Theoretical Estimation of the Autocorrelation Postulate for Nonlinear Sequences Generated from Threshold

Faez Hassan Ali
Sabah Mahmood Shaker
College of Science/Al-Mustansiriya University

Abstract

The Randomness is one of the basic criterions to measure Key Generator Efficiency. The key generator depends basically on Linear FeedBack Shift Register which is considered as one of the basic units of Stream Cipher Systems. In this paper, the autocorrelation postulate, which one of the basis of Randomness criteria, is calculated theoretically for non-linear key generator before it be implemented or constructed (software or hardware), this procedure save time and costs. A nonlinear key generator is chosen to apply the theoretical studies, this key generator is the Threshold.

1. Introduction

Linear Feedback Shift Register (LFSR) and Combining Function (CF) are considered as basic units to construct key generator (KG) that used in Stream Cipher Systems (SCS) [1]. Any weakness in any one of these units means clear weakness in KG sequence, so there are some conditions must be available in KG before it is constructed; therefore the KG efficiency is concluded.

In this paper, some studies are applied on the KG sequences to determine the sequence autocorrelation. The Basic efficiency for KG can be defined as the ability of KG and its sequence to withstand the mathematical analytic which the cryptanalyst applied on them, this ability measured by some basic criterions, the most important is the randomness, one of the randomness postulates is the autocorrelation postulate.

In the next part of this paper, the autocorrelation postulate of randomness criterion will be discussed in details and introduce the basic conditions to obtain efficient KG specially those related to autocorrelation. It's important to mention that the zero input sequences must be avoided, this done when the non-all zeros initial values for LFSR's are chosen.

Let KG consist of n-LFSR's have lengths $r_1,r_2,..,r_n$ respectively with $CF=F_n(x_1,x_2,\ldots,x_n)$, s.t. $x_i\in\{0,1\}$ $1\le i\le n$, represents the output of $LFSR_i$, let $S=\{s_0,s_1,\ldots\}$ be the sequence product from SCG and s_j, $j=0,1,\ldots$ represents elements of S. let S_i be the sequence i product from $LFSR_i$ with a_{ij} elements

$1 \le i \le n$, $j=0,1,\ldots,$.

2. Conditions of the Theoretical Estimation

Definition (1) [2]: Let $GCD_2=gcd(\prod_{i=1}^{1} m_i, m_2.GCD_1)=gcd(m_1,m_2)$, for convenient let $GCD_1=1$ and so on the general form of the recursion equation will be:

$$GCD_n=gcd(\prod_{i=1}^{n-1} m_i, mn.GCD_{n-1}) \quad \ldots(1)$$

where $n \ge 2$ s.t mi are positive integers, $\forall 1 \le i \le n$.

Let the sequence S has period P(S), the period of $LFSR_i$ denotes by $P(S_i)$, P(S) and $P(S_i)$ are least possible positive integers, so

$$P(S)=lcm(P(S_1),P(S_2),\ldots,P(S_n)) \quad \ldots(2)$$

$$P(S)=\frac{\prod_{i=1}^{n} P(S_i)}{GCD_n(P(S_i))} \quad \ldots(3)$$

s.t. $GCD_n(P(S_i))= gcd\left[\prod_{i=1}^{n-1} P(S_i), P(S_n)\cdot GCD_{n-1}(P(S_i))\right]$

If $P(S_i)$ are relatively prime with each other this mean $GCD_n(P(S_i))=1$ this implies [2]:

$$P(S)=\prod_{i=1}^{n} P(S_i) \quad \ldots(4)$$

It's known earlier that $P(S_i) \le 2^{r_i}-1$, and if the $LFSR_i$ has maximum period then $P(S_i)= 2^{r_i}-1$ [3].

Theorem (1) [2]

$P(S)=\prod_{i=1}^{n}(2^{r_i}-1)$ if and only if the following conditions are holds:

1. $GCD_n(P(S_i))=1$,.
2. the period of each LFSR has maximum period ($P(S_i)=2^{r_i}-1$).

3. Randomness

The sequence that is satisfied the 3-randomness properties called Pseudo Random Sequence (PRS) [3]. The randomness criterion depends on LFSR's and CF units, therefore from the important conditions to get PRS is, the sequence must be maximal and CF must be balance [4].

To guarantee the KG to produces PRS, the sequence must pass randomness tests with complete period, these tests applied into two ways, on: [1]

1. Global sequence for complete period and that is the right way (but it's hard to applied for high periods).
2. Local sequence for many times for various lengths less than the origin length.

In this part, the 1st way will be applied theoretically for any period.

If $GCD_n(P(S_i))=1$ then,

$$P(S)=2^{\sum_{i=1}^{n} r_i}+(-1)\cdot(2^{r_1+\cdots+r_{n-1}}+\cdots+2^{r_2+\cdots+r_n}+\cdots+(-1)^{n-1}\cdot(2^{r_1}+\cdots+2^{r_n})+(-1)^n \quad \ldots(5)$$

Let R_m^t denotes the combination to sum m of numbers r_i from n of the numbers r_i, R_m denotes the set of all possibilities of R_m^t s.t.

$$R_m^t=\begin{pmatrix} r_1, r_2,..., r_n \\ \sum_{j=1}^{m} r_{i_j} \end{pmatrix} \quad 0\le m\le n,\ 1\le i\le n,\ t\in\{1,2,\ldots,C_m^n\}$$

define $R_0=\{R_0^1\}$, $R_0^1=0$.

