The (a, b)-Shrinking Generator

Ali Adel Kanso

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

Abstract: A new construction of a pseudorandom generator based on a simple combination of two LFSRs is introduced. This construction allows users to generate a large family of sequences using the same initial states and the same characteristic feedback polynomials of the two combined LFSRs. The construction is related to the so-called shrinking generator that is a special case of this construction. The construction has attractive properties such as exponential period, exponential linear complexity, good statistical properties and security against correlation attacks. All these properties make it a suitable crypto-generator for stream cipher applications.

Keywords: Linear Feedback Shift Registers, Stream Ciphers, Clock-Controlled Registers, and Shrinking Generator.

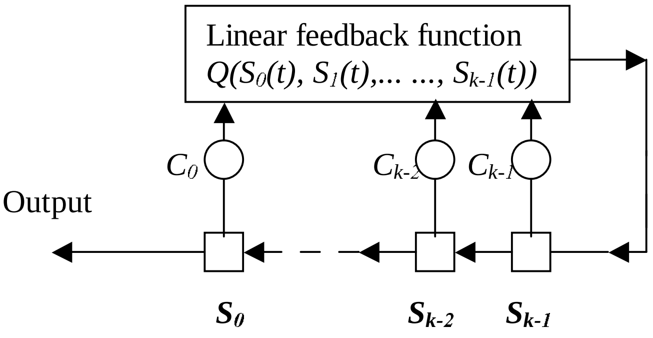
1 Introduction

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

An LFSR is made up of two parts: a shift register *S*, and a linear feedback function *Q*. The shift register *S* consists of *k* stages $\boldsymbol{S_0}$, $\boldsymbol{S_1}$, ..., $\boldsymbol{S_{k-1}}$ 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 linear feedback function *Q* is a function that maps the state of the register *S* to the bit 0 or 1 and it is of the form $Q(S_0(t), ..., S_{k-1}(t)) = (C_0S_0(t) \oplus .... \oplus C_{k-1}S_{k-1}(t))$ for some binary constants $C_0, C_1, ..., C_{k-1}$ called the feedback coefficients.

Figure 1. An k-stage LFSR



The feedback coefficients $C_0, C_1, C_2, C_3, ..., C_{k-2}, C_{k-1}$ determine a polynomial $C_0 \oplus C_1 x^1 \oplus C_2 x^2 \oplus .... \oplus C_{k-1} x^{k-1} \oplus x^k$ of degree k associated with the feedback function Q. We write $h(x)$ to denote this polynomial and call it *the characteristic feedback polynomial* of the linear feedback shift register.

Therefore, any k-stage LFSR can be uniquely described by a characteristic polynomial $h(x)$ over the finite field of order 2 of the form: $h(x) = C_0 \oplus C_1x^1 \oplus ... \oplus C_{k-1} x^{k-1} \oplus x^k$.

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 $\boldsymbol{S}_i$ is transferred into $\boldsymbol{S}_{i-1}$ $(i = 1, 2, .., k-1)$) and the new content of $\boldsymbol{S}_{k-1}$ is computed by applying the feedback function Q to the old contents of S.

The above can be expressed as follows:
$S_i(t + 1) = S_{i+1}(t)$ *for* $i = 0, 1, ..., k-2.$
$S_{k-1}(t + 1) = Q(S_0(t), ..., S_{k-1}(t)).$

The binary sequence (S_t) generated by this device is the sequence of contents of the 0th stage $\boldsymbol{S}_0$ 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 $\boldsymbol{S}$: $(\underline{S}_t) = \underline{S}_0, \underline{S}_1, \underline{S}_2, ...$ [Where $\underline{S}_t = S_0(t), S_1(t), ..., S_{k-1}(t)$ for $t = 0, 1, 2, ...$].

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

1.1 Construction

Linear feedback shift registers (LFSRs) are known to produce sequences with large period and good statistical properties. But inherent linearity of these sequences results in susceptibility to algebraic attacks that is the prime reason why LFSRs are not used directly for keystream generation. A well-known method for increasing the linear complexity preserving at the same time a large period and good statistical properties, is to control the clocking of the LFSR [1]. On the other hand, keystream generators based on regularly clocked LFSRs are susceptible to basic correlation attacks. Using irregular clocking reduces the danger from correlation attacks and provides practical immunity to fast correlation attacks.

