Frequency Postulate's Theoretical Calculation for the Sequences Produced by Modified Geffe Generator

الحساب النظري لفرضية التردد للمتتابعات المولدة من مولد جيف المطور

Hussein Ali Mohammed Al_Sharifi
M.Sc. in Mathematics /College of Education/Karbala University
Email: hussein7712a@yahoo.com

Abstract

The Randomness is one of the basic criterions to measure stream cipher efficiency. The stream cipher generator depends basically on Linear FeedBack Shift Register (LFSR) which is considered as one of the basic units of Stream Cipher Systems (SCS).

The basic idea of this paper is attacking and analysis of cryptosystems. So any developing in some kinds of stream cipher generators without taking in considers the basic criteria of efficiency may give no security to the generator, so this paper consists of two parts

First, the design part, this paper introduces developing of Geffe generator by increasing the LFSR's from (three) to (five) with new combining nonlinear function which has good statistical properties. The new generator called Modified Geffe generator.

Second, the attacking part, the frequency postulate of randomness criteria is calculated theoretically, this mean the generated sequence product by the new generator can estimated and this mean clear weakness in the suggested generator.

الخلاصة

تعتبر العشوائية من اهم مقاييس الاساسية لقياس كفاءة نظم التشفير الانسيابي. ان مولد التشفير الانسيابي يعتمد بشكل اساسي على المسجل الزاحف الخطي ذو التغذية الخلفية كونه أحد الوحدات الاساسية لنظم التشفير الانسيابي.

ان فكرة البحث الاساسية هي مهاجمة وتحليل نظم التشفير. لذلك فان اي تطوير في بعض انواع انظمة التشفير الانسيابي بدون الاخذ بنظر الاعتبار لمقاييس الكفاءة الاساسية لايعطي اي زيادة في امنية المولد، لذلك فان هذا البحث يتالف من جزئين.

اولا، جزء التصميم، في هذا البحث تم تطوير مولد جيف من خلال زيادة عدد المسجلات الزاحفة من (ثلاثة) الى (خمسة) مع استخدام دالة مركبة غير خطية لها خواص احصائية جيدة. المولد الجديد يدعى مولد جيف المطور.

ثانيا، جزء المهاجمة، تم في هذا البحث، حساب خاصية التردد للعشوائية نظريا، وهذا يعني يمكن تخمين المتتابعة الناتجة من المولد الجديد وبالتالي هذا يعطي ضعف واضح في تصميم المولد.

1. Introduction

Linear Feedback Shift Register (LFSR) and Combining Function (CF) are considered as basic units to construct Stream Cipher Generator (SCG) that used in stream cipher systems [14]. Any weakness in any one of these units means clear weakness in SCG sequence, so there are some conditions must be available in SCG before it is constructed; therefore the SCG efficiency is concluded.

In 1967 [8] Golomb deduced three theorems about the maximal sequence generated from LFSR. One of the three Golomb’s theorems deduced from the frequency postulate.

Although now dated, Rueppel in 1986 [10] provides a solid introduction to the analysis and design of stream ciphers. The results on the expected linear complexity and linear complexity profile of random sequences are from Chapter 4 of Rueppel.

In 1989, Staffelbach and Meier [13] presented two new so-called fast correlation attacks which are more efficient than Siegenthaler’s attack in the case where the component LFSRs have sparse feedback polynomials, or if they have low-weight polynomial multiples (e.g., each having fewer than 10 non-zero terms) of not too large a degree.

A comprehensive survey of correlation attacks on LFSR-based stream ciphers is the paper by Golić in 1994 [7]; the cases where the combining function is memoryless or with memory, as well as when the LFSRs are clocked regularly or irregularly, are all considered.

A PH. D. thesis which introduced by Al-Ageelee [2] in 1998, this work used Genetic Algorithm in cryptanalysis of class of stream cipher system depending on finding correlation between ciphertext and the output of some of LFSR.

In 2005, Ahmed [1] introduces a paper which contained the design of artificial neural networks for decryption i.e. getting distinguished polynomial for binary sequence with linear equivalence which is equal to 8 as well as getting the binary sequence that is related to the distinguished polynomial. And it was proved by the results that by using the ANN were very appropriate for the decryption of the stream cipher systems.

