t)})$ *is* L_1 *such that:* $m2^{\kappa-1} < L_1 \leq m2^\kappa$.

Proof. First, recall that if $gcd(r, 2^m - 1) = 1$ then $gcd(G_A(K), M) = gcd(r2^\kappa - 1, 2^m - 1) = 1$.

Upper Bound on L_l: If one starts at location i in the sequence $(B_{G_A(t)})$ for a fixed value of i with $0 \leq i < K$ and chooses every K^{th} element in the sequence $(B_{G_A(t)})$, then this is equivalent to starting at position $t = G_A(i)$ in (B_t) and choosing every $G_A(K)^{th}$ element. Such a sequence is a $G_A(K)$-decimation of (B_t). All the $G_A(K)$-decimation of (B_t) have the same minimal polynomial $I(x)$ whose roots are the $G_A(K)^{th}$ powers of the roots of $g(x)$ [5]. The final sequence $(B_{G_A(t)})$ consists of K such sequences interleaved. (In other words, if $(B_{G_A(t)})$ is written by rows into an array K columns wide, then each column is a sequence produced by $I(x)$). Hence, the sequence $(B_{G_A(t)})$ may be produced by an LFSR constructed as follows [6].

Take an LFSR with feedback polynomial $I(x)$ and replace each delay by a chain of K delays and only the left most of each such group of K delays is tapped and input to the feedback function with a non-zero feedback coefficient. Thus, $(B_{G_A(t)})$ is produced by an LFSR with the feedback polynomial $I(x^K)$. Hence, the minimal polynomial of $(B_{G_A(t)})$ divides $I(x^K) = I(x^{2^\kappa}) = I(x)^{2^\kappa}$. Hence, $(B_{G_A(t)})$ has linear complexity L_1 bounded from above by $mK = m2^\kappa$.

Furthermore, Chambers [6] has shown that, if $g(x)$ is irreducible, with degree m and exponent M and $gcd(G_A(K), M) = 1$, then the polynomial $I(x)$, like $g(x)$ is irreducible of degree m and exponent M.

Lower Bound on L_1**:** Let $Q(x)$ denote the minimal polynomial of $(B_{G_A(t)})$. The sequence $(B_{G_A(t)})$ satisfies $I(E)^{2^\kappa}(B_{G_A(t)}) = (0)$ for all t, where (0) is the all-zero sequence and E is the shift operator. Since the polynomial $I(x)$ is irreducible then the polynomial $Q(x)$ must be of the form $I(x)^\alpha$ for $\alpha \leq 2^\kappa$.

Assume $\alpha \leq 2^{\kappa-1}$. Then $Q(x)$ divides $I(x)^{2^{\kappa-1}}$. Since $I(x)$ is an irreducible polynomial of degree m it divides the polynomial $(1 + x^M)$. Therefore, $Q(x)$ divides $(1 + x^M)^{2^{\kappa-1}} = (1 + x^{2^{\kappa-1}M})$, but then the period of $(B_{G_A(t)})$ is at most $2^{\kappa-1}M$ [5] contradicting lemma 1. Therefore $\alpha > 2^{\kappa-1}$ and the lower bound follows. ■

Lemma 5 *If* $gcd(s, 2^n - 1) = 1$, *then the minimal polynomial of the sequence* $(C_{Q_A(t)})$ *is of the form* $J(x)^\beta$ *where* $2^{\kappa-1} < \beta \leq 2^\kappa$ *and* $J(x)$ *is an irreducible polynomial of degree* n. *In particular, the linear complexity of* $(C_{Q_A(t)})$ *is* L_2 *such that:* $n2^{\kappa-1} < L_2 \leq n2^\kappa$.

Proof. Similar to the proof of the above lemma. ■

Therefore, if $gcd(r, 2^m - 1) = gcd(s, 2^n - 1) = 1$ then the periods of $(B_{G_{A(t)}})$ and $(C_{Q_{A(t)}})$ are $P_G = 2^\kappa(2^m - 1)$ and $P_Q = 2^\kappa(2^n - 1)$ respectively and the minimal polynomials of $(B_{G_{A(t)}})$ and $(C_{Q_{A(t)}})$ are equal to $I(x)^\alpha$ and $J(x)^\beta$ respectively where $2^{\kappa-1} < \alpha, \beta \leq 2^\kappa$ and $I(x)$, $J(x)$ are irreducible polynomials of degree m and n respectively.