In this paper, a new clock-controlled generator that is called the *(a, b)*-Shrinking Generator (and referred to as *(a, b)-SHKG*) is introduced. The *(a, b)-SHKG* is a sequence generator composed of two linear feedback shift registers (LFSRs) [2], LFSR **A** and LFSR **B**; the first is clocked normally, but the second is clocked by the constant integer *"a"* if the content of the 0th stage of LFSR **A** is 1, otherwise, it is clocked by the constant integer *"b"*. LFSRs **A** and **B** are called "the control register" and "the generating register" of the *(a, b)-SHKG* respectively. The output bits of the *(a, b)-SHKG* are produced by shrinking the output of LFSR **B** under the control of

LFSR **A** as follows: At any time t, the output of LFSR **B** is taken if the current output of LFSR **A** is 1, otherwise it is discarded.

Suppose that LFSR **A** has m stages and characteristic feedback polynomial $f(x)$. Similarly, suppose that LFSR **B** has n stages and characteristic feedback polynomial $g(x)$. Let $\underline{A}_0 = A_0(0), A_1(0), ..., A_{m-1}(0)$ and $\underline{B}_0 = B_0(0), B_1(0), ..., B_{n-1}(0)$ be the non-zero initial states of **A** and **B** respectively. At time t, LFSR **A** is clocked once. LFSR **B** is clocked X_t times, where: $X_t = aA_0(t) + b[A_0(t) \oplus 1]$. {Where ⊕ denotes addition modulo 2.}

Define the "cumulative" function of **A** to be G_A: {0, 1, 2, …} → {0, 1, 2, …} where: $G_A(t) = \sum_{i=o}^{t-1} X_i$, for $t > 0$, and $G_A(0) = 0$.

The state of the *(a, b)-SHKG* at time t is given by: $\underline{S}_t = (\underline{A}_t, \underline{B}_{G_A(t)})$.

At any time t, the output of the *(a, b)-SHKG* of initial state $\underline{S}_0 = (\underline{A}_0, \underline{B}_0)$ is the content of the 0th stage of LFSR **B** [i.e. $B_0(G_A(t))$] if $A_0(t)$ is 1, otherwise there is no output.

*The (a, b)-SHKG may also be described in terms of the two output sequences (A_t) and (B_t) of the linear feedback shift registers **A** and **B** respectively.*

Acting on their own, suppose that LFSR **A** and LFSR **B** produce output sequences $(A_t) = A_o, A_1, ...$ and $(B_t) = B_0, B_1, ...$ respectively. The sequences (A_t) and (B_t) are called "the control sequence" and "the generating sequence" of the *(a, b)-SHKG* respectively and referred to these as component sequences.

Define a function Ge: {0, 1, 2, …} → {0, 1, 2, …} as follows: Let $t' \geq 0$ and suppose that $A_{t'} = 1$. Let t be the total number of ones in $A_0, A_1, ..., A_{t'}$ then $Ge(t - 1) = G_A(t')$.

The output sequence (Z_t) of the *(a, b)-SHKG* whose control sequence and generating sequence are (A_t) and (B_t) respectively is given by: $Z_t = B_{Ge(t)}$.

2 Properties of the Output Sequence *(Z_t)* of the *(a, b)-SHKG*

In this section some properties of the output sequences of the *(a, b)-SHKG* are established. Suppose that LFSRs **A** and **B** have initial states $\underline{A}_0$ and $\underline{B}_0$ respectively, and characteristic feedback polynomials $f(x)$ and $g(x)$ respectively. Let (A_t) and (B_t) denote the output sequences of **A** and **B** respectively. Suppose that (A_t) and (B_t) are periodic of periods M and N respectively. Let (Z_t) be the output sequence of this *(a, b)-SHKG.*

Let M_1 be the total number of ones in a full period M of (A_t). [i.e. M_1 is the total number of ones that appears in the 0th stage of **A** in a full period M.]

In the following lemmas, the period and the linear complexity of the output sequences are established. Finally, it is shown that output sequences of the *(a, b)-SHKG* have good statistical properties.

2.1 Period and Linear Complexity of (Z_t)

After $M[lcm(G_A(M), N)/G_A(M)]$ clock pulses have been applied to LFSR **A**, $lcm(G_A(M), N)$ clock pulses will have been applied to LFSR **B**. Then both **A** and **B** return to their initial states $\underline{A}_0$ and $\underline{B}_0$. Hence the state sequence of the generator is periodic with period dividing $M[lcm(G_A(M), N)/G_A(M)] = MN/gcd(G_A(M), N)$.