For instance let m=1 then $R_1=\{R_1^1,R_1^2,...,R_1^{C_1^n}\}, R_1^1=r_1,...,R_1^n=r_n$

If m=n then $R_n=\{R_n^1\}$, $R_n^1=\sum_{i=1}^{n} r_i$

So equation (5) can be written in compact formula:

$$P(S)=\sum_{k=0}^{n}(-1)^k\cdot\sum_{t=1}^{C_k^n}2^{R_{n-k}^t} \quad \ldots(6)$$

Golomb deduced three theorems about the maximal sequence generated from LFSR [3]. One of the three Golomb's theorems deduced from the autocorrelation postulate. In the next sections we will introduce new theorems, as Golomb did on LFSR.

4. Autocorrelation Postulate

Before we involve in details of calculating this part of randomness criterion we have to give some preliminaries.

Let $S_r=\{a_j\}_{j=0}^{p(S_r)-1}$ be the sequence generated from maximum LFSR, s.t. $a_j\in\{0,1\}$. In corresponding let $Q_r=\{b_j\}_{j=0}^{p(S_r)-1}$ denotes the transform sequence gotten from the following linear transform:

$b=1-2a$ ...(7)

Where $b_j\in\{-1,1\}$.

a=0,1, then is corresponding b=1, -1 respectively.

Definition (2) [3]**:** When the LFSR has maximum period s.t. $P(S_r)=2^r-1$, then its can generates k sequences A_k, $1\le k\le P(S_r)-1$, each generated using the initial vector v_k s.t. $A_k=\{a_{k_j}\}_{j=0}^{p(S_r)-1}$, ($A_0=\{0,0,\ldots,0\}$), then the set $A=\{A_0,A_1,\ldots,A_{P(Sr)-1}\}$ with XOR $\oplus$ operation $\langle A,\oplus\rangle$ form a group.

Golomb mentioned that for MS the $\sum_{i=0}^{P(S_r)-1}a_i=1$ and $\sum_{i=0}^{P(S_r)-1}b_i=-1$, and $P(S_r)=P(Q_r)=N_Q(1)+N_Q(-1)$.

Definition (3): Let $B_k=\{b_{k_j}\}_{j=0}^{P(S_r)-1}$ be the corresponding to A_k mentioned above when $0\le k\le P(S_r)-1$, ($B_0=\{1,1,\ldots,1\}$), then they form a set $B=\{B_0,B_1,\ldots,B_{P(Sr)-1}\}$.

Lemma (1) [3]:

Let $B=\{B_0,B_1,\ldots,B_{P(r)-1}\}$ be a non-empty set as defined above, then $\langle B,\cdot\rangle$ is a group.

As known:

$a_1+a_2=a_1\oplus a_2\oplus a_1a_2$ …(8)

and

$a_1\oplus a_2=a_1+a_2-2a_1a_2$ …(9)

Definition (4) [3]**:** Lets $C_r(\tau)$ be the auto correlation function of maximal sequence which is generated from LFSR with length r and shifted by integer τ s.t:

$$C_r(\tau)=\frac{1}{P(S_r)}d_r(\tau),\ \text{where}$$

$$d_r(\tau)=\sum_{k=0}^{P(S_r)-1}b_kb_{k+\tau}=\begin{cases}P(S) & \tau=0,P(S_r)\\ -1 & 0<\tau<P(S_r)\end{cases} \quad \ldots(12)$$

Remark (1): $d_r(\tau)$ can represents the difference between $N_r(1)$ and $N_r(-1)$ of the sequence Q_r after shifted by τ.

Definition (5) [2]: The auto correlation function $C_S(\tau)$ of the sequence S (or the corresponding sequence Q) which is generated from system of LFSR's can be defined as follows:

$$\left.\begin{aligned}C_S(\tau)&=\frac{1}{P(S)}d_S(\tau),\ \text{where}\\ d_S(\tau)&=\sum_{k=0}^{P(S)-1}q_kq_{k+\tau}\end{aligned}\right\} \quad \ldots(13)$$

Where $q_k\in\{-1,1\}$ is the element k of the sequence Q.

Remark (2): $d_S(\tau)$ represents the difference between $N_S(1)$ and $N_S(-1)$ of the sequence Q after shifted τ.

Definition (6) [2]**:** Let T_k^t denotes the combination to multiply k of $P(S_i)$ from the total number n of $P(S_i)$, $1\le i\le n$.

Let T_k denotes the set of all possibilities of T_k^t, s.t.

$$T_k^t=\left(\prod_{i=1}^{k}P(S_{i_j})\right)^{P(S_1),\ldots P(S_n)},\ 0\le k\le n,\ t\in\{1,2,\ldots,C_k^n\},$$

we defined $T_0=\{T_0^1\}$, $T_0^1=1$

For instance, let k=1, then $T_1=\{T_1^1,T_1^2,\ldots,T_1^n\}$, $T_1^i=P(S_i)$, $1\le i\le n$.

When k=n, then $T_1=\{T_n^1\}$, s.t. $T_n^1=\prod_{j=1}^{n}P(S_j)$

Definition (7) [2]**:** Let the CF be F_n, s.t. $F_n:A\rightarrow\{0,1\}$, let H_n be the corresponding function of F_n s.t. $H_n:B\rightarrow\{-1,1\}$.

Lemma (2): If F_n is the linear function s.t. $s=F_2(a_1,a_2)=a_1\oplus a_2$, then $q=H_2(b_1,b_2)=b_1.b_2$

Where s and q are the output element of the functions F_n and H_n respectively.

