GENERALIZE THE RANDOMNESS TESTS TO TEST THE DIGITAL SEQUENCES PRODUCED FROM DIGITAL STREAM CIPHER SYSTEMS

Faez H. A. Al-Azawi, Sahar A. M. Al-Bassam, *Mahmood A. Shamran

Department of Mathematics, College of Science, University of Al Mustansiriyah. Baghdad- Iraq.
*Department of Mathematics, College of Science for Women, University of Baghdad. Baghdad- Iraq.

Abstract

In this paper, first, the Golomb's postulates are generalized to construct a good mathematical base, and then generalize the binary standard randomness tests to be suitable to be applied on digital sequences. This paper includes some tables describe the tests results of the digital sequences generated from some digital generators, like the Multiplicative Cyclic Group System (MCGS) generator.

تعميم اختبارات العشوائية لاختبار المتتابعات الرقمية الناتجة من نظم التشفير الانسيابي الرقمية

فائز حسن علي العزاوي، سحر احمد محمد البصام، *محمود عريبي شمران الرفيعي

قسم الرياضيات ، كلية العلوم، الجامعة المستنصرية. بغداد-العراق.

*قسم الرياضيات، كلية العلوم للبنات، جامعة بغداد. بغداد-العراق.

الخلاصة

في هذا البحث تم تعميم بديهيات Golomb ثم تم تعميم الاختبارات الثنائية العشوائية وتحويرها لتلائم اختبار المتتابعات الرقمية. في هذا البحث تم عرض نتائج الاختبارات للمتتابعات الرقمية المولدة من بعض المولدات العشوائية مثل مولد الزمرة الضربية الدوارة.

1. Introduction

In general, any sequence generated from any generator considered a statistical experiment, for this reason the randomness tests called **statistical random tests**. Since they are statistical experiments then the proof of that the sequence is random is called **probabilistic** proof, that means when the generated sequence is random in high ratio for all experiments, then we can judge that the sequence is random, and vice versa. The randomness judgment done by two conditions [1]:

1. The length of the tested sequence must be as high as possible.
2. The number of repeating the test (experiments) must be as high as possible.

Now we will introduce some relevant basic concepts.

Cryptography is the study of principles and techniques by which information can be concealed in ciphertexts and later revealed by legitimates users employing the secret key, but in which it is either impossible or computationally infeasible for an unauthorized

person to do so. **Cryptanalysis** is the science (and art) of recovering information from ciphertexts without knowledge of the key. Both terms are subordinate to the more general term **Cryptology**. The cryptography concerned in **Encryption** and **Decryption** processes. Figure 1 demonstrates the relation between these concepts [2].

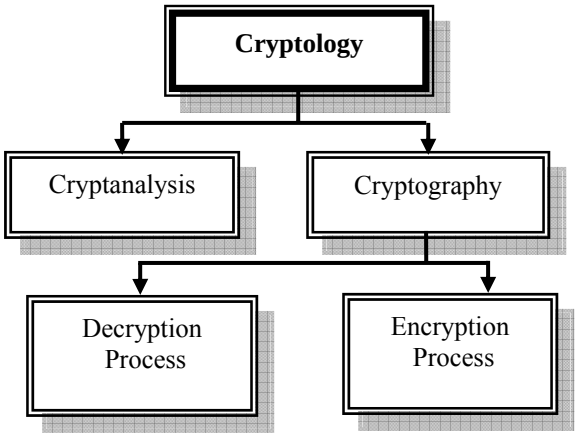

Fig. 1: Cryptology branches

In **stream ciphers**, the message units are bits, and the key is usual produced by a **random bit generator**. The plaintext is encrypted on a bit-by-bit basis.

The key is fed into random bit generator to create a long sequence of binary signals. This "key-stream" k is then mixed with plaintext m, usually by a bit wise XOR (Exclusive-OR modulo 2 addition) to produce the ciphertext stream, using the same random bit generator and seed. Stream ciphers are generally faster than block ciphers in hardware, and have less complex hardware circuitry. They are also more appropriate, and in some cases mandatory (e.g., in some telecommunications applications), when buffering is limited or when characters must be individually processed as they are received [3].