The *(a, b)-SHKG* produces an output whenever the 0th stage of **A** contains a 1. Therefore, after $lcm(G_A(M), N)$ clock pulses have been applied to **B** the *(a, b)-SHKG* produces $M_1N/gcd(G_A(M), N)$ output bits. Hence, the period P_Z of the sequence (Z_t) is a divisor of $M_1N/gcd(G_A(M), N)$.

In the following lemma, it is shown that the maximum period of the output sequence (Z_t) is attained under simple conditions.

Lemma 1

*Let c be a non-negative integer such that $c = max(a, b)$. If the length m of LFSR **A** satisfies $m < N/c$, and $gcd(G_A(M), N) = 1$ then the period P_Z of the sequence (Z_t) is equal to M_1N.*

Proof. The proof of this lemma is given in the appendix.

Definition 1

The linear complexity L of a periodic sequence (Z_t) is equal to the degree of its minimal polynomial. The minimal polynomial is defined as the characteristic feedback polynomial of the shortest LFSR that can generate the sequence (Z_t).

The following lemma establishes that if (B_t) is an m-sequence [2] (i.e. **B** is a primitive *n*-stage LFSR) then under simple conditions the sequence (Z_t) has gap by a factor of 2 between the lower and upper bound of its linear complexity.

Lemma 2

Suppose that (B_t) is an m-sequence of period $N = (2^n - 1)$ and the number of ones in a full period of (A_t) is a power of 2 i.e. $M_1 = 2^r$ for some r. Let c be a non-negative integer such that $c = max(a, b)$.

*If the length m of LFSR **A** satisfies $m < N/c$, and $gcd(G_A(M), N) = 1$ then (Z_t) has linear complexity L such that: $n2^{r-1} < L \leq n2^r$.*

Proof.

Upper Bound on *L*: To show an upper bound on the linear complexity of (Z_t) it suffices to present a polynomial $P(x)$ of degree *d* (for some positive integer *d*) for which the coefficients of P represent a linear relation satisfied by the elements of (Z_t) [3]. That is,

$$P(x) = \sum_{i=0}^{d} P_i x^i \text{ then } \sum_{i=0}^{d} P_i Z_{i+t} = 0 \quad \forall t \geq 0.$$

For $0 \leq k < M_1$ let $(Z_{k+t}{}^{M_1})$ denote the k^{th} translate of the sequence (Z_t) decimated by M_1 (i.e. $Z_{k+j}{}^{M_1} = Z_{k+jM_1}$, for $j = 0, 1, \ldots.$). Fact 1 in the proof of (lemma 1) states that this translate and decimation, written in terms of the sequence (B_t) is $Z_{k+jM_1} = B_{Ge(k)+jG_A(M)}$ i.e. $(Z_{k+t}{}^{M_1})$ is a translate of the sequence (B_t) decimated by $G_A(M)$. Since $gcd(G_A(M), N) = 1$ and (B_t) is an m-sequence, each sequence $(Z_{k+t}{}^{M_1})$ has the same linear complexity as the original sequence (B_t), and it satisfies a polynomial $Q(x)$ of degree n [3] i.e. $Q(E)(Z_{k+t}{}^{M_1}) = Q(E)(Z_{k+tM_1}) = (0)$ $\forall t$, where E is the shift operator and (0) is the all-zero sequence of length N.

Let $Q(x) = \sum_{i=0}^{n} Q_i x^i$ then for each $k = 0, 1, 2, \ldots., (M_1 - 1)$, $\sum_{i=0}^{n} Q_i Z_{k+(t+i)M_1} = 0$ $\forall t.$

Then in terms of the bits of (Z_t) one can write $\sum_{h=0}^{nM_1} P_h Z_{t+h} = 0$ $\forall t$, where $P_h = 0$ when $h \pmod{M_1} \neq 0$, and $P_{iM_1} = Q_i$ for $i = 0, 1, \ldots, n.$

Hence, a linear recurrence for the sequence (Z_t) is found. Therefore, a polynomial $P(x) = Q(x)^{M_1}$ of degree nM_1 is found, such that $P(E)(Z_t) = Q(E)^{M_1}(Z_t) = (0)$, and then the linear complexity of (Z_t) is at most nM_1.

Lower Bound on *L*: Let $R(x)$ denote the minimal polynomial of (Z_t). The sequence (Z_t) satisfies $Q(E)^{M_1}(Z_t) = (0)$, where M_1 is a power of 2. Since the polynomial $Q(x)$ is irreducible then the polynomial $R(x)$ must be of the form $Q(x)^q$ for $q \leq 2^r$.

