Stochastic Non-Linear Pseudo-Random Sequence Generator

Mahmood A. Shamran *

Date of acceptance 30/6 / 2009

Abstract:

Many of the key stream generators which are used in practice are LFSR-based in the sense that they produce the key stream according to a rule *y = C(L(x))*, where *L(x)* denotes an internal linear bit stream, produced by small number of parallel linear feedback shift registers (LFSRs), and C denotes some nonlinear compression function. In this paper we combine between the output sequences from the linear feedback shift registers with the sequences out from non linear key generator to get the final very strong key sequence

Key words:Stochastic process, birth and death key generator (BDG), linear feedback shift register (LFSR), Pseudo- noise sequence, Auto-correlation function, Periodicity.

Introduction:

Random numbers (in some sense) is important in many applications such as computer simulation, Monte Carlo integration, cryptography, randomized computation, ranging. In each case we need a sequence of numbers (or of bits) that "appears randomly", yet is repeatable. There is often a trade off- in order to pass many tests that may be necessary to make the sequence generators very complex, making it hard to analyze the sequence with respect to the randomness measures. [1].

In 1999, we introduce a new type of random sequence key generator for stream cipher purpose, based on the stochastic process specially the birth and death process named “Birth and Death Key Generator”,[BDG for short]. [2]

In this paper we introduce a new system for the same type, in fact we make a combination between the (BDG) and the linear feedback shift register. This system has a many strong points which make it very hard for attack methods.

Preliminaries:

1. <u>Definition (Stochastic process):</u>[3]

A stochastic process [S.P. for short] $\{X_t\}$ is a collection $\{X_t : t \in I\}$ of random variables. Typically, *I* is an interval in R (in such case we say that $\{X_t\}$is a continuous time stochastic process), or a subset of $\{1,2,...,n,...\}$(in such case we say that $\{X_t\}$is a discrete time stochastic process. We also call $t \longleftrightarrow X_t(w)$ the sample function (or sample path) of the S.P.

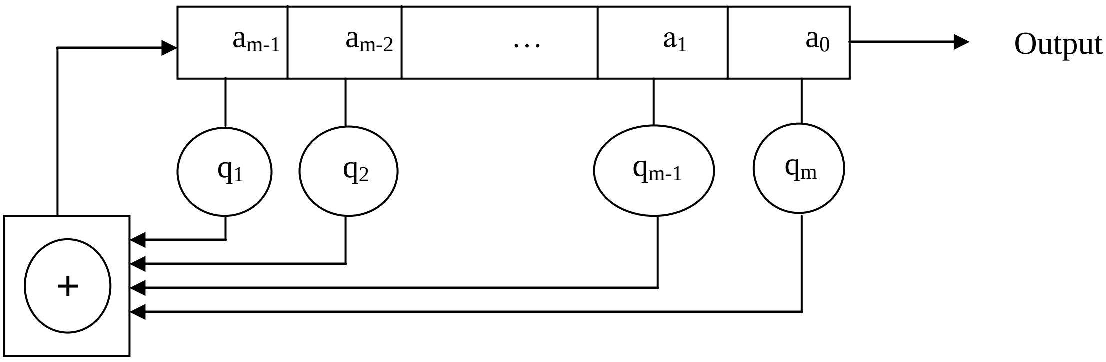
2. <u>Definition (Linear feedback shift register):</u> [1]

A linear feedback shift register [LFSR for short] of length m over a ring *R* with coefficients $q_1, q_2, ..., q_m \in R$ is a sequence generator whose state is an element S=(a_{m-1},a_{m-2},…,a_0)$\in R^m = \sum$, where $\sum$ is the set of the input state And whose state change operation *r* is given by

(a_{m-1},a_{m-2},…,a_0)$\longrightarrow (\sum q_i a_{m-i}, a_{m-1}, ..., a_1)$

See fig(1).

*Mathematics Dept./College of Science for Women/Baghdad University



Figure(1) A linear feedback Shift register of length m

3. Definition (Linear feedback shift register sequence): [4]

A linear feedback shift register sequence modulo *n* of length k>0 is a sequence $s_1, s_2, \dots$ such that $s_1, s_2, \dots, s_k$ are given and

$$s_{k+i} \equiv a_1 s_i + a_2 s_{i+1} + \dots + a_k s_{k+i-1} \pmod n \quad ,$$

where $a_1, a_2, \dots, a_k$ are given integers.

4. Definition (Period): [1]

Let *A* be a set and let a=(a_0,a_1,a_2,…) be a sequence of elements $a_i \in A,$ the sequence a is periodic if there exists an integer *P*>0 such that $a_i \equiv a_{i+p}$ for all i=0,1,2,… such that *P* is called a period of the sequence a and the least *P* is called period.