Proof: By using equation (7):

$s=\frac{1}{2}(1-q)$ …(14)

From equation (9):

$s= F_2(a_1,a_2)=a_1\oplus a_2= a_1+a_2-2a_1a_2$

Reuse equation (7) in a_i and b_i we get:

$s = \frac{1}{2}[(1-b_1)+(1-b_2)]-2(1-b_1)(1-b_2)$ …(15)

When simplify equation (15) we get:

$s = \frac{1}{2}[1-b_1b_2]$ …(16)

Compare equation (16) with (14) we get:

$q=H_2(b_1,b_2)=b_1.b_2$ ■

The next lemma discusses the behavior of H_2 when F_2 is the product function.

Lemma (3): If F_2 is the product function s.t.

$s =F_2(a_1,a_2) =a_1.a_2$, then:

$q=H_2(b_1,b_2)=\bigtriangleup_{i=1}^{2}b_i = 1-\frac{1}{2}(1-b_1)(1-b_2)$ …(17)

Where Δ (read delta) represents the multiple of (×) operation (it can be denoted by *).

Proof:

$s = a_1.a_2 = \frac{1}{2}(1-b_1).\frac{1}{2}(1-b_2) = \frac{1}{4}(1-b_1).\frac{1}{2}(1-b_2)$

Since $s =\frac{1}{2} (1-q)$, then $q=1-2s$

$\therefore 2s=\frac{1}{2}\prod_{i=1}^{2}(1-b_i)$

$\therefore q=\bigtriangleup_{i=1}^{2}b_i=1-\frac{1}{2}\prod_{i=1}^{2}(1-b_i)$ ■

Of course, if $GCD_n(P(S_i))=1$, then:

$q_m=H_n(b_{1m},b_{2m},\ldots,b_{nm})$, m=0,1,…,P(S)-1.

Golomb [3] mentioned that if the algebraic system ({0,1},⊕,•) form a filed, then algebraic system ({-1,1},×,*) is a filed too [3], s.t. 1 and -1 are identity elements of the operations × and * respectively, s.t.

*	1	-1
1	1	1
-1	1	-1

Because of 3rd Golomb's theorem, two states can be concluded where $0<\tau\leq P(S)-1$, when we focus in the state which we notice that the frequency of $\tau\not\equiv 0(\mathrm{mod}\, T_k^t)$, $\forall 1\leq k\leq n$, is more than other states, this state occurs exactly $\Phi(P(S))$ times, since its represents the number of the relatively prime numbers with P(S). Actually, we know that $P(S)=\prod_{i=1}^{n} P_i=\prod_{i=1}^{n}(2^{r_i}-1)=\prod_{i=1}^{n} p_i^{q_i}$, where p_i are primes chosen as large as possible and q_i are non-negative integers, then p_i-1 approaches p_i, that implies $\Phi(P(S))$ approaches P(S), and that what will proved in the next lemma.

Lemma (4): The proportion of $\Phi(P(S))$ to P(S) is approach 1.

Proof:

$$\frac{\Phi(P(S))}{P(S)}=\frac{\prod_{i=1}^{n} p_i^{q_i-1}(p_i-1)}{\prod_{i=1}^{n} p_i^{q_i}}=\frac{\prod_{i=1}^{n}(p_i-1)}{\prod_{i=1}^{n} p_i} \qquad \ldots(18)$$

In equation (18) as p_i be large as possible $\Rightarrow p_i\text{-}1\rightarrow p_i$.

$\therefore \frac{\Phi(P(S))}{P(S)}\approx 1.$ ■

Example (1):

Table (1) shows the proportion of $\Phi(P(S))$ to P(S) for various lengths.

Table (1) the proportion of $\Phi(P(S))$ to P(S) for various lengths.

n	r_i	$P(S_i)$	P(S)	$\Phi(P(S))$	Proportion
2	2,5	3,31	93	60	65%
	3,4	7,15	105	48	46%
3	2,3,5	3,7,31	651	360	55%
	3,5,7	7,31,127	27559	22580	82%
	5,7,13	31,127,8191	32247967	30958200	96%

5. Threshold System (3-TSCG)

This SCG as usual using CF called Majority function which is balance and symmetric (which expect that the 3-TSCG will produces PRS). This generator illustrated in figure (1) tries to get around the security problems by using a variable number of LFSR's [4]. The theory is that if you use a lot of LFSRs, it's harder to break the cipher.

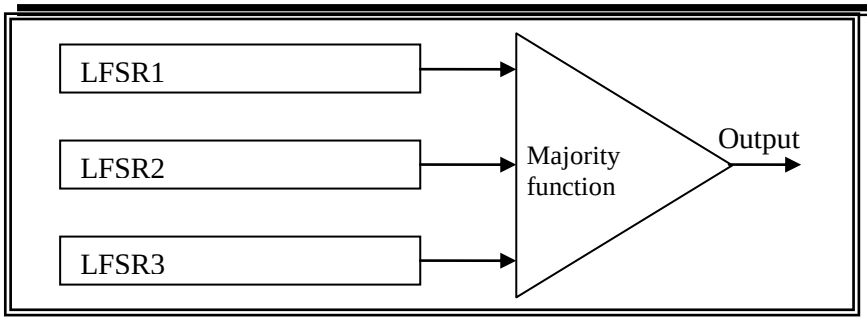

Figure (1) Threshold CSG [4].

Take the output of a large number of LFSRs (use an odd number of them). Make sure the lengths of all the LFSRs are relatively prime and all the feedback polynomials are primitive: maximize the period. If more than half the output bits are 1, then the output of the generator is 1. If more than half the output bits are 0, then the output of the generator is 0.