Theorem 6 *If m, n are positive integers greater than 1 satisfying $gcd(m, n) = 1$ and r, s satisfy $gcd(r, 2^m - 1) = gcd(s, 2^n - 1) = 1$, then the output sequence (Z_t) has period $P_Z = 2^\kappa(2^m - 1)(2^n - 1)$ and linear complexity L such that: $(m + n)2^{\kappa-1} < L \leq (m + n)2^\kappa$.*

Proof. From the above lemmas, the minimal polynomials of $(B_{G_{A(t)}})$ is $I(x)^\alpha$ and that of $(C_{Q_{A(t)}})$ is $J(x)^\beta$ where $2^{\kappa-1} < \alpha, \beta \leq 2^\kappa$. Since $I(x)$ and $J(x)$ are irreducible of different degrees then $gcd(I(x), J(x)) = 1$, hence $gcd(I(x)^\alpha, J(x)^\beta) = 1$ [5]. Therefore, the period of (Z_t) is $P_Z = lcm(P_G, P_Q)$ [5, theorem 3.9] and the minimal polynomial of (Z_t) is $I(x)^\alpha J(x)^\beta$ of degree $L = (m\alpha + n\beta)$ [5, theorem 6.57].

Hence, the period of (Z_t) is $P_Z = lcm(2^\kappa(2^m - 1), 2^\kappa(2^n - 1)) = \{[2^\kappa(2^m - 1)(2^n - 1)]/(2^{gcd(m,n)} - 1)\}$ [7, lemma 5.9]. Thus, the period of (Z_t) is $P_Z = 2^\kappa(2^m - 1)(2^n - 1)$, and the linear complexity of (Z_t) is L such that: $(m + n)2^{\kappa-1} < L \leq (m + n)2^\kappa$. ■

2.2. The Statistical Properties of (Z_t)

In this section, the number of ones and zeroes in a full period $P_Z = 2^\kappa(2^m - 1)(2^n - 1)$ of the sequence (Z_t) are counted. It also shown that when m and n are positive integers greater than 1 satisfying $gcd(m, n) = 1$ and the positive integers r and s satisfy $gcd(r, 2^m - 1) = gcd(s, 2^n - 1) = 1$, then any pattern of length $q \leq \min(\psi, \delta)$ where ψ, δ are positive integers such that $\psi = \lfloor (m-1)/r + 1 \rfloor$ and $\delta = \lfloor (n-1)/s + 1 \rfloor$ occurs with probability $2^{-q} + O(1/2^{m-q}) + O(1/2^{n-q})$, where $\lfloor W \rfloor$ is the integer part of W for any real number W.

Since (B_t) and (C_t) are m-sequences then in a full period $M = (2^m - 1)$ of (B_t) the number of ones and zeroes is $M_1 = 2^{m-1}$ and $M_0 = (2^{m-1} - 1)$ respectively, and in a full period $N = (2^n - 1)$ of (C_t) the number of ones and zeroes is $N_1 = 2^{n-1}$ and $N_0 = (2^{n-1} - 1)$ respectively [2].

If the period of (Z_t) attains its maximum value $P_Z = 2^\kappa(2^m - 1)(2^n - 1)$, then it is obvious that the number of ones and zeroes in a full period of (Z_t) is $2^\kappa[(2^m - 1)2^{n-1} - 2^{m-1}]$ and $2^\kappa[(2^m - 1)2^{n-1} - (2^{m-1} - 1)]$ respectively.

In the following theorem, similar techniques to the ones used by Gunther [8] are applied to determine the distribution of short patterns in the output sequences of the ASG(r, s).