5. Definition (Auto-correlation Function): [5]

The auto-correlation function is a way to quantize how random a sequence is and is defined by

$$AC(k) = \frac{1}{p}\sum_{i=1}^{p} a_i . a_{i+k}$$

Where *p* is the period of the sequence $\{a_i\}_{i=1}^{\infty}$ and when 0<*k*<*p*, *AC*(*k*) is close to zero (meaning there is very little correlation of the sequence with itself) and *AC*(*k*)=1/2 when *k*=0, indicating that the number of 1's is equal to the number of 0's.

6. Definition(Pseudo-noise Sequence): [6]

$i>0$ Let $\{a_i\}_{i=1}^{\infty}$ be a binary sequence satisfying the Golomb's postulates:

1- The number of 1's in every period differ from the number of 0's by at most one.

2- In every period, at least half of the runs must have length 1, at least one-fourth length 2, etc., as long as the number of runs so indicated exceeds one. Moreover, for each of these lengths, there must be (almost) equally many runs of 0's and 1's.

3- $\{a_i\}_{i=1}^{\infty}$ is a 2-level autocorrelation sequence. That is

$$AC(k)=\begin{cases} N & \text{if } k \equiv 0 \pmod N \\ -1 & \text{otherwise} \end{cases}$$

A binary sequence that satisfies the Golomb's postulates is called a pseudo-noise sequence or a pn-sequence.

Remark:

The linear feedback shift register sequence is pseudo-random sequence.

7. **Definition (m-sequence): [1]**

A sequence a= $\{a_i\}^{\infty}_{i=1}$ is called m-sequence (over a ring R) of degree r if it can be generated by a linear feedback shift register with length r, and if every nonzero block of length r occurs exactly once in each period of a.

In other words, the sequence a is the output sequence of a LFSR that cycles through all possible nonzero states before it repeats.

Birth and Death Key Generator (BDG) [2]

Let a and b be two prime numbers such that a-1 and b-1 represent the order of cyclic group Z_a, Z_b respectively, so that each of these groups has a generator elements.

$$Y = \frac{(e^{-a} - b.e^{-a/b}) + T(e^{-a/b} - e^{-a})}{1-b} \qquad \ldots\ldots(1)$$

Where $T \in$(set of the generator elements of Z_b), and a , b represent the birth and death rate respectively, which are two generator elements in Z_a , Z_b .

The relation (1) represents the equation of straight line within the time interval [1,b]. And

$$P(x(t)=0) = \frac{d}{a}(1-Y) + 0.5Y \qquad \ldots\ldots(2)$$

Where $d \in$(set of the generator elements of Z_a).

If $P(x(t)=0) \le 0.5$ then the generated bit is "1" and if $P(x(t)=0) > 0.5$ Then the generated bit is "0", repeat this process for each generator element of Z_a with all the generator elements of Z_b then we get the sequence of bits represent the output of the (BDG) key generator.

Stochastic Non linear Key generator :

If we add the output sequence from the linear feedback shift register (LFSR) having a suitable length with delay by 1bit to the output sequence from the (BDG). Then this new suggested key generator called Stochastic Non Linear Key Generator [SNG for short].
See fig (2).

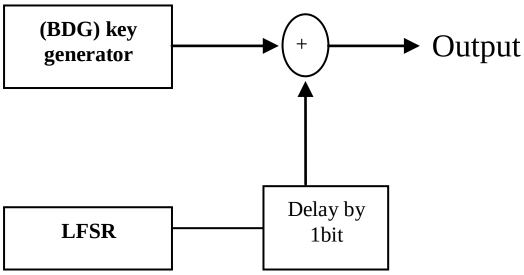


Figure (2) Stochastic Non linear key generator

Main Results:

The resulting sequences from the new suggested key generator (SNG) have the efficient criteria as shown in the sequent tables:

1 Periodicity:

Having a period is clearly a statistical defect that distinguishes a sequence from a random one. A cipher with a too small period is obviously easy to predict. The period must be large enough to ensure that is never repeated. It is usually done by using a building block that can be proven to have a large period, for instance a maximum-length LFSR.[6]
The period for the sequence output from this system is:
per(SNGsequence)= LCM(per (LFSR), per(BDG))

2 Complexity:

The different complexities try to measure how hard a sequence is to produce. For a complexity measure to be practically useful for cryptographic purposes, two requirements have to be fulfilled.