Assume $q \leq 2^{r-1}$. Then $R(x)$ divides $Q(x)^{2^{r-1}}$. Since $Q(x)$ is an irreducible polynomial of degree n it divides the polynomial $(1 + x^N)$. Therefore, $R(x)$ divides $(1 + x^N)^{2^{r-1}} = (1 + x^{2^{r-1}N})$, but then the period of (Z_t) is at most $2^{r-1}N$ [3] contradicting (lemma 1). Therefore, $q > 2^{r-1}$ and the lower bound follows.

2.2 The Statistical Properties of (Z_t)

In this section, the number of ones and zeroes in a full period $P_Z = M_1N$ of the sequence (Z_t) are counted. It is also discussed that when the initial state and the primitive characteristic feedback polynomial of the n-stage linear feedback shift register **B** of the (a, b)-*SHKG* are chosen with uniform probability over all non-zero initial states of length n and among all primitive polynomials of degree n respectively, then a collection of the output sequences of the (a, b)-*SHKG*'s has good statistical properties.

Let M_1 be the total number of ones that appears in the 0th stage of **A** in a full period M. Let N_1 and N_0 be the total number of ones and zeroes respectively that appears in the 0th stage of **B** in a full period N. If the period of (Z_t) attains its maximum value $P_Z = M_1N$, then it is obvious that the number of ones in a full period of (Z_t) is M_1N_1, and the number of zeroes is M_1N_0.

Next, consider some sample spaces that are used in the following theorem.

Let Γ_1^A be the sample space of all characteristic feedback polynomials of LFSR **A**.
Let Γ_2^A be the sample space of all non-zero initial states of LFSR **A**.
Let Ω_1^B be the sample space of all primitive characteristic feedback polynomials of LFSR **B**.
Let Ω_2^B be the sample space of all non-zero initial states of LFSR **B**.

Define the sample spaces Γ^A and Ω^B such that $\Gamma^A = \Gamma_1^A \times \Gamma_2^A = \{(f(x), \underline{A}_0) \mid f(x) \in \Gamma_1^A, \underline{A}_0 \in \Gamma_2^A\}$, and $\Omega^B = \Omega_1^B \times \Omega_2^B = \{(g(x), \underline{B}_0) \mid g(x) \in \Omega_1^B, \underline{B}_0 \in \Omega_2^B\}$.

Coppersmith et al [4] have stated that if the initial state and the primitive characteristic feedback polynomial of an LFSR of length n are chosen with uniform probability over all non-zero initial states of length n and among all primitive polynomials of degree n respectively, then the distribution of any non-consecutive k bits produced by this LFSR is almost uniform. Since the output sequence of an *(a, b)-SHKG* is just some non-consecutive bits of LFSR **B** selected according to the control register LFSR **A**, then the output sequences of the *(a, b)-SHKG* whose generating register LFSR **B** has initial state and primitive characteristic feedback polynomial chosen with uniform probability over all non-zero initial states of length n and among all primitive polynomials of degree n respectively, inherit these properties.

The following theorem established in [5] states that the distribution of patterns in the outputs of a collection of *(a, b)-SHKG*'s is almost uniform.

Theorem 1

*Let (Z_t) denote the output sequence of an (a, b)-SHKG whose control register LFSR **A** is the m-stage linear feedback shift register with non-zero initial state $\underline{A}_0$ and characteristic feedback polynomial f(x), and whose generating register LFSR **B** is the n-stage linear feedback shift register with non-zero initial state $\underline{B}_0$ and primitive characteristic feedback polynomial g(x). Let the distribution on Ω^B be uniform [i.e. $P(g(x), \underline{B}_0)) = 1/|\Omega^B|$, for all $(g(x), \underline{B}_0) \in \Omega^B$]. Let c be a non-negative integer such that $c = \max(a, b)$. Let k be a positive integer satisfying $m(k-1)c < (2^n - 1)$ i.e. $k < [(2^n - 1)/(m\,c) + 1]$. Let t_0 be a positive integer and let R_k be the $\mathbf{Z}^k{}_2$-valued random variable on $\Gamma^A \times \Omega^B$ that maps the elementary event $(f(x), \underline{A}_0, g(x), \underline{B}_0)$ to the k consecutive output bits of (Z_t) beginning at t_0 i.e. $R_k(f(x), \underline{A}_0, g(x), \underline{B}_0) = Z_{t_0}, Z_{t_0+1}, \ldots, Z_{t_0+k-1}$. Let ϑ be any binary pattern of k bits. The probability that $R_k = \vartheta$ is in the range $2^{-k} \pm [(m(k-1)c) + 1]/2^n$.*