Theorem 7 *Let m, n be positive integers greater than 1 satisfying $gcd(m, n) = 1$ and let r, s satisfy $gcd(r, 2^m - 1) = gcd(s, 2^n - 1) = 1$. Let ψ and δ be positive integers such that $\psi = \lfloor (m - 1)/r + 1 \rfloor$ and $\delta = \lfloor (n - 1)/s + 1 \rfloor$.*

The probability of occurrence of any pattern $\sigma = (\sigma_0, \sigma_1, ..., \sigma_{q-1}) \in \{0, 1\}^q$ of length $q \leq \min(\psi, \delta)$ in the sequence (Z_t) is 2^{-q} up to an error of order $O(1/2^{m-q}) + O(1/2^{n-q})$.

Proof. The proof is given in the appendix. ■

Clearly, the smaller the values for r and s compared to m and n are, the better the above result is. This does not mean that it is suggested to take r and s to be very small, for example $r = s = 1$. For more security it is better to irregularly clock the generating registers by large values, so that the gap between the bits selected from the generating sequences is large.

Experiments have shown that if $gcd(m, n) = 1$, then for any values of r and s satisfying $gcd(r, 2^m - 1) = gcd(s, 2^n - 1) = 1$, the output sequences of the ASG(r, s) have good statistical properties.

Therefore, when m and n are positive integers greater than 1 satisfying $gcd(m, n) = 1$ and r, s satisfy $gcd(r, 2^m - 1) = gcd(s, 2^n - 1) = 1$, then an ASG($r, s$) with a de Bruijn sequence as the control sequence and m-sequences as the generating sequences generates sequences with period $P_Z = 2^\kappa(2^m - 1)(2^n - 1)$, linear complexity L such that $(m+n)2^{\kappa-1} < L \leq (m+n)2^\kappa$, and these sequences have good statistical properties.

In practice, one can choose r and s to be powers of 2 in which case $gcd(r, 2^m - 1) = gcd(s, 2^n - 1) = 1$.

In the following section, some correlation attacks on the ASG(r, s) are considered.

3. Attacks

A suitable stream cipher should be resistant against a "known-plaintext" attack. In a known-plaintext attack the cryptanalyst is given a plaintext and the corresponding cipher-text (in another word, the cryptanalyst is given a keystream), and the task is to reproduce the keystream somehow.

The most important general attacks on LFSR-based stream ciphers are correlation attacks. Basically, if a cryptanalyst can in some way detect a correlation between the known output sequence and the output of one individual LFSR, this can be used in a divide and conquer attack on the individual LFSR [9, 10, 11, 12].

The output sequence of the ASG(r, s) is an addition modulo 2 of its two irregularly decimated generating sequences $(B_{G_A(t)})$ and $(C_{Q_A(t)})$. Thus, one would not expect a strong correlation to be obtained efficiently, especially, if primitive feedback polynomials of high Hamming weight are associated with the feedback functions of the registers **B** and **C** [11], and the values of r and s which are used to clock the generating registers are considered as part of the key (i.e., r and s are kept secret).

If the characteristic feedback functions of **A**, **B** and **C** are known then a cryptanalyst can exhaustively search for initial state of **A**, each such state can be expanded to a prefix of the control sequence (A_t) using the characteristic feedback function of **A**. Suppose that one expands the sequence (A_t) until its p^{th} 1 and 0 are produced where $p = \max(m, n)$. From this prefix, and from the knowledge of a corresponding p-long prefix of the output sequence of (Z_t), one can derive the value of p non-consecutive bits of the generating sequences (B_t) and (C_t) using the following relation:

$$Z_t \oplus Z_{t+1} = \begin{cases} B_{G_A(t)} \oplus B_{G_A(t+1)} & \text{if } A_t = 1, \\ C_{Q_A(t)} \oplus C_{Q_A(t+1)} & \text{if } A_t = 0. \end{cases}$$

Since the characteristic feedback functions of **B** and **C** are known, then the initial states of **B** and **C** can be revealed given these non-consecutive p-bits of (B_t) and (C_t) respectively by solving a system of linear equations, but first one has to reveal the values of r and s in order to determine the locations of these non consecutive p-bits in (B_t) and (C_t). Therefore, the attack takes approximately $O(\Phi 2^\kappa m^3 n^3)$ steps where $\Phi = \Phi_1\Phi_2$, Φ_1 is the number of possible values for r such that $gcd(r, 2^m - 1) = 1$ and Φ_2 is the number of possible values for s such that $gcd(s, 2^n - 1) = 1$.