A **feedback shift register** is made up of two parts: a shift register and a **feedback function**. The shift register is a sequence of bits, (the length of a shift register is figured in bits). Each time a bit is needed; all of the bits in the shift register are shifted 1 bit to the right [4]. Cryptographers have liked stream ciphers made up of shift registers. We will only touch on the mathematical theory. Solomon Golomb [5], an NSA mathematician, wrote a book with Selmers results and some of his own [6]. The simplest kind of feedback shift register is a **Linear Feedback Shift Register** (LFSR). The feedback function is simply the XOR of certain bits in the register.

Universal tests were presented by Schrift and Shamir in 1993 [7] for verifying the assumed properties of a pseudorandom generator whose output sequences were not necessarily uniformly distributed.

Gustafson et al. in 1994 [12] describe a computer package which implements various statistical tests for assessing the strength of a pseudorandom bit generator.

In 1996, Gustafson [13] considered alternative statistics for the runs test and the autocorrelation test. Gustafson, Dawson, and Golić [14] proposed a new repetition test which measures the number of repetitions of l-bit blocks. The test requires a count of the number of patterns repeated, but does not require the frequency of each pattern.

2. Randomness

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 [7, 8].

Definition (1): A random bit generator is a device or algorithm which outputs a sequence of statistically independent and unbiased binary digits.

Remark (1): (random bits vs. random numbers) A random bit generator can be used to generate (uniformly distributed) random numbers. For example, a random integer in the interval [0,n] can be obtained by generating a random bit sequence of length log2|n+1| bits, and converting it to an integer; if the resulting integer exceeds n, one option is to discard it and generate a new random bit sequence.

Definition (2): A Pseudo Random Bit Generator (PRBG) is a deterministic algorithm which, given a truly random binary sequence of length k, outputs a binary sequence of length L , k which "appears" to be random. The input to the PRBG is called the seed, while the output of the PRBG is called a pseudorandom bit sequence.

Remark (2): The $\chi 2$ (chi-square) distribution can be used to compare the goodness-of-fit of the observed frequencies of events to their expected frequencies under a hypothesized distribution. The $\chi 2$ distribution with υ degrees of freedom arises in practice when the squares of υ independent random variables having standard normal distributions are summed [13].

3. Golomb′s Concept of Randomness

Definition (3): Let S be a periodic sequence of period N. Golomb′s randomness postulates are the following[3]:

R1: In the cycle S^N of S, the number of 1′s differs from the number of 0′s by at most 1.

R2: In the cycle S^N at least half the runs have length 1, at least one-fourth have length 2, at least one-eighth have length 3, etc., as long as the number of runs so indicated exceeds 1. Moreover, for each of these lengths, there are (almost) equally many gaps and blocks.

R3: The autocorrelation function C(t) is two-valued. That is for some integer K:

$$N.C(t)=\sum_{i=0}^{N-1}(2s_i-1)\cdot(2s_{i+t}-1)=\begin{cases}N & ,t=0\\ K & ,0\le t\le N-1\end{cases}$$

Definition (4): A binary sequence which satisfies Golomb′s randomness postulates is called a pseudo-noise sequence or a pn-sequence.

Pseudo-noise sequences arise in practice as output sequences of maximum-length linear feedback shift registers.

4. Generalize the Golomb's Postulates

We showed before that the Golomb's postulates are applied on binary sequences. In this section we try to generalize these postulates in order to be suitable to be applied on digital sequences. We can define the digital sequence which is satisfied the new generalized postulates, as **Pseudo Random Digital Sequence** (PRDS).

Let S be a sequence has m distinct digits (0..m-1) with period P. Let i_j be the digit j of S, s.t. $0\le i_j\le m-1$, j=0, 1,… ,. In the next subsections we will introduce the new digital postulates.

4.1 Digital Frequency Postulate

Its obvious that if the frequency N_i of each distinct digit i is approximate to other frequencies, then the digital sequence is satisfies this postulate, so must be:

$N_0\approx N_1\approx\ldots\approx N_{n-1}$

Statistically, N_i represents the observed number occurrence of digit i, $0\le i\le m-1$.