In 2009 [3] Al-Shammari, A. G., introduces the estimation of the four basic criterions which are: Periodicity, Linear Complexity, Randomness and Correlation Immunity used as basic criterions to measure Key Generator Efficiency. He can calculated these basic criterions theoretically for any key generator before it be implemented or constructed (software or hardware). This work introduces the mathematical proof of the good efficiency of the linear key generator deterministically.

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

In the next part of this paper, the frequency postulate of randomness criterion will be discussed in details and introduce the basic conditions to obtain efficient SCG especially those related to frequency. 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 SCG consists of n-LFSR's with lengths $r_1,r_2,..,r_n$ respectively with $CF=F_n(x_1,x_2,\dots,x_n)$, s.t. $x_i\in\{0,1\}$ $1\le i\le n$, represents the output of $LFSR_i$, let $S=\{s_0,s_1,\dots\}$ be the sequence product from SCG and s_j, $j=0,1,\dots$ represents elements of S. let S_i be the sequence i product from $LFSR_i$ with a_{ij} elements $i=1,\dots,n$, $j=0,1,\dots,$.

2. Conditions of the Theoretical Estimation

In the next definition we want to generalize the using of gcd function.

Definition (2.1) [3]: 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:

$$GCD_n=\gcd(\prod_{i=1}^{n-1} m_i ,m_n.GCD_{n-1}) \qquad \dots(1)$$

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

Theorem (2.2) [3]: Let $m_i\in Z+$, $\forall 1\le i\le n$ then:

$$lcm(m_1,m_2,\dots,m_n)=\frac{\prod_{i=1}^{n} m_i}{GCD_n(m_i)} \qquad \dots(2)$$

where $GCD_n(m_i)$ defined in (1).

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:

$$P(S)=lcm(P(S_1),P(S_2),\dots,P(S_n)) \qquad \dots(3)$$

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

s.t. $GCD_n(P(S_i))=\gcd\left[\prod_{i=1}^{n-1} P(S_i), P(S_n)\cdot GCD_{n-1}(P(S_i))\right]$ [8]

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

$$P(S)=\prod_{i=1}^{n} P(S_i) \qquad \ldots(5)$$

It's known earlier that $P(S_i)\leq 2^{r_i}-1$, and if the $LFSR_i$ has maximum period then $P(S_i)=2^{r_i}-1$ [5].

Theorem (2.3) [3]

$P(S)=\prod_{i=1}^{n}(2^{r_i}-1)$ 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) [8]. The randomness criterion depends on LFSR's and CF units, therefore from the important conditions to get PRS is that the sequence must be maximal and CF must be balance [5].

For our purposes, a sequence generator is pseudo-random if it has this property: It looks random. This means that it passes all the statistical tests of randomness that we can find [8].

Definition (3.1) [14]: A random bit generator is a device or algorithm which outputs a sequence of statistically independent and unbiased binary digits.

To guarantee the SCG to produces PRS, the sequence must passes randomness tests with complete period, these tests applied into two ways, on: [6]

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,

$$P(S)=2^{\sum_{i=1}^{n} r_i}+(-1)\cdot(2^{r_1+\cdots+r_{n-1}}+\cdots+2^{r_2+\cdots+r_n}+\cdots+(-1)^{n-1}\cdot(2^{r_1}+\cdots+2^{r_n})+(-1)^n \qquad \ldots(6)$$

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.

$$R_m^t=\binom{r_1, r_2, \ldots, r_n}{\sum_{j=1}^{m} r_{i_j}} \quad 0\leq m\leq n,\ 1\leq i\leq n,\ t\in\{1,2,\ldots,C_m{}^n\}$$

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

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

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

So equation (6) can be written in compact formula:

$$P(S)=\sum_{k=0}^{n}(-1)^{k}\cdot\sum_{t=1}^{C_k^n}2^{R_{n-k}^{t}} \qquad \ldots(7)$$

In the next sections we will introduce new theorems, as Golomb do on LFSR, to show the frequency distribution for the new proposed generator.