With three LFSRs, the output generator can be written as:

The Threshold SCG using the non-linear CF s.t:

$F_3(x_1,x_2,x_3)=x_1x_2\oplus x_1x_3\oplus x_2x_3$.

From the combing function of this generator, except that it has a larger linear complexity [5]:

$LC(S) = r_1r_2 + r_1r_3 + r_2r_3$

where r_1, r_2, and r_3 are the lengths of the first, second, and third LFSRs.

6. Implementation of Autocorrelation Postulate on 3-TSCG

Notice that:

$q_m=(b_{1m}*b_{2m})\,(b_{1m}*b_{3m})\,(b_{2m}*b_{3m})$, $m=0,\ldots,P(S)-1$.

Before involved in calculating $\sum_{m=0}^{P(S)-1} q_m$ we have to prove the following facts:

Fact (1): if $a\in\{-1,1\}$, then $a^2=1$.

Proof: is trivial.

Fact (2): The distributive law of the operation (*) on (·) is satisfied s.t.

$a*(b\cdot c)=(a*b)\cdot(a*c)$, where $a,b,c\in\{-1,1\}$.

Proof:

$(a*b)(a*c)= [1-½(1-a)(1-b)]\,[1-½(1-a)(1-c)]=1-½[(1-a)(1-b)+ (1-a)(1-c)]$
$+¼(1-a)^2(1-b)(1-c)=1-½[1-(a+b)+ab+1-(a+c)+ac]$

Using fact (1), then:

$(a*b)(a*c)=1-½[2-(2a+b+c)+ab+ac]+¼[1-(a+b+c)+ab+ac+2bc-abc]$
$=1-½[1-(a+bc)+abc]=1-½(1-a)(1-bc)=a*(b\cdot c)$ ■

In the next facts and lemmas, let $\{a_i\}$, $\{b_i\}$ and $\{c_i\}$ be maximal sequence generated from MLFSR with length r (or r_1,r_2 and r_3).

Fact (3): $\sum_{i=0}^{P(r)-1} a_i * b = (2^{r-1}-1)+b\cdot 2^{r-1}$, where $a_i, b \in \{-1,1\}$.

Proof:

$\sum_{i=0}^{P(S_r)-1} a_i * b = a_0 * b + a_1 * b + \ldots + a_{P(Sr)-1} b$.

Since $a_i * b = 1-\frac{1}{2}(1-a_i)(1-b) = \begin{cases} 1, & \text{if } a_i=1 \\ b, & \text{if } a_i=-1 \end{cases}$

and since $N(1)=2^{r-1}-1$ and $N(-1)=2^{r-1}$, then:

$$\sum_{i=0}^{P(r)-1} a_i * b = \underbrace{1+1+\ldots+1}_{2^{r-1}-1 \text{ times}} \underbrace{+ b+b+\ldots+b}_{2^{r-1} \text{ times}} = (2^{r-1}-1)+b\cdot 2^{r-1} = 2^{r-1}(1+b)-1$$ ■

Fact (4): $\sum_{i=0}^{P(S_1)-1} \sum_{j=0}^{P(S_2)-1} a_i * b_j = 2^{r_1+r_2-1} - (2^{r_1}+2^{r_2}) + 1$, where $a_i, b_j \in \{-1,1\}$.

Proof:

$$\sum_{i=0}^{P(S_1)-1} \sum_{j=0}^{P(S_2)-1} a_i * b_j = \sum_{i=0}^{P(S_1)-1} \sum_{j=0}^{P(S_2)-1} [1-\tfrac{1}{2}(1-a_i)(1-b_j)]$$

$$= \sum_{i=0}^{P(S_1)-1} [P(r_2) - \tfrac{1}{2}(1-a_i) \sum_{j=0}^{P(S_2)-1} (1-b_j)] = \sum_{i=0}^{P(S_1)-1} [P(r_2) - \tfrac{1}{2}(1-a_i).2^{r_2}]$$

$$= P(r_1)P(r_2) - 2^{r_2-1} \sum_{i=0}^{P(S_1)-1} (1-a_i) = 2^{r_1+r_2} - (2^{r_1}+2^{r_2}) + 1 - 2^{r_1+r_2-1} = 2^{r_1+r_2-1} - (2^{r_1}+2^{r_2}) + 1$$ ■

Fact (5): $\sum_{i=0}^{P(r)-1} (a_i * b)\cdot a_i = (2^{r-1}-1) - b.2^{r-1}$.

Proof:

Since $a_i * b = \begin{cases} 1, & \text{if } a_i=1 \\ b, & \text{if } a_i=-1 \end{cases}$, then $(a_i * b)a_i = \begin{cases} 1, & \text{if } a_i=1 \\ -b, & \text{if } a_i=-1 \end{cases}$

$\sum_{i=0}^{P(r)-1} (a_i * b)\cdot a_i = (2^{r-1}-1) - b2^{r-1}$

(using Fact (1) and Fact(3)). ■

Now we are ready to calculate $\sum_{m=0}^{P(S)-1} q_m$ by using the next lemma.