The expected number of occurrence is:

$$E^*_F=\frac{P}{m} \qquad \ldots(1)$$

Then, $N_i\approx E^*_F$, $\forall i$

Where P is the period of the sequence.

4.2 Digital Run Postulate

The digital run here can be defined as the number of similar digits which are lie between two different digits. Now we can depend on mathematical deduction to deduce the two new conditions of run postulates:

The number (R_{ij}) of kind i runs with length j is approximately equal to 1/m of the number of runs of length j-1; $R_{ij}\approx(1/m)R_{ij-1}$, where $2\le j\le M_i$, M_i denotes the length of maximum run of kind i. The all kinds of runs of length j are approximate to each other, s.t. $R_{0j}\approx R_{1j}\approx\ldots\approx R_{m-1,j}$, where $1\le j\le M_i$, its obvious that:

$$\sum_{j=1}^{M_i} j.R_{ij}=N_i, 0\le i\le m-1 \qquad \ldots(2)$$

$R_{ij}\approx E_{Rj}$, $\forall i$, where E_{Rj} is the expected number of the runs with length j which can be calculated from the next theorem.

Theorem (1): *Let S be a sequence satisfies the run postulate, then the expected number of runs with length j is:*

$$E_{Rj}=\frac{P(m-1)^2}{m^{j+2}}, 1\le j\le M, \text{ where}$$

$M=max(M_0,M_1,\ldots,M_{m-1})$.

Proof

From equation (2), and $N_i\approx E^*_F$, *then*

$$E^*_F=\frac{P}{m}=E_{R1}+2E_{R2}+\ldots+M.E_{RM}=\sum_{j=1}^{M} j\cdot E_{Rj} \qquad \ldots(3)$$

And since $R_{i2}=\frac{R_{i1}}{m}$ then $E_{R2}=\frac{E_{R1}}{m}$, since S satisfies the runs postulate, so in general,

$$E_{Rj}=\frac{E_{R1}}{m^{j-1}},\ 2\le j\le M \qquad \ldots(4)$$

substitute equation (3) in equation (4), we get:

$$\frac{P}{m}=\sum_{j=1}^{M}\frac{j\cdot E_{R1}}{m^{j-1}}=mE_{R1}\sum_{j=1}^{M}\frac{j}{m^j} \qquad \ldots(5)$$

By using the ratio test:

as $M\to\infty$ then $S'=\lim_{j\to\infty}\sum_{j=1}^{M}\frac{j}{m^j}=\sum_{j=1}^{\infty}\frac{j}{m^j}$ is convergence series.

$$\rho=\lim_{j\to\infty}\frac{u_{j+1}}{u_j}=\lim_{j\to+\infty}\frac{j+1}{m^{j+1}}\cdot\frac{m^j}{j}=$$

$$\frac{1}{m}\lim_{j\to+\infty}\frac{j+1}{j}=\frac{1}{m}$$

So that $\rho<1$ since $m\ge2$

For S′ we have,

$s_1=\frac{1}{m}$

$s_2=\frac{1}{m}+\frac{2}{m^2}$, so

$$s_M=\frac{1}{m}+\frac{2}{m^2}+\ldots+\frac{M}{m^M} \qquad \ldots(6)$$

$$\frac{1}{m}s_M=\frac{1}{m^2}+\frac{2}{m^3}+\ldots+\frac{M}{m^{M+1}} \quad \ldots(7)$$

Subtract equation (7) from (6)

$$s_M-\frac{1}{m}s_M=\frac{1}{m}+\frac{1}{m^2}+\ldots+\frac{1}{m^M}-\frac{M}{m^{M+1}}$$

$$s_M(\frac{m-1}{m})=\sum_{k=1}^{M}\frac{1}{m^k}-\frac{M}{m^{M+1}}=\frac{1}{m}(\sum_{k=1}^{M}\frac{1}{m^{k-1}}-\frac{M}{m^M}),$$

then

$$s_M=\frac{1}{m-1}(\sum_{k=1}^{n}\frac{1}{m^{k-1}}-\frac{M}{m^M})$$

let $n\to\infty$, the series

$$S'=\lim_{M\to\infty}s_M=\frac{1}{m-1}(\sum_{k=1}^{\infty}\frac{1}{m^{k-1}}-\lim_{M\to\infty}\frac{M}{m^M})\ldots(8)$$