1st Golomb's theorem says that if LFSR with length r has maximal sequence then $N_r(0)=2^{r-1}-1$ and $N_r(1)=2^{r-1}$, where $N_r(a)$ denotes the number of bit "a" in the maximal sequence [8] s.t.:

$$P(r)=2^r-1=(2^{r-1}-1)+2^{r-1}=\sum_{a=0}^{1}N_r(a)$$

Let $N_S(a)$ be the frequency of bit "a" in S which generates from SCG then:

$$P(S)=\sum_{a=0}^{1}N_S(a)=N_{r_1}(0)\cdots N_{r_n}(0)+N_{r_1}(0)\cdots N_{r_n}(1)+\cdots+N_{r_1}(1)\cdots N_{r_n}(1) \qquad \ldots(8)$$

From this equation the act of CF will starts to distribute the ratio of "0" and "1" in S. If the terms of equation (8) rearranged s.t. $0=F(a_{i1},a_{i2},..,a_{in})$, $1\le i\le m_0$ for the 1st m_0 terms, and $1=F(a_{i1},a_{i2},..,a_{in})$, $1\le i\le m_1$ for 2nd m_1 terms $2^n=m_0+m_1$ then,

$$N_S(a)=\sum_{i=1}^{m_a}\prod_{j=1}^{n}N_{r_j}(a_{ij}) \qquad \ldots(9)$$

subject to $a=F(a_{i1},a_{i2},..,a_{in})$ s.t. $1\le i\le m_a$, a=0,1.

Where m_a denotes the number of states which are subject to the above condition [3].

4. Modified Geffe Generator

4.1 Geffe Generator

The Geffe generator [6] is defined by three maximum-length LFSRs whose lengths r_1, r_2, r_3 are pair wise relatively prime, with nonlinear combining function:

$F_3(x_1,x_2,x_3) = x_1*x_2\oplus(1\oplus x_2)*x_3 = x_1*x_2\oplus x_2*x_3\oplus x_3$

(see figure (1)).

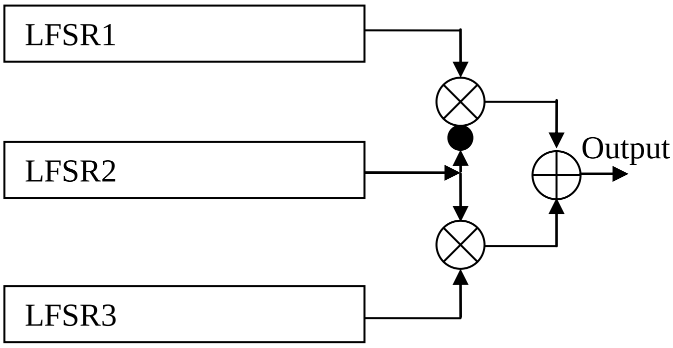

Figure (1) Geffe generator [6].

The keystream generated has period $(2^{r_1}-1)(2^{r_2}-1)(2^{r_3}-1)$ and linear complexity $LC=r_1r_2+r_2r_3+r_3$. The Geffe generator is cryptographically weak because information about the states of LFSR1 and LFSR3 leaks into the output sequence. Despite having high period and moderately high linear complexity, the Geffe generator succumbs to correlation attacks [11].

4.2 Modified Geffe Generator (5-MGG) Description

Know we would improve this generator by choosing 5 LFSR's instead of 3 LFSR's, if the output of LFSR3 is 0 then we choose the xoring of LFSR1 and LFSR2, otherwise we choose the xoring of LFSR4 and LFSR5. The CF of this generator is:

$$F_5(x_1,x_2,x_3,x_4,x_5)=(x_1 \oplus x_2)*(x_3 \oplus 1) \oplus (x_4 \oplus x_5)*x_3 \qquad \ldots(10\text{-a})$$

Or it can be written as follows:

$$F_5(x_1,x_2,x_3,x_4,x_5)=x_1 \oplus x_2 \oplus x_1x_3 \oplus x_2x_3 \oplus x_3x_4 \oplus x_3x_5 \qquad \ldots(10\text{-b})$$

so we called this system Modified-Geffe generator.

4.3 Efficiency Criteria of 5-MGG

1**. Periodicity**

From equation (5) we can find general formula to calculate the Periodicity of 5-MGG:

$$P(S)=\prod_{i=1}^{5}(2^{r_i}-1).$$

Example (1)

if r_i=2,3,…,6 for i =1,...,5, then:

$$P(S)=\text{l.c.m}(3,7,15,31,63)$$
$$=\text{l.c.m}(3^1.5^0.7^0.31^0,\ 3^0.5^0.7^1.31^0,\ 3^1.5^1.7^0.31^0, 3^0.5^0.7^0.31^1,\ 3^2.5^0.7^1.31^0)$$
$$= 3^{\max(0,1,2)}.5^{\max(0,1)}.7^{\max(0,1)}.31^{\max(0,1)}= 3^2.5^1.7^1.31^1=9765.$$

2. Randomness

From the truth table of CF of modified Geffe, notice the ratio of number of 0's to the total output of the function = 32 (2^5=32) is 0.5, this mean the number of 0's=16 and so as number of 1's, that's indicates that this generator can generates random sequence. The truth table of CF is shown in table (1).