Lemma (5):

$$\sum_{m=0}^{P(S)-1} q_m = -(2^{r_1+r_2-1} + 2^{r_1+r_3-1} + 2^{r_2+r_3-1}) + (2^{r_1} + 2^{r_2} + 2^{r_3}) - 1 \quad \ldots(19)$$

Proof:

$$\sum_{m=0}^{P(S)-1} q_m = \sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}\sum_{k=0}^{P(S_3)-1}(b_{1i} * b_{2j})(b_{1i} * b_{3k})(b_{2j} * b_{3k})$$

$$= (2^{r_3-1}-1)\sum_{i=0}^{P(S_1)-1}\left[(2^{r_3-1}-1)+2^{r_3-1}b_{1i}\right]+2^{r_3-1}\sum_{i=0}^{P(S_1)-1}\left[(2^{r_2-1}-1)b_{1i}-2^{r_2-1}\right]$$

$$= (2^{r_3-1}-1)(2^{r_2-1}-1)(2^{r_1}-1)-(2^{r_3-1}-1)2^{r_2-1}-2^{r_3-1}(2^{r_2-1}-1)-2^{r_3-1}2^{r_2-1}(2^{r_1}-1)$$

$$\sum_{m=0}^{P(S)-1} q_m = -(2^{r_1+r_2-1}+2^{r_1+r_3-1}+2^{r_2+r_3-1})+(2^{r_1}+2^{r_2}+2^{r_3})-1 \quad \blacksquare$$

Equation (19) can be written as:

$$\sum_{m=0}^{P(S)-1} q_m = -\sum_{j=1}^{3} 2^{R_2^j-1}+\sum_{j=1}^{3} 2^{r_j}-1$$

Before we calculating $d_S(\tau)$ we need the following Lemmas.

Lemma (6):

$$\sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}(b_{1i} * b_{2j})(b_{1i+\tau_1}+b_{2j+\tau_2})(b_{1i}b_{2j}+b_{1i+\tau_1}b_{2j+\tau_2})$$

$$= 2-2^{r_1+r_1}-(d_{r_1}(\tau_1)+1)(d_{r_2}(\tau_2)+1)-\tfrac{1}{2}(2^{r_1}+1+d_{r_1}(\tau_1)(2^{r_2}+1+d_{r_2}(\tau_2)) \qquad ...(20)$$

Proof:

$$\sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}(b_{1i} * b_{2j})(b_{1i+\tau_1}+b_{2j+\tau_2})(b_{1i}b_{2j}+b_{1i+\tau_1}b_{2j+\tau_2})$$

$$= \sum_{i=0}^{P(S_1)-1} b_{1i}\sum_{j=0}^{P(S_2)-1} b_{2j}+\sum_{i=0}^{P(S_1)-1} b_{1i+\tau_1}\sum_{j=0}^{P(S_2)-1} b_{2j+\tau_2}-\tfrac{1}{2}[\sum_{i=0}^{P(S_1)-1}(1-b_{1i})b_{1i}\sum_{j=0}^{P(S_2)-1}(1-b_{2i})b_{2i}$$

$$+\sum_{i=0}^{P(S_1)-1}(1-b_{1i})b_{1i+\tau_1}\sum_{j=0}^{P(S_2)-1}(1-b_{2i})b_{2i+\tau_2}+\sum_{i=0}^{P(S_1)-1}(1-b_{1i+\tau_1})b_{1i}\sum_{j=0}^{P(S_2)-1}(1-b_{2i+\tau_2})b_{2i}$$

$$+\sum_{i=0}^{P(S_1)-1}(1-b_{1i+\tau_1})b_{1i+\tau_1}\sum_{j=0}^{P(S_2)-1}(1-b_{2i+\tau_2})b_{2i+\tau_2}]+\tfrac{1}{4}[\sum_{i=0}^{P(S_1)-1}(1-b_{1i})(1-b_{1i+\tau_1})b_{1i}\sum_{j=0}^{P(S_2)-1}(1-b_{2i})(1-b_{2i+\tau_2})b_{2i}$$

$$+\sum_{i=0}^{P(S_1)-1}(1-b_{1i})(1-b_{1i+\tau_1})b_{1i+\tau_1}\sum_{j=0}^{P(S_2)-1}(1-b_{2i})(1-b_{2i+\tau_2})b_{2i+\tau_2}]$$

$$= 1+1-\tfrac{1}{2}[(P(r_1)+1)(P(r_2)+1)+(d_{r_1}(\tau_1)+1)(d_{r_2}(\tau_2))$$

$$+(d_{r_1}(\tau_1)+1)(d_{r_2}(\tau_2)+1)+(P(r_1)+1)(P(r_2)+1)]$$

$$+\tfrac{1}{4}[(P(r_1)+2+d_{r_1}(\tau_1))(P(r_2)+2+d_{r_2}(\tau_2))(P(r_1)+2+d_{r_1}(\tau_1))(P(r_2)+2+d_{r_2}(\tau_2))]$$

$$= 2-2^{r_1+r_1}-(d_{r_1}(\tau_1)+1)(d_{r_2}(\tau_2)+1)-\tfrac{1}{2}(2^{r_1}+1+d_{r_1}(\tau_1))(2^{r_2}+1+d_{r_2}(\tau_2)) \quad \blacksquare$$

Lemma (7): $\sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}(b_{1i} * b_{2j})(b_{1i+\tau_1} * b_{2j+\tau_2})(b_{1i}b_{2j}b_{1i+\tau_1}b_{2j+\tau_2})$

$$= d_{r_1}(\tau_1)d_{r_2}(\tau_2)-(d_{r_1}(\tau_1)+1)(d_{r_2}(\tau_2)+1)-\tfrac{1}{4}(2^{r_1}+1+d_{r_1}(\tau_1))(2^{r_2}+1+d_{r_2}(\tau_2)) \ldots(21)$$