From the above theorem, any pattern of length k occurs with probability in the range $2^{-k} \pm [(m(k-1)c) + 1]/2^n$ among any of the $|\Gamma^A| \times |\Omega^B|$ k-tuples consisting of a specified set of k consecutive output bits of *(a, b)-SHKG*'s satisfying the conditions of theorem 1.

Clearly, the smaller the numbers *"a"* and *"b"* compared to n are, the better the above result is. This does not mean that it is suggested to take *"a"* and *"b"* to be very small, for example *"a" = "b" = 1*. For more security it is better to irregularly clock the generating register LFSR **B** by large values, so that the gap between the bits selected from the output of LFSR **B** is large.

In the next section, an *(a, b)-SHKG* with primitive LFSRs is considered.

3 Applications

*Suppose that **A** and **B** are primitive m and n stages LFSRs respectively with period $M = (2^m - 1)$ and $N = (2^n - 1)$ respectively. Let (Z_t) denote the resulting output sequence of the (a, b)-SHKG whose control and generating registers are **A** and **B** respectively.*

Since **A** is a primitive m-stage LFSR of period $M = (2^m - 1)$, then $G_A(M) = a2^{m-1} + b(2^{m-1} - 1)$.

If $gcd(a2^{m-1} + b(2^{m-1} - 1), 2^n - 1) = 1$ and $m < (2^n - 1)/c$, then by lemmas 1 and 2 (Z_t) has period $P_Z = 2^{m-1}(2^n - 1)$, and linear complexity L such that: $n2^{m-2} < L \leq n2^{m-1}$.

From the discussion and the theorem of section 2.2, the number of ones and zeroes in a full period of (Z_t) is 2^{m+n-2} and $2^{m-1}(2^{n-1} - 1)$ respectively and the collection of the output sequences of this family of *(a, b)-SHKG*'s has good statistical properties.

In the next section, some correlation attacks on the *(a, b)-SHKG* are considered.

4 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 [6, 7, 8, 9].

The output sequence of the *(a, b)-SHKG* is an irregular decimation of its generating sequence. Thus, one would not expect a strong correlation to be obtained efficiently, especially, if the primitive characteristic feedback polynomials of the LFSRs are of high hamming weight [8], and the values for *"a"* and *"b"* which are used to clock the generating register are considered as part of the key.

The following attack on Coppersmith et al's Shrinking Generator *SG* introduced in [4] allows a cryptanalyst to reconstruct the initial states of the *SG* in a running time upper bounded by $O(2^m n^3)$ provided that the characteristic feedback polynomials of LFSRs **A** and **B** are known. In this attack, a cryptanalyst can exhaustively search for **A**'s initial state; each such state can be expanded to a prefix of the control sequence (A_t) using the characteristic feedback polynomial of **A**. Suppose that the sequence (A_t) is expanded until its n^{th} "1" is produced. From this prefix, and from the knowledge of a corresponding *n*-long prefix of the output sequence (Z_t), one can derive the value of *n* non-consecutive bits of the generating sequence (B_t). Since the characteristic feedback polynomial of **B** is known, then **B**'s initial state can be revealed given these non-consecutive *n*-bits by solving a system of linear equations. Therefore, the attack's complexity is exponential in *m* and polynomial in *n*, or more precisely, $O(2^m n^3)$.

The above attack can also be applied on the *(a, b)-SHKG* with an additional condition that is, a cryptanalyst has also to exhaustively search for the values for *"a"* and *"b"* in order to reveal the location of the *n* non-consecutive bits in the sequence (B_t), so he/she can solve the system of linear equations. Therefore, the attack takes approximately $O(\Phi\, 2^m n^3)$ where Φ is approximately the number of possible values for *"a"* and *"b"* such that $gcd(G_A(M), N) = 1$ and $m < N/c$.

If the characteristic feedback polynomials of **A** and **B** are kept secret, another attack is introduced on the *SG* in [4] which takes $O(2^{2m}\, m\, n)$ steps to recover the secret key [10].