Table (1) Truth table of CF of 5-MGG.

x_1	x_2	x_3	x_4	x_5	F_5
0	0	0	0	0	0
0	0	0	0	1	1
0	0	0	1	0	1
0	0	0	1	1	0
0	0	1	0	0	0
0	0	1	0	1	0
0	0	1	1	0	0
0	0	1	1	1	0
0	1	0	0	0	0
0	1	0	0	1	1
0	1	0	1	0	1
0	1	0	1	1	0
0	1	1	0	0	1
0	1	1	0	1	1
0	1	1	1	0	1
0	1	1	1	1	1
1	0	0	0	0	0
1	0	0	0	1	1
1	0	0	1	0	1

1	0	0	1	1	0
1	0	1	0	0	1
1	0	1	0	1	1
1	0	1	1	0	1
1	0	1	1	1	1
1	1	0	0	0	0
1	1	0	0	1	1
1	1	0	1	0	1
1	1	0	1	1	0
1	1	1	0	0	0
1	1	1	0	1	0
1	1	1	1	0	0
1	1	1	1	1	0
0.5	0.5	0.5	0.5	0.5	0.5
Correlation Probability (CP_i) for each LFSR					Ratio of "0"

Note: the shaded cells means the similarity between x_i and the output of CF.

3. Linear Complexity

The Linear Complexity is defined as the length, of the shortest LFSR (which is equivalent LFSR) that can mimic the generator output. Any sequence generated by a finite-state machine over a finite field has a finite linear complexity [9].

The Linear Complexity (LC) for the generated sequence can by calculated by:

$$LC(S)= r_1+r_2+r_1r_3+r_2r_3+r_3r_4+r_3r_5 \qquad \ldots(11)$$

Example (2)

Let's use the same information mentioned in example (1), then:

LC(S)=2+3+2*4+3*4+4*5+4*6 = 69

4. Correlation Immunity

Correlation can be defined as the relation between the sequence of CF=F_n from the key generator and the sequences that are combined each other by CF. This relation caused because of the non-linearity of the function F_n. The correlation probability CP(x), in general, represents the ratio between the number of similar binaries of two sequences to the length of the compared part of them. F_n has m^{th} order CI, if the output z of F_n is statistically independent from m output from m-sequences $(x_1,x_2,...,x_m)$, of n combined sequences s.t. $m \le n$.

Notes from table (1) (from the shaded cells) that the number of similarity between x_i and the output of CF is 16 bits from the total number 32 bits $\forall i$, then the correlation probability (CP_i) can be calculated as:

$CP_i = 16/32 = 0.5$, for i = 1,2,…,5.

Let's denotes the Correlation Immunity for the generated sequence by CI(S), then it can by calculated by:

CI(S) = 5,

since the number of immune $x_i = 5$.

This indicates that 5-MGG is immune and it cannot be attacked by correlation attack or fast correlation attack, while Geffe generator is not immune [12].

5. Estimation of Frequency Postulate for 5-Threshold Generator

5.1 Theoretical Estimation of $N_S(1)$

Let $N_S(1)$ be the number of bit (1) in the sequence S generated from 5-MGG.
Recall equations (8) and (9), and when n=5:
P(S)= $N_{r_1}(0).N_{r_2}(0).N_{r_3}(0).N_{r_4}(0).N_{r_5}(0)+...+N_{r_1}(1).N_{r_2}(1).N_{r_3}(1).N_{r_4}(1).N_{r_5}(1)$
Suppose that x_i, i=1,…,5 is the output of LFSR's of MGG. and we know that the x_3 will be control on the ($x_1 \oplus x_2$) and ($x_4 \oplus x_5$).
To calculate $N_S(1)$ we add all terms which contain the following states:
$N_{r_1}(0)\cdot N_{r_2}(1)\cdot N_{r_3}(0)\cdot\cdot, N_{r_1}(1)\cdot N_{r_2}(0)\cdot N_{r_3}(0)\cdot\cdot,$
and
$\cdot\cdot N_{r_3}(1)\cdot N_{r_4}(0)\cdot N_{r_5}(1), \cdot\cdot N_{r_3}(1)\cdot N_{r_4}(1)\cdot N_{r_5}(0).$
Then:

$$N_S(1)=[N_{r_1}(0)\cdot N_{r_2}(1)+N_{r_1}(1)\cdot N_{r_2}(0)]\cdot N_{r_3}(0)\cdot\sum_{i=0}^{1}\sum_{j=0}^{1}N_{r_4}(i)\cdot N_{r_5}(j)$$
$$+[\sum_{i=0}^{1}\sum_{j=0}^{1}N_{r_1}(i)\cdot N_{r2}(j)]\cdot N_{r_3}(1)\cdot[N_{r_4}(0)\cdot N_{r_5}(1)+N_{r_4}(1)\cdot N_{r_5}(0)]\cdot$$

After simplify the above equation we obtain:

$$N_S(1)=[2^{r_1+r_2-1}-(2^{r_1-1}+2^{r_2-1})]\cdot(2^{r_3-1}-1)[2^{r_4+r_5}-(2^{r_4}+2^{r_5})+1]$$
$$+[2^{r_1+r_2}-(2^{r_1}+2^{r_2})+1]\cdot 2^{r_3-1}\cdot[2^{r_4+r_5-1}-(2^{r_4-1}+2^{r_5-1})] \quad \ldots(12)$$

Note: we believe that formula (12) is the simplest form.

5.2 Theoretical Estimation of NS(0)

in the same way we can calculate $N_S(0)$.
Let $N_S(0)$ be the number of bit (0) in the sequence S generated from 5-MGG.
Recall equations (8) and (9), and when n=5:
To calculate $N_S(0)$ we add all terms which contain the following states:
$N_{r_1}(0)\cdot N_{r_2}(0)\cdot N_{r_3}(0)\cdot\cdot, N_{r_1}(1)\cdot N_{r_2}(1)\cdot N_{r_3}(0)\cdot\cdot,$
and
$\cdot\cdot N_{r_3}(1)\cdot N_{r_4}(0)\cdot N_{r_5}(0), \cdot\cdot N_{r_3}(1)\cdot N_{r_4}(1)\cdot N_{r_5}(1).$
Then:

$$N_S(0)=[N_{r_1}(0)\cdot N_{r_2}(0)+N_{r_1}(1)\cdot N_{r_2}(1)]\cdot N_{r_3}(0)\cdot\sum_{i=0}^{1}\sum_{j=0}^{1}N_{r_4}(i)\cdot N_{r_5}(j)$$
$$+[\sum_{i=0}^{1}\sum_{j=0}^{1}N_{r_1}(i)\cdot N_{r2}(j)]\cdot N_{r_3}(1)\cdot[N_{r_4}(0)\cdot N_{r_5}(0)+N_{r_4}(1)\cdot N_{r_5}(1)]\cdot$$

After simplify the above equation we obtain:

$$N_S(0)= [2^{r_1+r_2-1}-(2^{r_1-1}+2^{r_2-1})+1]\cdot(2^{r_3-1}-1)[2^{r_4+r_5}-(2^{r_4}+2^{r_5})+1]$$
$$+[2^{r_1+r_2}-(2^{r_1}+2^{r_2})+1]\cdot 2^{r_3-1}\cdot[2^{r_4+r_5-1}-(2^{r_4-1}+2^{r_5-1})+1] \quad \ldots(13)$$

Remark (1)

Notes that if we add formulas (12) and (13) we obtain:

$$P(S)= 2^{R_5^1}-\sum_{i=1}^{C_1^5}2^{R_4^i}+\sum_{i=1}^{C_2^5}2^{R_3^i}-\sum_{i=1}^{C_3^5}2^{R_2^i}+\sum_{i=1}^{C_4^5}2^{R_1^i}-1 \quad \ldots(14)$$

Example (3):

Table (2) shows the values of $N_S(0)$ and $N_S(1)$ for different r_i of 5-MGG.

Table (2) the values of $N_S(0)$ and $N_S(1)$ for different r_i of 5-MGG.