Proof:

$$\sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}(b_{1i} * b_{2j})(b_{1i+\tau_1} * b_{2j+\tau_2})(b_{1i}b_{2j}b_{1i+\tau_1}b_{2j+\tau_2})$$

$$= \sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}\{1-\tfrac{1}{2}[(1-b_{1i})(1-b_{2j})+(1-b_{1i+\tau_1})(1-b_{2j+\tau_2})]$$

$$+\tfrac{1}{4}[(1-b_{1i})(1-b_{2j})(1-b_{1i+\tau_1})(1-b_{2j+\tau_2})]\}(b_{1i}b_{2j}b_{1i+\tau_1}b_{2j+\tau_2})$$

$$= \sum_{i=0}^{P(S_1)-1}b_{1i}b_{1i+\tau_1}\sum_{j=0}^{P(S_2)-1}b_{2j}b_{2j+\tau_2} - \tfrac{1}{2}[\sum_{i=0}^{P(S_1)-1}(1-b_{1i})b_{1i}b_{1i+\tau_1}\sum_{j=0}^{P(S_2)-1}(1-b_{2j})b_{2j}b_{2j+\tau_2}$$

$$+\sum_{i=0}^{P(S_1)-1}(1-b_{1i+\tau_1})b_{1i}b_{1i+\tau_1}\sum_{j=0}^{P(S_2)-1}(1-b_{2j+\tau_2})b_{2j}b_{2j+\tau_2}]$$

$$+\tfrac{1}{4}\sum_{i=0}^{P(S_1)-1}(1-b_{1i})(1-b_{1i+\tau_1})b_{1i}b_{1i+\tau_1}\sum_{j=0}^{P(S_2)-1}(1-b_{2j})(1-b_{2j+\tau_2})b_{2j}b_{2j+\tau_2}$$

$$= d_{r_1}(\tau_1)d_{r_2}(\tau_2) - \tfrac{1}{2}[(d_{r_1}(\tau_1)+1)(d_{r_2}(\tau_2)+1)+(d_{r_1}(\tau_1)+1)(d_{r_2}(\tau_2)+1)]$$

$$+\tfrac{1}{4}(P(r_1)+2+d_{r_1}(\tau_1))(P(r_1)+2+d_{r_1}(\tau_1))$$

$$= d_{r_1}(\tau_1)d_{r_2}(\tau_2) - (d_{r_1}(\tau_1)+1)(d_{r_2}(\tau_2)+1) - \tfrac{1}{4}(2^{r_1}+1+d_{r_1}(\tau_1))(2^{r_2}+1+d_{r_2}(\tau_2)) \quad \blacksquare$$

Now we will shifting Q_m by τ to find $d_S(\tau)$ by using the next theorem.

Theorem (2):

$$d_S(\tau)= \tfrac{1}{4}[d_{r_1}(\tau_1)(P(r_2)P(r_3)-2)+d_{r_2}(\tau_2)(P(r_1)P(r_3)-2)+d_{r_3}(\tau_3)(P(r_1)P(r_2)-2)$$
$$+2(P(r_1)+P(r_2)+P(r_3))+d_{r_1}(\tau_1)d_{r_2}(\tau_2)d_{r_3}(\tau_3)] \quad \ldots(22)$$

Proof:

$$d_S(\tau)= \sum_{m=0}^{P(S)-1}Q_m = \sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}\sum_{k=0}^{P(S_3)-1}(b_{1i}*b_{2j})(b_{1i}*b_{3k})(b_{2j}*b_{3k})$$
$$(b_{1i+\tau_1}*b_{2j+\tau_2})(b_{1i+\tau_1}*b_{3k+\tau_3})(b_{2j+\tau_2}*b_{3k+\tau_3})$$

$$= \sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}(b_{1i}*b_{2j})(b_{1i+\tau_1}*b_{2j+\tau_2})\sum_{k=0}^{P(S_3)-1}[1-\tfrac{1}{2}(1-b_{1i}b_{2j})(1-b_{3k})][1-\tfrac{1}{2}(1-b_{1i+\tau_1}b_{2j+\tau_2})(1-b_{3k+\tau_3})]$$

$$= \sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}(b_{1i}*b_{2j})(b_{1i+\tau_1}*b_{2j+\tau_2})\{P(r_3)-2^{r_3}+2^{r_3-1}b_{1i}b_{2j}$$
$$+\tfrac{1}{4}(1-b_{1i}b_{2j})(1-b_{1i+\tau_1}b_{2j+\tau_2})(2^{r_3}+1+d_{r_3}(\tau_3))\}$$

$$= 2^{r_3-1}\sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}(b_{1i}*b_{2j})(b_{1i+\tau_1}*b_{2j+\tau_2})(b_{1i}b_{2j}+b_{1i+\tau_1}b_{2j+\tau_2})$$

$$+\tfrac{1}{4}(2^{r_3}+1+d_{r_3}(\tau_3))\{\sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}(b_{1i}*b_{2j})(b_{1i+\tau_1}*b_{2j+\tau_2})$$

$$-\sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}(b_{1i}*b_{2j})(b_{1i+\tau_1}*b_{2j+\tau_2})(b_{1i}b_{2j}+b_{1i+\tau_1}b_{2j+\tau_2})$$

$$+\sum_{i=0}^{P(S_1)-1}\sum_{j=0}^{P(S_2)-1}(b_{1i}*b_{2j})(b_{1i+\tau_1}*b_{2j+\tau_2})(b_{1i}b_{2j}b_{1i+\tau_1}b_{2j+\tau_2})\} \quad \ldots(23)$$