- An efficient algorithm to calculate the complexity has to be known.
- The distribution of the complexity for random sequences has to be known

If these requirements are fulfilled, the complexity measure can be used as a statistical

test. Complexity measures are often more interesting in cryptology than other statistical

17 tests, as some of them give methods to recreate the sequence using building blocks commonly used in stream ciphers. For the linear complexity, the distribution for a random sequence has been exactly calculated. It has been approximated for the maximum order complexity. Some complexity measures are only of theoretical interest because there is no efficient algorithm to calculate them.[6]

Then the linear complexity for (SNG) system is:

LC(SNG)=LC(LFSR)+LC(BDG)

3 Auto- Correlation:

We can treatment this subject by using the delay by (1) bit to the sequence from the (LFSR) before adding it to the sequence from (BDG), this treatment deletes the strong correlation relationship between the output sequence with the sequence out from the (LFSR).

4 Other Statistical tests:

Regarding other tests such as run test, poker test, frequency test…etc. we test the output sequence from the (SNG)system to measure how hard a produce sequence. This results as shown in the tables below.

Table(1) shows The Periodicity, Linear Complicity[6] and Randomness (Frequency, Run and auto correlation) tests for some binary sequences which are generate from (SNG) with different prime numbers and LFSR length.

Table (1) efficiency criterions for (SNG) output results.

LFSR length	*OP*	*OT*	*L(S)*	*LC*	*Randomness*		
					Freq.	*Run*	*AC*
37	*53*	*89*	*10000*	*4992*	*P*	*P*	*F P P P P P P P P F*
	83	*103*	*10000*	*4989*	*P*	*P*	*P P P P P P P P P P*
43	*211*	*163*	*50000*	*24985*	*P*	*P*	*P P P P F F P P P F*
	59	*61*	*50000*	*24966*	*P*	*P*	*P P P P P P P P P P*
501	*457*	*367*	*100000*	*49986*	*P*	*P*	*P P P P F P F P P P*
601	*83*	*691*	*250000*	*125012*	*P*	*P*	*P P P P P P P P P P*

Table(2) shows output results of various (SNG) system tested by CRYPT -X'98 using Periodicity, Linear Complicity, Frequency, Binary Derivative, Change point, Sub block, Run and Sequence Complicity tests [6].

Table (2) tests results of SNG system for using XOR-CF.

LFSR length	*Primes*	*L(S)*	*LC*	*FT*	*BDT*	*CPT*	*SBT*	*RT*	*SCT*
103	***OP** 101* ***OT** 997*	*25000*	*11989*	*P*	*P*	*P*	*P*	*P*	*P*
523	*199* *1103*	*100000*	*49991*	*P*	*P*	*P*	*P*	*P*	*P*
613	*149* *509*	*500000*	*250103*	*P*	*P*	*P*	*P*	*P*	*P*

Reference:

1.Mark, G. and Andrew, K. 2008." Algebraic Shift Register Sequences": University of Toronto. pp482.

2.Mahmood, A. Sh. 1999.Thesis," Application of Markov Process To Generate Binary Sequences Which is Used in Stream Cipher Systems",AL-Mustansiriyah, University, Iraq.

3.Amir, D. 2008."Stochastic Process": Standford University. standford. pp131

4.Schneier, B. 1996. “Applied Cryptography: Protocols, Algorithms, and Source Code in C”, John Wiley & Sons, New York, 2nd edition,pp1020.

5.Brock, B.T.2006."Linear Feedback Shift Registers and Cyclic Codes in SAGE": Mathematics Dept. USNA.pp221

6.MATTSSON, J. 2006.Thesis," Stream Cipher Design": Royal Institute of Technology School of Computer Science and Communication Stockholm, Sweden.

مولد المتسلسلات شبه العشوائية اللاخطي التصادفي

محمود عريبي شمران*

*كلية العلوم للبنات/قسم الرياضيات/جامعة بغداد

الخلاصة:

معظم مولدات المفاتيح التي تستند الى فلسفة الزواحف الخطية تنتج متسلسلات مفاتيح ناتجة من المتسلسلة الداخلية الصادرة عن عدد قليل من الزواحف الخطية المتوازية وبعض الدوال اللاخطية والتي تعطي قوة لمتسلسلة المفتاح الناتجة ضد المهاجمة والكسر.

في هذا البحث تم المزج بين متسلسلة مفتاح من مولد لاخطي هو (Birth and Death Key Generator) ومتسلسلة مفتاح ناتجة من زاحف خطي بطول مناسب مع تأخير بمقدار بت واحد لتجاوز مشكلة الارتباط الذاتي(Autocorrelation) والحصول على متسلسلة مفتاح نهائية تحمل مواصفات شبه عشوائية جيدة ودورية عالية جداً وتجتاز كافة الاختبارات الاحصائية المطلوبة بنجاح وتكون قوية ضد طرق المهاجمة والكسر وتلبي متطلبات الاستخدام الآمن لأغراض التشفير الانسيابي.