Ex	r_i					$P(S_i)$					$N_S(a)$		P(S)
	i					i					$N_S(0)$	$N_S(1)$	
	1	2	3	4	5	1	2	3	4	5			
1	2	3	5	7	11	3	7	31	127	2047	86569845	82669974	169239819
2	2	5	7	9	11	3	31	127	511	2047	6210205905	6144300882	12354506787
3	3	5	7	8	11	7	31	127	255	2047	7209141779	7176242836	14385384615

5.3 Calculate the Proportion of $N_S(1)$ to P(S)

$$\frac{N_S(1)}{P(S)}=\frac{[2^{r_1+r_2-1}-(2^{r_1-1}+2^{r_2-1})]\cdot(2^{r_3-1}-1)[2^{r_4+r_5}-(2^{r_4}+2^{r_5})+1]}{\prod_{i=1}^{n}(2^{r_i}-1)}+$$

$$\frac{[2^{r_1+r_2}-(2^{r_1}+2^{r_2})+1]\cdot 2^{r_3-1}\cdot[2^{r_4+r_5-1}-(2^{r_4-1}+2^{r_5-1})]}{\prod_{i=1}^{n}(2^{r_i}-1)}$$

$$\frac{N_S(1)}{P(S)}=\frac{2^{\sum_{i=1}^{5}r_i-1}-(2^{\sum_{i=1}^{4}r_i-1}+\cdots+2^{\sum_{i=2}^{5}r_i-1})+(2^{r_1+r_2+r_3-2}+\cdots+2^{r_3+r_4+r_5-2})-(2^{r_1+r_2-1}+\cdots+2^{r_3+r_5-2})+(2^{r_1-1}+2^{r_2-1})}{\prod_{i=1}^{n}(2^{r_i}-1)}$$

As r_i be as large as possible, for 1≤i≤5, then $2^{r_i}-1\rightarrow 2^{r_i}$ (ignore 1), then $P(S)\approx\prod_{i=1}^{5}2^{r_i}=2^{\sum_{i=1}^{5}r_i}$, then:

$$\frac{N_S(1)}{P(S)}=\frac{2^{\sum_{i=1}^{5}r_i-1}}{2^{\sum_{i=1}^{5}r_i}}-\frac{(2^{\sum_{i=1}^{4}r_i-1}+\cdots+2^{\sum_{i=2}^{5}r_i-1})}{2^{\sum_{i=1}^{5}r_i}}+\frac{(2^{r_1+r_2+r_3-2}+\cdots+2^{r_3+r_4+r_5-2})}{2^{\sum_{i=1}^{5}r_i}}-\frac{(2^{r_1+r_2-1}+\cdots+2^{r_3+r_5-2})}{2^{\sum_{i=1}^{5}r_i}}+\frac{(2^{r_1-1}+2^{r_2-1})}{2^{\sum_{i=1}^{5}r_i}}$$

$$\frac{N_S(1)}{P(S)}=\frac{1}{2}-\frac{1}{2^{r_5+1}+\cdots+2^{r_1+1}}+\frac{1}{2^{r_4+r_5+2}+\cdots+2^{r_1+r_2+2}}-\frac{1}{2^{r_3+r_4+r_5+1}+\cdots+2^{r_1+r_2+r_4+1}}+\frac{1}{2^{\sum_{i=2}^{5}r_i+1}+2^{r_1+r_3+r_4+r_5+1}}$$ as $r_i\rightarrow\infty$,

then: $2^{r_i+a}\rightarrow\infty$, and $\frac{1}{2^{r_i+a}}\rightarrow 0$, where a=1,2, for 1≤i≤5, and so on for all dominator of the above equation, then:

$$\therefore\ \frac{N_S(1)}{P(S)}\approx\frac{1}{2}\text{-0+0-0+0=0.5} \qquad \ldots(15)$$

Remark (2)

Notes from equation (15) that: $N_S(0)\approx N_S(1)$.

Example (3):

Table (3) shows the proportion of $N_S(1)$ to P(S) for various 5-MGG.

Table (3) the proportion of $N_S(1)$ to P(S) for various 5-MGG.

Ex	r_i					$N_S(a)$		P(S)	Proportion of $N_S(a)$		
	r_1	r_2	r_3	r_4	r_5	$N_S(0)$	$N_S(1)$		Exp.	Observed	
										$N_S(0)$	$N_S(1)$
1	2	3	5	7	11	86569845	82669974	169239819	0.5	0.512	0.488
2	2	5	7	9	11	6210205905	6144300882	12354506787		0.503	0.497
3	4	5	11	8	11	7209141779	7176242836	14385384615		0.501	0.499

6. Applying of Chi-Square Test on 5-MGG

In this section we will apply chi-square test on the results gotten from calculations of frequency postulate on 5-MGG.