By substitute equations (20) and (21) in equation (23) and simplify them we get:

$$d_S(\tau)= 2^{r_3-1}[2-2^{r_1+r_2}-(d_{r_1}(\tau_1)+1)(d_{r_2}(\tau_2)+1)+\tfrac{1}{2}(2^{r_1}+1+d_{r_1}(\tau_1))(2^{r_2}+1+d_{r_2}(\tau_2))]$$

$$+\tfrac{1}{4}(2^{r_3}+1+d_{r_3}(\tau_3))[P(r_1)P(r_2)-2+d_{r_1}(\tau_1)d_{r_2}(\tau_2)]$$
$$-[P(r_1)P(r_2)-2^{r_1+r_2}+\tfrac{1}{4}(2^{r_1}+1+d_{r_1}(\tau_1))(2^{r_2}+1+d_{r_2}(\tau_2))]$$

After reformulate the above equation we get:

$$d_S(\tau)=\tfrac{1}{4}[d_{r_1}(\tau_1)(P(r_2)P(r_3)-2)+d_{r_2}(\tau_2)(P(r_1)P(r_3)-2)+d_{r_3}(\tau_3)(P(r_1)P(r_2)-2)$$
$$+2(P(r_1)+P(r_2)+P(r_3))+d_{r_1}(\tau_1)d_{r_2}(\tau_2)d_{r_3}(\tau_3)]$$ ■

equation (31) can be written in the form:

$$d_S(\tau)=\tfrac{1}{4}[d_{r_1}(\tau_1)(2^{r_2+r_3}-2^{r_2}-2^{r_3}-1)+d_{r_2}(\tau_2)(2^{r_1+r_3}-2^{r_1}-2^{r_3}-1)$$
$$+d_{r_3}(\tau_3)(2^{r_1+r_2}-2^{r_1}-2^{r_2}-1)+2(2^{r_1}+2^{r_2}+2^{r_3}-1)$$
$$+d_{r_1}(\tau_1)d_{r_2}(\tau_2)d_{r_3}(\tau_3)]-1 \qquad \ldots(24)$$

According to the values of $d_{r_i}(\tau_i)$, $1\le i\le 3$, there are different values to $d_S(\tau)$. Table (2) shows the different phases of equation (24) where τ divides T_m^t (or not), $1\le m\le 3$, $1\le t\le C_m^t$.

Table (2) Different phases of equation (24) for 3-TSCG.

St	m	t	$t\mid T_m^t$	$d_{r_i}(\tau_i)$	$d_S(\tau)$
1	1	1	$T_1^1=P(r_1)$	$d_{r_2}(\tau_2)=d_{r_3}(\tau_3)=-1$	$2^{R_3^1-2}-(2^{R_2^1-1}+2^{R_2^2-1}+2^{R_2^3-2})+\sum_{i=1}^{3}2^{r_i}-1$
2	1	2	$T_1^2=P(r_2)$	$d_{r_1}(\tau_1)=d_{r_3}(\tau_3)=-1$	$2^{R_3^1-2}-(2^{R_2^1-1}+2^{R_2^2-2}+2^{R_2^3-1})+\sum_{i=1}^{3}2^{r_i}-1$
3	1	3	$T_1^3=P(r_3)$	$d_{r_1}(\tau_1)=d_{r_2}(\tau_2)=-1$	$2^{R_3^1-2}-(2^{R_2^1-2}+2^{R_2^2-1}+2^{R_2^3-1})+\sum_{i=1}^{3}2^{r_i}-1$
4	2	1	$T_2^1=P(r_1)\,P(r_2)$	$d_{r_3}(\tau_3)=-1$	$2^{R_3^1-1}-(2^{R_2^1}+2^{R_2^2-1}+2^{R_2^3-1})+\sum_{i=1}^{3}2^{r_i}-1$
5	2	2	$T_2^2=P(r_1)$ $P(r_3)$	$d_{r_2}(\tau_2)=-1$	$2^{R_3^1-1}-(2^{R_2^1-1}+2^{R_2^2}+2^{R_2^3-1})+\sum_{i=1}^{3}2^{r_i}-1$
6	2	3	$T_2^3=P(r_2)$ $P(r_3)$	$d_{r_1}(\tau_1)=-1$	$2^{R_3^1-1}-(2^{R_2^1-1}+2^{R_2^2-1}+2^{R_2^3})+\sum_{i=1}^{3}2^{r_i}-1$
7	3	1	$T_3^1=P(S)$	--------	$2^{R_3^1}-\sum_{i=1}^{3}2^{R_2^i}+\sum_{i=1}^{3}2^{r_i}-1=P(S)$
8	-	-	$t\nmid T_m^t$	$d_{r_i}(\tau_i)=-1,\ \forall i$	$-\sum_{i=1}^{3}2^{R_2^i-2}+\sum_{i=1}^{3}2^{r_i}-1$

Example(2):

Let n=3, $r_1=2$, $r_2=3$, $r_3=5$, then in table (2) the different values of $d_S(\tau)$ for 3-TSCG are appeared.

Table (2) the different values of $d_S(\tau)$ of example (2) for 3-TSCG.

States	1	2	3	4	5	6	7	8
$\mathbf{d_S(\tau)}$	155	123	99	331	283	219	651	-61

Notice that the proportions of states 1,2 and 3 are approach to 0.25, the proportions of states 4,5 and 6 are approach 0.5, while the proportion of state 8 is approach 0. Of course we focus in state 8 only.

Example (3):

Table (3) shows the proportions of each states of table (2) with many examples of lengths of the combined LFSR's.