Notice that $\lim_{M\to\infty}\frac{M}{m^M}=\lim_{M\to\infty}M\cdot\lim_{M\to\infty}\frac{1}{m^M}=\lim_{M\to\infty}M.0=0$

Since the series is $\sum_{k=1}^{\infty}\frac{1}{m^{k-1}}$ geometric series with a=1 and $r=\frac{1}{m}$ [4], and since $|r|=\frac{1}{m}<1$ because $m\geq 2$, then the series is convergence series and the sum

$$S=\sum_{k=1}^{\infty}\frac{1}{m^{k-1}}=\frac{a}{1-r}=\frac{1}{1-1/m}=\frac{m}{m-1} \quad \ldots(9)$$

substitute equation (9) in equation (8)

$$S'=\frac{1}{m-1}\cdot\frac{m}{m-1}=\frac{m}{(m-1)^2} \quad \ldots(10)$$

Substitute equation (10) in equation (5) we get:

$$\frac{P}{m}=m.E_{R1}.\frac{m}{(m-1)^2}$$

$$\therefore E_{R1}=\frac{P(m-1)^2}{m^3}=\frac{P(m-1)^2}{m^{1+3}}$$

In general, and by using the mathematical induction, we have

$$E_{Rj}=\frac{P(m-1)^2}{m^{j+2}}, \text{ where } 1\leq j\leq M$$

Theorem (2): The expected number of total number E^*_R of runs of any kind i, where $0\leq i\leq m-1$ is $E^*_R=\frac{P(m-1)}{m^2}$.

Proof

$$E^*_R=\frac{P(m-1)^2}{m^3}+\frac{P(m-1)^2}{m^4}+\ldots+\frac{P(m-1)^2}{m^{M+2}}$$

$$=\frac{P(m-1)^2}{m^3}\sum_{j=1}^{M}\frac{1}{m^{j-1}} \quad \ldots(11)$$

The series $\sum_{j=1}^{M}\frac{1}{m^{j-1}}$ is geometric convergence series as $M\to\infty$, then

$$\sum_{j=1}^{\infty}\frac{1}{m^{j-1}}=\frac{m}{m-1} \quad \ldots(12)$$

Using equation (12) in (11), we get:

$$E^*_R=\frac{P(m-1)^2}{m^3}\cdot\frac{m}{m-1}=\frac{P(m-1)}{m^2}$$

By using theorem (2) we can estimate the expected number E^*_{SR} of sum of E^*_R of sum of all runs by:

$$E^*_{SR}=\sum_{i=0}^{m-1}E^*_R=m.\frac{P(m-1)}{m^2}=\frac{P(m-1)}{m}$$

Notice that $E^*_{SR}=P-\frac{P}{m}=P-E^*_F$

4.3 Digital Auto correlation Postulate

As mentioned before, that this postulate found to specify that if the tested digital sequence has a repetition with itself. Let $N_0(\tau)$ denotes the number of similar digits in S after shifting it by τ, let $N_1(\tau)$ denotes the number of distinct digits in S after shifting it by τ, where $1\leq\tau\leq P-1$, s.t.

$N_0(\tau)=\#\{s_i=s_{i+\tau}:\forall 1\leq i\leq P\}$

$N_1(\tau)=\#\{s_i\neq s_{i+\tau}:\forall 1\leq i\leq P\}$,

s.t. $1\leq\tau\leq P-1$.

Where $N_0(\tau)+N_1(\tau)=P-\tau$.