This attack can also be applied on the *(a, b)-SHKG* with the same additional condition as in the previous attack that is, a cryptanalyst has also to exhaustively search for the values of *"a"* and *"b"*. Therefore, to recover the secret key, this attack takes $O(\Phi\, 2^{2m}\, m\, n)$ steps.

There is also another attack that can be applied to the *SG* and the *(a, b)-SHKG* through the linear complexity, but this attack requires $(2^m\, n)$ consecutive bits of the output sequence.

It is mentioned in [10] that a Shrinking Generator *SG* with secret primitive characteristic feedback polynomials, and their length satisfy $gcd(m, n) = 1$, has a security level approximately equal to 2^{2l} for $m \approx l$ and $n \approx l$. Thus, if $m \approx 64$ and $n \approx 64$, the *SG* appears to be secure against all presently known attacks mentioned in [6, 7, 8, 9, 11].

The *(a, b)-SHKG* produces bits from the generating register LFSR **B** using the same techniques as the *SG*. The only difference is that the generating register is clocked X_t times at time *t*. For *"a"* = *"b"* = *1*, the *(a, b)-SHKG* becomes a *SG*.
Like the *SG*, for $m \cong 64$ and $n \cong 64$, the *(a, b)-SHKG* appears to be secure against the correlation attacks introduced in [6, 7, 8, 9, 11, 12, 13, 14, 15, 16]. Moreover, the *(a, b)-SHKG* is more secure than the *SG* against all the above attacks if the values for *"a"* and *"b"* are taken to be part of the secret key.

For maximum security, the *(a, b)-SHKG* should be used with secret *"a"*, *"b"*, secret primitive characteristic feedback polynomials, and *"a"*, *"b"*, *m* and *n* should satisfy $gcd(a2^{m-1} + b(2^{m-1} - 1),\ 2^n - 1) = 1$ and $m < (2^n - 1)/c$. Subject to these constraints, if $m \cong l$ and $n \cong l$, the *(a, b)-SHKG* has a security level approximately equal to $\Phi\, 2^{2l}$.

Remark. *When "a" = "b" = d, the output sequence of the (d, d)-SHKG whose control and generating sequences are (A_t) and (B_t) may be seen as an (1, 1)-SHKG whose control and generating sequences are (A_t) and (C_t) where (C_t) is the d^{th} decimation of (B_t). Since any two distinct (generating) sequences (of equal period) decimated by d_i and d_k respectively (for $i \neq k$) may produce a same sequence. Then, for some values for "a" and "b" where "a" = "b" = d_j for j = 0, 1, 2, ..., an (d_i, d_i)-SHKG may generate the same output sequence as an (d_k, d_k)-SHKG where i and k ∈ {0, 1, 2, ...}. Therefore, it is suggested to avoid the case where "a" = "b".*

Example: Consider an (a, b)-SHKG with a 3-stage control register and a 4-stage generating register. For "a" = "b" = 2, the (2, 2)-SHKG with a non-zero initial state for the control register and the initial state 1111 for the generating register produces the same output sequence as the (4, 4)-SHKG with the same initial state for the control register and the initial state 1010 for the generating register. This arises since the 2nd decimation of the output sequence of the generating register with initial state 1111 is the same as the 4th decimation of the output sequence of the generating register with initial state 1010.

5 Related Work

Interesting examples of existing LFSR-based constructions for comparison with the *(a, b)-SHKG* are Coppersmith et al's Shrinking Generator *SG* [4] and some of the clock-controlled generators introduced in [1], in particular Xian and Guozhen's *(k_0, k_1)*-Clock-Controlled Generator *(k_0, k_1)-CCG* [17]. The *SG* and the *(k_0, k_1)-CCG* have similar proven properties as the *(a, b)-SHKG*.

The *(k_0, k_1)-CCG* is built up from two LFSRs **A** and **B**, and it works the same way as the *(a, b)-SHKG*, the only difference is that the output of **B** under the control of **A** is taken to be the output of the *(k_0, k_1)-CCG* regardless of the current output of **A**. One advantage of the *(k_0, k_1)-CCG* is that it generates an output bit each time **A** is clocked. On the other hand, the omission of bits, which is important in LFSR-based constructions [4] is significantly more superior for the *(a, b)-SHKG* than the *(k_0, k_1)-CCG*. For the *(k_0, k_1)-CCG* one of any $c = max(a, b)$ consecutive bits originally output by **B** appears in the output sequence of the *(k_0, k_1)-CCG*, whereas for the *(a, b)-SHKG* one of any *(m c)* consecutive bits originally output by **B** appears in the output sequence of the *(a, b)-SHKG*. Also if *k* bits from the control sequence are required to determine the original locations of *k* bits in the generating sequence of an *(k_0, k_1)-CCG*, then *2k* bits of the control register (on average) are required to determine the locations of *k* bits in the generating sequence of an *(a, b)-SHKG*.