The probability of two random numbers being relatively primes is 60.8% [13]. Thus, $\Phi_1 \approx 2^{m-1}$ and $\Phi_2 \approx 2^{n-1}$, and the above attack takes approximately $O(2^{\kappa+m+n-2}m^3n^3)$ steps.

For $\kappa \approx 64$, $m \approx 64$ and $n \approx 64$, the attack takes approximately $O(2^{226})$ steps. Thus, this ASG(r, s) appears to be secure against this attack. Moreover, it appears to be secure against all correlation attacks introduced in [9, 10,11,12,14, 15, 16, 17,18,19].

There is also another attack that can be applied to the ASG(r, s) through the linear complexity, but this attack requires $(m+n)2^\kappa$ consecutive bits of the output sequence.

For maximum security, the ASG(r, s) should be used with secret initial states, secret characteristic feedback functions, secret r, s satisfying $gcd(r, 2^m - 1) = gcd(s, 2^n - 1) = 1$, and m, n greater than 1 satisfying $gcd(m, n) = 1$. Subject to these constraints, an ASG(r, s) with $\kappa \approx 64$, $m \approx 64$ and $n \approx 64$ appears to be secure against all presently known attacks.

4. Related Work

An interesting example of existing FSR-based construction for comparison with the ASG(r, s) is the *Alternating Step Generator* (ASG) of Gunther [8].

The ASG is a special case of the ASG(r, s); it is actually an ASG($1, 1$). Although the ASG(r, s) is slower than the ASG, its advantage is that it provides more security. For an ASG with $\kappa \approx l$, $m \approx l$ and $n \approx l$, if the characteristic feedback functions of **A**, **B** and **C** are known, then in order to reveal the initial states of the three registers the attack mentioned in section 3 takes approximately $O(2^l l^6)$ steps, whereas for the ASG(r, s), the attack takes approximately $O(\Phi 2^l l^6)$ steps. Moreover, for the ASG in order to produce a new sequence, one has to choose a new initial state and/or a new characteristic feedback function for at least one of the FSRs, whereas for the ASG(r, s) in order to produce a new sequence, it suffices to assign new value(s) for r and/or s.

5. Conclusion

From the theoretical results established, it is concluded that an ASG(r, s) whose control FSR generates a de Bruijn sequence and generating FSRs generate m-sequences produces sequences with large periods, high linear complexities, good statistical properties, and they are secure against correlation attacks. Furthermore, using the same initial states and the same characteristic feedback functions, the ASG(r, s) produces a new sequence each time one assigns new value(s) for r and/or s. These characteristics and properties enhance its use as a suitable crypto-generator for stream cipher applications.

Acknowledgement

I would like to thank Prof. Peter Wild for his helpful comments.

References

[1] D. Gollmann and W. Chambers. Clock-controlled shift register: A review. *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 4, May 1989, pp. 525–533.

[2] S. W. Golomb. *Shift Register Sequences*. Aegean Park Press, 1982.

[3] Steven A. Tretter. Properties of PN2 sequences. *IEEE Transactions on Information Theory*, vol. IT-20, March 1974, pp. 295–297.

[4] T. Beth and F. Piper. The stop and go generator. In *Advances in Cryptology: Proceedings of EuroCrypt 84*, Lecture Notes in Computer Science, Berlin: Springer-Verlag 1985, vol. 209, pp. 88–92.

[5] R. Lidl, H. Niederreiter. *Introduction to Finite Fields and Their Applications*. Cambridge University Press, 1986.

[6] W. Chambers. Clock-controlled shift registers in binary sequence generators. *IEE Proceedings E*, vol. 135, Jan 1988, pp. 17–24.

[7] R. Ruppell. *Analysis and Design of Stream Ciphers*. Berlin, Heidelberg, New York: Springer-Verlag, 1986.

[8] C. G. Gunther. Alternating step generators controlled by de Bruijn sequences. In *Advances in Cryptology: Proceedings of EuroCrypt 87*, Lecture Notes in Computer Science, Berlin: Springer-Verlag, vol. 309, 1988, pp. 5–14.