The probability of similarity of one digit from n digits is 1/m, then the expected number of similarity is $E^*_0=\frac{P-\tau}{m}$, and expected number of difference is:

$$E^*_1=\frac{(P-\tau)\cdot(m-1)}{m}=(P-\tau)-E^*_0.$$

5. Evolving the Digital Randomness Test

In this section we will reformulate the three main testing laws to be suitable to apply on digital sequence. We called the new digital

randomness tests by the **Main Digital Standard Randomness Tests** (MDSRT).
Let S be the digital sequence, which want to be tested, with length L has the element(s) i, ranged 0≤i≤m-1.

5.1 Digital Frequency Test

Let N_i represents the observed number of occurrence of digit i, where 0≤i≤m-1, and the expected number of occurrence of digit i is $E^*_F=\frac{L}{m}$, then:

$$T_F=\sum_{i=0}^{m-1}\frac{(N_i-E^*_F)^2}{E^*_F}=\sum_{i=0}^{m-1}\frac{(N_i-L/m)^2}{L/m} \quad \ldots(13)$$

With freedom degree υ=m-1.
The following Lemma (1) gives more simple formula for T_F of frequency test using formula (13).

Lemma (1): For frequency test of digital sequence S, $T=\frac{m}{L}\sum_{i=0}^{m-1}N_i^2-L$.

Proof:

$$T=\sum_{i=0}^{m-1}\frac{(N_i-L/m)^2}{L/m}=$$

$$\frac{m}{L}(\sum_{i=0}^{m-1}N_i^2-2\frac{L}{m}\sum_{i=0}^{m-1}N_i+\sum_{i=0}^{m-1}\frac{L^2}{m^2})=$$

$$\frac{m}{L}\sum_{i=0}^{m-1}N_i^2-2L+L$$

$$\therefore T=\frac{m}{L}\cdot\sum_{i=0}^{m-1}N_i^2-L \quad \ldots(14)$$

5.2 Digital Run Test

Let R_{ij} represents the observed number of runs of kind i with length j, and let E_{Rj} be the expected number of runs of any kind with length j, then

$$T_{Ri}=\sum_{j=1}^{M_i}\frac{(R_{ij}-E_{Rj})^2}{E_{Rj}}=\sum_{j=1}^{M_i}\frac{(R_{ij}-\frac{L\cdot(m-1)^2}{m^{j+2}})^2}{\frac{L\cdot(m-1)^2}{m^{j+2}}},$$

$0\le i\le m-1 \quad \ldots(15)$

With freedom degree $\upsilon_i=M_i-1$.
Formula (15) can be reformulated in another face:

$$R_i=\frac{1}{L}\cdot(\frac{m}{m-1})^2\cdot\sum_{j=0}^{M_i}m^j\cdot R_{ij}^2-2\sum_{j=0}^{M_i}R_{ij}+\frac{L(m-1)}{m^2} \quad ..(16)$$

5.3 Digital Auto correlation Test

Let $N_0(\tau)$ and $N_1(\tau)$ represent the number of similar and distinct digits of the sequence S respectively, after shifting it by τ, then the expected number of similarity and difference respectively are:

$E_0(\tau)=\frac{L-\tau}{m}$ and $E_1(\tau)=\frac{(m-1)\cdot(L-\tau)}{m}$, the following lemma proofs that Chi square of auto correlation test for the digital sequence S is:

$$T_A(\tau)=\frac{(mN_0(\tau)-(L-\tau))^2}{(m-1)(L-\tau)} \quad \ldots(17)$$

With freedom degree υ=1.

Lemma (2): The Chi square of auto correlation test for the digital sequence S shifted by τ is:

$$T_A(\tau)=\frac{(mN_0(\tau)-(L-\tau))^2}{(m-1)(L-\tau)}.$$

Proof: for simplicity, let n=m-1, then m=n+1, L′=L-τ, $N_0=N_0(\tau)$ and $N_1=N_1(\tau)$.