The *SG* is a special case of our construction, it is actually an *(a, b)-SHKG* with *"a" = "b" = 1*. Although the *(a, b)-SHKG* is slower than the *SG*, its advantages is that, it provides more security as mentioned in the previous section. Moreover, for the *SG* in order to produce a new sequence, one has to at least choose another initial state or another characteristic feedback polynomial, whereas for the *(a, b)-SHKG* in order to produce a new sequence, it suffices to choose another value(s) for *"a"* and/or *"b"*.

6 Conclusion

From the theoretical results established, it is concluded that an *(a, b)-SHKG* with primitive LFSRs generates 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 polynomials, the *(a, b)-SHKG* produces a new sequence each time new value(s) are assigned for *"a"* and/or *"b"*. 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 J. Sel. Ar. Comm. vol. 7, NO. 4, May 1989, pp. 525-533.
[2]- S. W. Golomb, "Shift Register Sequences", Aegean Park Press, 1982.
[3]- R. Lidl, H. Niederreiter, " Introduction to Finite Fields and Their Applications", UK: Cambridge University Press, 1986.
[4]- D. Coppersmith, H. Krawczyk, and Y. Mansour, "The Shrinking Generator", Proceedings of Crypto 93, Springer-Verlag, 1994, pp 22-39.
[5]- A. Kanso, "Clock-Controlled Generators", PhD thesis, University of London 1999, pp. 177.
[6]- J. Golic, M. Mihaljevic, "A Generalized Correlation Attack on a Class of Stream Ciphers Based on the Levenstein Distance", Journal of Cryptology, 3, 1991, pp. 201-212.
[7]- J.Golic, "Towards Fast Correlation Attacks on Irregularly Clocked Shift Registers", Lecture Notes in Computer Science 921 (EuroCrypt'95), 1995, pp. 248-262.
[8]- W. Meir, O. Staffelbach, "Fast Correlation Attacks on Certain Stream Ciphers", Journal of Cryptology, 1, 1989, pp. 159-176.
[9]- T. Siegenthaler, "Correlation-Immunity of Non-linear Combining Functions for Cryptographic Applications", IEEE Trans On Information Theory, 30, 1984, pp.776-780.
[10]- A. Menezes, P. Van Oorshot, and S. Vanstone, "Handbook of Applied Cryptography", CRC Press, 1996.
[11]- J. Golic, "On the Security of Shift Register Based Keystream Generators", R. Anderson, Editor, Fast Software Encryption, Cambridge Security Workshop (LNCS 809), Springer-Verlag, 1994, pp. 90-100.
[12]- T. Johansson, "Reduced Complexity Correlation Attacks on Two Clock-Controlled Generators", Advances of Cryptology (AsiaCrypt 98), Lecture Notes in Computer Science, vol. 1514, 1998, pp. 342-356.
[13]- M. Mihaljevic, "An Approach to the Initial State Reconstruction of a Clock-Controlled Shift Register Based on a Novel Distance Measure", Advances in Cryptology (AusCrypt 92), Lecture Notes in Computer Science, vol. 178, 1993, pp. 349-356

[14]- J. Golic, L. O.Connor, “Embedding Probabilistic Correlation Attacks on Clock-Controlled Shift Registers”, Advances in Cryptology (EuroCrypt 94), Lecture Notes in Computer Science, vol. 950, 1995, pp. 230-243.
[15]- T. Johansson, F.Jonsson, “Improved Fast Correlation Attacks on Certain Stream Ciphers via Convolutional Codes”, Advances in Cryptography (EuroCrypt 99), Lecture Notes in Computer Science, vol. 1592, Springer-Verlag, 1999, pp. 347-362.
[16]- T. Johansson, F.Jonsson, “Fast Correlation Attacks Through Reconstruction of Linear Polynomials”, Advances in Cryptology (Crypto 2000), Lecture Notes in Computer Science, vol. 1880, Springer-Verlag, 2000, pp. 300-315.
[17]- L. Xian and Guozhen, “Analysis of (k_0, k_1) Clock-Controlled Sequences”, AAECC 6, Springer-Verlag 1995, pp. 159-169.