[9] J. Golic, M. Mihaljevic. A generalized correlation attack on a class of stream ciphers based on the Levenstein distance. *Journal of Cryptology*, vol. 3, 1991, pp. 201–212.

[10] J.Golic. Towards fast correlation attacks on irregularly clocked shift registers. In *Advances in Cryptology: Proceedings of EuroCrypt 95*, Lecture Notes in Computer Science, Berlin: Springer-Verlag, vol. 921, 1995, pp. 248–262.

[11] W. Meir, O. Staffelbach. Fast correlation attacks on certain stream ciphers. *Journal of Cryptology*, vol. 1, 1989, pp. 159–176.

[12] T. Siegenthaler. Correlation-immunity of non-linear combining functions for cryptographic applications. *IEEE Transactions On Information Theory*, vol. IT-30, no. 5, 1984, pp.776–779.

[13] `www.utm.edu/research/primes/notes/relprimr.html`.

[14] J. Golic. On the security of shift register based keystream generators. In R. Anderson, Editor, *Fast Software Encryption, Cambridge Security Workshop*, Lecture Notes in Computer Science, Berlin: Springer-Verlag, vol. 809, 1994, pp. 90–100.

[15] T. Johansson. Reduced complexity correlation attacks on two clock-controlled generators. In *Advances of Cryptology: Proceedings of AsiaCrypt 98*, Berlin: Lecture Notes in Computer Science, vol. 1514, 1998, pp. 342–356.

[16] M. Mihaljevic. An approach to the initial state reconstruction of a clock-controlled shift register based on a novel distance measure. In *Advances in Cryptology: AusCrypt 92*, Berlin: Lecture Notes in Computer Science, vol. 178, 1993, pp. 349–356.

[17] J. Golic, L. O'Connor. Embedding probabilistic correlation attacks on clock-controlled shift registers. In *Advances in Cryptology: EuroCrypt 94*, Berlin: Lecture Notes in Computer Science, vol. 950, 1995, pp. 230–243.

[18] T. Johansson, F.Jonsson. Improved fast correlation attacks on certain stream ciphers via convolutional codes. In *Advances in Cryptology: EuroCrypt 99*, Berlin: Lecture Notes in Computer Science, vol. 1592, Springer-Verlag, 1999, pp. 347–362.

[19] T. Johansson, F.Jonsson. Fast correlation attacks through reconstruction of linear polynomials. In *Advances in Cryptology: Crypto 2000*, Berlin: Lecture Notes in Computer Science, vol. 1880, Springer-Verlag, 2000, pp. 300–315.

A. Proof of Theorem 7

Proof. Since $gcd(m, n) = 1$ and $gcd(r, 2^m - 1) = gcd(s, 2^n - 1) = 1$, then the period of (Z_t) $P_Z = 2^\kappa(2^m - 1)(2^n - 1)$.

Let $t \in \{0, 1, .., P_{Z-1}\}$ be represented in the form $t = u + (v + yM)2^\kappa$, $u \in \{0, 1, .., K - 1\}$, $v \in \{0, 1, .., M-1\}$, $y \in \{0, 1, .., N-1\}$ and let us first consider the frequency of patterns among subsequences $Z_t, Z_{t+1}, ..., Z_{t+q-1}$ for a fixed $u \in \{0, 1, ..., K - 1\}$.

Let $\rho = \rho(u)$ and $\theta = \theta(u)$ be defined by:

$$\begin{aligned}
\rho_0 &= 0 \\
\theta_0 &= \sigma_0 \\
\rho_{i+1} &= \rho_i \oplus A_{u+i}(\sigma_{i+1} \oplus \sigma_i) \\
\theta_{i+1} &= \theta_i \oplus (1 \oplus A_{u+i})(\sigma_{i+1} \oplus \sigma_i)
\end{aligned} \tag{1}$$

for $i \in \{0, 1, ..., q-2\}$.

Then σ can be written as

$$\sigma_i = \rho_i \oplus \theta_i \tag{2}$$

for $i \in \{0, 1, ..., q-1\}$.

The matching condition at time t is:

$$Z_{t+i} = \sigma_i \tag{3}$$

for $i \in \{0, 1, ..., q-1\}$.