Table (3) the proportions of each states with many examples of combined LFSR's.

r_i	Percentage of States 1,2,3	Exp.	Percentage of States 4,5,6	Exp.	Percentage of State 8	Exp.	P(S)
2,3,5	0.24,0.19,0.15		0.51,0.43,0.34		0.0940		651
3,5,7	0.24,0.22,0.21		0.5,0.48,0.43		0.0400		27559
5,7,11	0.25,0.24,0.25	0.25	0.5,0.5,0.48	0.5	0.0100	0.0	8059039
5,9,11	0.25,0.24,0.24		0.5,0.5,0.48		0.0086		32426527
7,9,11	0.25,0.25,0.25		0.5,0.5,0.5		0.0026		132844159

Where Exp. is the Expected value.

7. Applying of Chi-Square Tests on 3-TSCG

In this section we will apply chi-square test on the results gotten from calculations of autocorrelation postulate on 3-TSCG.

Let M be the number of categories in the sequence S, c_i be the category i, $N(c_i)$ be the observed frequency of the category c_i, Pr_i the probability of occurs of the category c_i, then the expected frequency E_i of the category c_i is $E_i=P(S)\cdot Pr_i$, the T (chi-square value) can be calculated as follows [6]:

$$T=\sum_{i=1}^{K}\frac{(N(c_i)-E_i)^2}{E_i} \quad \ldots(25)$$

Assuming that T distributed according to chi-square distribution by υ=M-1 freedom degree by α as significance level (as usual α=0.05%), which it has T_0 as a pass mark. If $T\le T_0$ then the hypothesis accepted and the sequence pass the test, else we reject the hypothesis and the sequence fails to pass the test, this mean that T not distributed according to chi-square distribution.

In order to test our results we have to suggest an example suitable to our study case. Let n=3, r_1=7, r_2=9 and r_3=11. P(S)= 132844159 (taken from last experiment of table (3)).

In **Auto correlation** test, take υ= 1, with α=0.05%, then T_0=3.84 (see chi-square table). Since $d_S(\tau)$ represent between the $N_S(-1)$ and $N_S(1)$ for the sequence S when its shifted by τ, then its can be used to estimate the statistic value T of chi square test by using the proportion of state 8 mentioned in table (3). We can reformulate equation (25) to be suitable to autocorrelation test, so it can write as follows:

$$T(\tau)=\frac{d_S^2(\tau)}{P(S)} \quad \ldots(26)$$

For the 3-TSCG, from equation (24) and because of equation (12) then (26) will be:

$$T=(0.0026)^2/132844159 \quad \ldots(27)$$

When substitute the information of the chosen example in equation (27), then:

$T = 5.088669 \times 10^{-14} << 3.84$, then 3-TSCG **passes** this test.

8. Conclusions

1. In this work we successes to calculate the autocorrelation property deterministically, for nonlinear generator (Threshold generator), while it was calculated before this work probabilistically.
2. As known before the chosen samples are really random although we see that they fail to pass the autocorrelation postulate for complete period cycle of the output sequence, but if we choose local sequence with chosen period, then the results of autocorrelation postulate will **passes** this test.
3. These theoretical studies can be applied on other kinds of stream cipher generators to calculate the autocorrelation to them.
4. As future work we may apply other properties of randomness criteria like, frequency and run on linear or non-linear stream cipher generators.

References

[1]. Stallings, W., "***Cryptography and Net-work Security: Principles and Practices***", Pearson Prentice-Hall, 4th Edition, 2006.

[2]. Al-Shammari, A. G. N, "***Mathematical Modeling and Analysis Technique of Stream Cipher Cryptosystems***", Ph. D. Thesis, University of Technology, Applied Sciences, 2009.

[3]. Golomb, S. W., "***Shift Register Sequences***" San Francisco: Holden Day 1967.

[4]. Brüer, J. O., "***On Nonlinear Combinations of Linear Shift Register Sequences***" Internal Report LITH-ISY-1-0572,1983.

[5]. Schneier B., "***Applied Cryptography, Second Edition: Protocols, Algorithms, and Source Code in C*** ", Wiley Computer Publishing, John Wiley & Sons, Inc., 1996.

[6]. Bluman, A. G., "***Elementary Statistic: Step by Step Approach***", 6th ed., McGraw-Hill Companies Inc., New York, NY10020, 2007.

التخمين النظري لفرضية الارتباط الذاتي للمتتابعات غير الخطية المولدة من مولد ثريشولد

م.فائز حسن علي

م.م صباح محمود شاكر

قسم الرياضيات/كلية العلوم/الجامعة المستنصرية

مستخلص

تعتبر العشوائية (Randomness) من أهم مقاييس الكفاءة الأساسية لقياس كفاءة مولدات المفاتيح. مولد المفاتيح يعتمد بشكل أساسي على المسجل الزاحف الخطي ذو التغذية الخلفية (Linear Feedback Shift Register) كونه أحد الوحدات الأساسية لنظم التشفير الانسيابي (Stream Cipher Systems). في هذا البحث، تم حساب خاصية الارتباط الذاتي، باعتبارها احد أسس العشوائية، لمتتابعة مولدة من مولد مفاتيح غير خطي **نظريا** قبل بناء او تنفيذ النظام عمليا (برمجيا او مادياً)، وهذا الأسلوب سوف يوفر الوقت والجهد والكلفة لمصمم الشفرة. تم اختيار مولد مفاتيح غير خطي لتطبيق الدراسة النظرية للبحث هي منظومة ثريشولد (Threshold).