Appendix

Proof of lemma 1: Recall that $Z_i = B_{Ge(i)}$.

Fact 1: Recall that in a full period of (A_t) the number of 1’s is M_1, so when considering a full period of (A_t) there are M_1 outputs Z_i and the sequence (B_t) advances $G_A(M)$ places, so $\forall j \geq 0,\ Z_{i + jM_1} = B_{Ge(i) + jG_A(M)}$. (eq. 1)

Fact 2: Let k, k' be any pair of indices. If $\forall j\!: B_{k + jG_A(M)} = B_{k' + jG_A(M)}$, then N divides $(k - k')$.
Proof of fact 2: Define a sequence (C_t) where $C_t = B_{tG_A(M)}\ \forall t \geq 0$. The sequence (C_t) is a decimation of the sequence (B_t) by $G_A(M)$. As $gcd(G_A(M), N) = 1$ and (B_t) has period N, then the sequence (C_t) also has period N.

Now if $B_{k + jG_A(M)} = B_{k' + jG_A(M)}\ \forall j \geq 0$, then the translates $(C_{t + h})$ and $(C_{t + h'})$ are equal where $k = hG_A(M)\ (mod\ N)$ and $k' = h'G_A(M)\ (mod\ N)$. Hence, N divides $(h - h')$ so that N divides $(h - h')G_A(M)$ i.e. N *divides* $(k - k')$.

Proceeding with the main proof. Let P_Z be the period of the sequence (Z_t). By the argument given above P_Z must divide $M_1N/gcd(G_A(M), N) = M_1N$.

Proceeding to show that M_1N divides P_Z. By definition $Z_i = Z_{i + Pz}$.
In particular, $\forall i, j,\ Z_{i + jM_1} = Z_{i + Pz + jM_1}$.
Using (eq. 1), $\forall i, j\!: B_{Ge(i) + jG_A(M)} = B_{Ge(i + Pz) + jG_A(M)}$.
Using (fact 2), $\forall i$, N *divides* $Ge(i + P_Z) - Ge(i)$. (eq. 2)

Next step is to show that (eq. 2) is possible only if M_1 divides P_Z.

Rewrite (eq. 2) as follows:
$\forall i, \exists j_i\!: Ge(i + P_Z) = Ge(i) + j_iN$. (eq. 3)
Putting $(i + 1)$ instead of (i) in (eq. 3):
$Ge(i + 1 + P_Z) = Ge(i + 1) + j_{i+1}N$. (eq. 4)
Subtracting (eq. 3) from (eq. 4):
$\forall i,\ Ge(i + 1 + P_Z) - Ge(i + P_Z) = Ge(i + 1) - Ge(i) + (j_{i+1} - j_i)N$. (eq. 5)

Notice that, $Ge(i + 1) - Ge(i) \leq (mc)$ since one can not have more than $(m - 1)$ consecutive zeroes in the m-stage LFSR **A**.

If $j_{i+1} - j_i$ were different than zero, it would imply that $N \leq (mc)$, which is impossible assuming $m < (N/c)$. Therefore, $(j_{i+1} - j_i = 0)$, and then
$\forall i,\ Ge(i + 1 + P_Z) - Ge(i + P_Z) = Ge(i + 1) - Ge(i).$

The latter implies that the translate of (A_t) starting at $A_{i'}$ [where $Ge(i) = G_A(i')$, and $A_{i'} = 1$] is identical to the translate starting at $A_{(i+Pz)'}$. This means that M divides $[(i + P_Z)' - i']$, or equivalently, that the number of elements in the sequence (A_t) between $A_{i'}$ and $A_{(i+Pz)'}$ is a multiple of its period. But then the number of 1's in this segment is a multiple of M_1. On the other hand, the number of 1's is exactly P_Z, thus proving that M_1 divides P_Z.

Let h be such that: $P_Z = hM_1$. (eq 6)
$\forall j,\ B_{Ge(0)} = Z_0 = Z_{jPz} = Z_{jhM_1} = B_{Ge(0) + jhG_A(M)}$. (eq 7)

The last equality follows from (eq 1). So $\forall j$: $B_{Ge(0)} = B_{Ge(0) + jhG_A(M)}$. This implies that N divides $hG_A(M)$, and since $gcd(G_A(M), N) = 1$, then N divides h. From (eq. 6) M_1N divides P_Z.

Hence, the period P_Z of (Z_t) is equal to M_1N.