This is equivalent to:

$$B_{G_A(t+i)} \oplus C_{Q_{A(t+i)}} = \rho_i \oplus \theta_i \tag{4}$$

for $i \in \{0, 1, ..., q-1\}$.

Using the following relations:

$$\begin{aligned} G_A(u+i+1) &= G_A(u+i) + rA_i \\ Q_A(u+i+1) &= Q_A(u+i) + s(A_i \oplus 1) \end{aligned} \tag{5}$$

the sum of Equation (4) and of the corresponding equation for $(i+1)$ becomes:

$$\begin{aligned} B_{G_A(t+i+1)} \oplus B_{G_A(t+i)} &= \rho_{i+1} \oplus \rho_i \\ C_{Q_A(t+i+1)} \oplus C_{Q_A(t+i)} &= \theta_{i+1} \oplus \theta_i \end{aligned} \tag{6}$$

since, when $A_i = 1$, $\theta_{i+1} \oplus \theta_i = C_{Q_A(t+i+1)} \oplus C_{Q_A(t+i)} = 0$, and when $A_i = 0$, $\rho_{i+1} \oplus \rho_i = B_{G_A(t+i+1)} \oplus B_{G_A(t+i)} = 0$.

This has two solutions:

$$\begin{aligned} B_{G_A(t+i)} &= \rho_i \\ C_{Q_A(t+i)} &= \theta_i \end{aligned} \tag{7}$$

and

$$\begin{aligned} B_{G_A(t+i)} &= 1 \oplus \rho_i \\ C_{Q_A(t+i)} &= 1 \oplus \theta_i \end{aligned} \tag{8}$$

for $i \in \{0, 1, ..., q-1\}$.

The number of solutions to this equation is equal to the number of occurrences of the pattern σ in the sequence (Z_t) (where $t = u + (v + yM)2^\kappa$, $v \in \{0, 1, .., M-1\}$, $y \in \{0, 1, .., N-1\}$), i.e., to the quantity we want to determine.

Without restricting ourselves we consider the solution of Equation (7). Making use of the fact that $K = 2^\kappa$ and that $G_A(K) = r2^{\kappa-1}$ and $Q_A(K) = s2^{\kappa-1}$, this equation becomes:

$$B_{G_A(u+i)+vr2^{\kappa-1}} = \rho_i \tag{9}$$

where $i \in \{0, 1, .., q-1\}$. (The term yM is omitted since (B_t) has period M.)

$$C_{Q_A(u+i)+(v+yM)s2^{\kappa-1}} = \theta_i \tag{10}$$

Let $\varphi(u) = [G_A(u+q-1) - G_A(u)]/r$ which is less than m since $q \leq \min(\psi, \delta)$ where $\psi = \lfloor (m-1)/r + 1 \rfloor$ and $\delta = \lfloor (n-1)/s + 1 \rfloor$, then the assumptions that (B_t) is an m-sequence imply that Equation (9) has $2^{m-\varphi(u)-1}$ solutions if $\rho \neq 0$.

Let $\varkappa(u) = [Q_A(u+q-1) - Q_A(u)]/s$, then similarly, (C_t) is an m-sequence and $gcd(m, n) = 1$ imply that Equation (10) has $2^{n-\varkappa(u)-1}$ solutions if $\theta \neq 0$.

This remains true for $\rho = 0$ and/or $\theta = 0$ if we accept an error at most $O(1/2^{m-q}) + O(1/2^{n-q})$. Note that $\varphi(u) + \varkappa(u) = (q-1)$.

Clearly, the same result also holds for Equation (8).

Hence, the total number of solutions to Equation (3) is:

$$2(2^{m-\varphi(u)-1})(2^{n-\varkappa(u)-1}) = 2^{m+n-q}$$

which is independent of u.

This finally implies that the frequency of the pattern σ is given by:

$$\frac{2^{m+n-q}}{MN} + O(1/2^{m-q}) + O(1/2^{n-q})$$

Therefore, in a full period of (Z_t) any pattern of length $q \leq \min(\psi, \delta)$ occurs with a probability $(1/2^q) + O(1/2^{m-q}) + O(1/2^{n-q})$. ■