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, p_i the probability of occurs of the category c_i, then the expected frequency E_i of the category c_i is $E_i=P(S)\cdot p_i$, the T (chi-square value) can be calculated as follows [4]:

$$T=\sum_{i=1}^{K}\frac{(N(c_i)-E_i)^2}{E_i} \qquad \ldots(16)$$

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 $T\leq T_0$ 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. Let $N(c_a)=\mu_a$=Ns(a)/P(S), for a=0,1, μ_a is the mean of $N_S(a)$. To apply Hypothesis test:

H_0: $\mu_0 \approx \mu_1$, while,

H_1: there are a big difference between μ_0 and μ_1.

Then we apply the hypothesis test for the difference between two means using chi-square distribution:

$$T=\frac{(\mu_0-\mu_1)^2}{\mu_0+\mu_1}\sim \chi^2(1), \text{ s.t. } \upsilon=1.$$

In order to test our results we have to suggest an example (we choose worst case from table (3)). Let r_1=2, r_2=3, r_3=5, r_4=7 and r_5=11. P(S)=169239819, E_i=84619909.5. In **Frequency** test υ= 1, with α=0.05%, then T_0=3.84 (see chi-square table). it's clear that $\mu_0 + \mu_1$=1. From equation (12), we get $N_S(0)$= 86569845 and $N_S(1)$= 82669974, then: μ_0 =0.512 and μ_1 =0.488, then:

T=0.000531<<T_0=3.81, then S generated from 5-MGG **passes** the test this means we accept the hypothesis H_0 and refuse H_1.

7. Conclusions

1. In this work we prove deterministically that the Modified Geffe cryptosystem has good statistical frequency properties.
2. We notice that if we apply chi-square on the values of $N_S(a)$ not on means of them, then the results fail to passes the test since the difference between $N_S(0)$ and E_i is too big and when the difference squared will be bigger although its divided on E_i added to the result of the same thing done for $N_S(1)$, so T will fail to passes the test.
3. These theoretical studies can be applied on other kind of SCG,s to calculate the frequency of these SCG,s which are use combining functions with some combinations of variables.
4. As future work we may apply other properties of randomness criterion like, run and autocorrelation on non-linear SCG.
5. The frequency test not enough to judge on the sequence that has good randomness tests we still have the run and autocorrelation test.
6. We recommended not using MGG in cryptography since it's still weak even it passes the randomness tests.

References

[1]. Ahmad I. A., "**Using Neural Networks In Cryptanalysis Stream Cipher Systems**", Mansoura Journal for Computer Science and Information Systems, Volume 1, Number 0, Jan. 2005.

[2]. Al-Algeelee S. A., "**Use of GA in Cryptanalysis of a Class of Stream Cipher System**", PH. D. Thesis, University of Technology, 1998.

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

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

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

[6]. Geffe, P. R., "**How to Protect Data with Ciphers that are Really Hard to Break**", Electronics pp. 99-101, Jan. 4, 1973.

[7]. Golić, J., "**On the Security of Shift Register Based Keystream Generators**", R. Anderson, editor, Fast Software Encryption, Cambridge Security Workshop (LNCS 809), 90–100, Springer-Verlag, 1994.

[8]. Golomb, S. W., "**Shift Register Sequences**" San Francisco: Holden Day 1967 (Reprinted in 1982).

[9]. Massey, J. L., "**Cryptography and System Theory**", Proceedings of the 24th Allerton Conference on Communication, Control, and Computers, 1-3 Oct. 1986.

[10]. Rueppel, R. A., "**Analysis and Design of Stream Ciphers**" Springer-Verlag, Berlin, 1986Whitesitt, J. E., "**Boolean Algebra and its Application**", Addison-Wesley, Reading, Massachusetts, April, 1995.

[11]. Schneier B., "**Applied Cryptography, Second Edition: Protocols, Algorithms, and Source Code in C** ", Wiley Computer Publishing, John Wiley & Sons, Inc., 1997.

[12]. Siegenthaler, T., "**Correlation-Immunity of Nonlinear Combining Functions for Cryptographic Applications**", IEEE Transactions on Information Theory, v. IT-30, n. 5, pp. 776-780, Sep. 1984.

[13]. Staffelbach, O. and Meier, W., "**Fast Correlation Attacks on Certain Stream Ciphers**", Journal of Cryptology, 1 (1989), 159–176.

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