$$T_A(\tau)=\sum_{i=0}^{1}\frac{(N_1(\tau)-E_1(\tau))^2}{E_1(\tau)}=\frac{(N_0-\frac{L'}{m})^2}{\frac{L'}{m}}+$$

$$\frac{(N_1-\frac{nL'}{m})^2}{\frac{nL'}{m}}$$

$$=\frac{n(N_0-\frac{L'}{m})^2+(N_1-\frac{nL'}{m})^2}{\frac{nL'}{m}}=$$

$$\frac{nN_0^2-2nN_0\frac{L'}{m}+n\frac{L'^2}{m^2}+N_1^2-2n\frac{L'}{m}+n^2\frac{L'^2}{m^2}}{\frac{nL'}{m}} \quad (18)$$

Since $N_1=L'-N_0$, then

$$-2n\,L'(N_0+N_1)=-2nL'^2 \quad \ldots(19)$$

And

$$n\frac{L'^2}{m}+n^2\frac{L'^2}{m}=nL'^2 \quad \ldots(20)$$

And

$$m(nN_0^2+N_1^2)=m^2N_0^2+mL'^2+2mN_0L' \quad \ldots(21)$$

adding equation (19) to (20) and adding the result to (21), then substitute the final result in (18) using the fact that m-n=1, we get:

$$T_A(\tau)=\frac{m^2N_0^2-2mN_0L'+L'^2}{nL'}=\frac{(mN_0-L')^2}{nL'}$$

$$\therefore T_A(\tau)=\frac{(mN_0(\tau)-(L-\tau))^2}{(m-1)(L-\tau)}$$

Remark (3): In the Lemma (1) and (2) note that we have no need to calculate the expected value of any sample in the three tests, so we don't need the equations (13) and (15) any more.

6. Implementation of MDSRT on Digital Sequences

In this section we will show how we can applied the MDSRT on arbitrarily sequence with 30 digits length, using the new statistic digital laws. Let the sequence S consists of the following digits:

S="102311033321000121221301112301".

Its obvious that m=4.

1. **Frequency test**: the following table shows the frequency values N_i:

i	**0**	**1**	**2**	**3**
$\mathbf{N_i}$	7	11	6	6

By using formula (14) we get:

$$T_F=\frac{4}{30}(49+121+36+36)-30=2.266$$

This value compared with $T_0=7.81$ where $\upsilon=3$. The sequence S passes this test.

2. **Run test**: the following table shows the run values R_{ij}:

	i	**Runs (R_{ij})**			
j		**0**	**1**	**2**	**3**
Length	**1**	4	6	4	3
	2	0	1	1	0
	3	1	1	0	1

$M_0=3$, $M_1=3$, $M_2=2$, $M_3=3$ and $M=3$, by using formula (17) we get:

$$T^*_0=0.059(4(16)+4^2(0)+4^3(1))-2(4+0+1)+5.625=3.366$$

This value compared with $T_0=5.99$ at $\upsilon_0=2$.

Using the same formula, we find:

$T^*_{R1}=2.841$ compared with $T_1=5.99$ at $\upsilon_1=2$.

$T^*_{R2}=0.345$ compared with $T_2=3.84$ at $\upsilon_2=1$.

$T^*_{R3}=3.525$ compared with $T_3=5.99$ at $\upsilon_3=2$.

Since $T^*_{Ri} \le T_{Ri}\ \forall i$, then the sequence S passes this test.

1. **Auto correlation test**: first, let us shift the sequence S by one shift ($\tau=1$), then:

1	0	2	3	1	1	0	3	3	3	2	1	0	0	0	1
	1	0	2	3	1	1	0	3	3	3	2	1	0	0	0
	→	2	1	2	2	1	3	0	1	1	1	2	3	0	1
	→	1	2	1	2	2	1	3	0	1	1	1	2	3	0

We notice that there are 8 similar digits (shaded cells), so $N_0(1)=8$, and by using formula (17), then:

$$T_A(1)=\frac{(4(8)-29)^2}{3(29)}=0.103$$

So we can find the other $N_0(\tau)$ by the following table:

τ	**1**	**2**	**3**	**4**	**5**	**6**	**7**
$N_0(\tau)$	8	5	3	5	6	9	4
$T_A(\tau)$	0.1	0.76	2.78	0.46	0.01	2.0	0.71
τ	**8**	**9**	**10**	**11**	**12**	**13**	**14**
$N_0(\tau)$	6	5	4	3	5	4	5
$T_A(\tau)$	0.06	0.02	0.27	0.86	0.07	0.02	0.33
τ	**15**	**16**	**17**	**18**	**19**	**20**	**21**
$N_0(\tau)$	2	5	2	4	3	4	2
$T_A(\tau)$	0.56	0.86	1.26	1.78	0.76	1.2	0.33
τ	**22**	**23**	**24**	**25**	**26**	**27**	**28**
$N_0(\tau)$	1	1	4	2	0	1	0
$T_A(\tau)$	0.0	0.43	2.0	3.27	0.0	1.0	0.68

Notice that $0.016 \le T_A(\tau) \le 3.266$, for $1 \le \tau \le 29$, compare all values of $T_A(\tau)$ with $T_A(\tau)=3.84$, where $\upsilon=1$. So S passes the test.

7. Testing the Digital Sequences Generated from MCGS

The sequences generated from MCG system mentioned in [14] is being tested for binary randomness test (m=2) only. In this paper we can test these sequences by using the MDSRT.

Now we will test three different digital sequences for m=3, 5 and 7, with different length L=2000, 5000 and 8000 digits respectively. All these sequences are generated from different linear MCGS's (CF is XOR function) have the initial keys described in table 1.

Table 1: the Three MCGS's initial key.

MCGS	n	q_i	α_{1i}	α_{2i}	k_i	m
1	2	101 997	2 7	8 855	1 1	3
2	3	199 1103 3607	3 5 5	44 125 3125	1 1 1	5
3	5	149 509 1051 1301 2003	2 2 7 2 5	8 8 567 8 125	1 1 1 1 1	7

The three following tables (table 2, 3 and 4) are show the randomness test results of the three digital sequences mentioned above by using MDSRT.

Table 2: MDSRT results of MCGS output with L=2000 for m=3.

Test	T^* Value	υ	Pass Value T_0
Freq.	2.428	2	6.01
Run	6.971 7.229 6.63	6 6 5	12.31 12.31 10.97
A.C.	No# of fail values 0.0≤T(τ)≤14.238 0.05% for 500 shift	1	3.81

Table 3: MDSRT results of MCGS output with L=5000 for m=5.

Test	T^* Value	υ	Pass Value T_0
Freq.	2.294	4	9.52
Run	1.4 3.49 5.73 6.62 10.99	3 4 4 3 4	7.84 9.52 9.52 7.84 9.52
A.C.	No# of fail value 0.0≤T(τ)≤9.465 0.07% for 500 shift	1	3.81

Table 4: MDSRT results of MCGS output with L=8000 for m=7.

Test	T^* Value	υ	Pass Value T_0
Freq.	6.992	6	12.309
Run	2.997 3.458 7.088 5.982 6.283 1.823 3.429	4 3 4 3 3 3 3	9.52 7.84 9.52 7.84 7.84 7.84 7.84
A.C.	No# of fail values 0.0≤T(τ)≤15.899 0.068% for 500 shift	1	3.81

8. Comparison Study of MDSRT Results between MCGS and Other Generators

In this section we try to make a comparison study between MCGS and other generators, for digital sequences with L=5000 and m=10 for the compared generators. Of course, the 1st generator is the MCGS number two which is mentioned in table 1 of the previous section. The 2nd is the binary LFSR with length 31 and the 3rd stage tapping as a connection function, in order to get digital sequences, we have to choose 4 bits from four different positions from the LFSR. The four bits transformed to hex, if we take mod 10, we get a digital sequences with m=10. Table 5 show the MDSRT results of digital sequences generated from MCGS and LFSR.

Table 5: MDSRT results of DS generated from MCGS and LFSR.

Test	MCGS			LFSR		
	T^*	υ	T_0	T^*	υ	T_0
Freq	4.05	9	16.9	402.6	9	16.9
Run	1.30	2	6.01	39.28	2	6.01
	0.93	2	7.84	39.06	2	6.01
	1.88	3	7.84	27.22	4	9.52
	1.99	2	6.01	7.49	2	6.01
	2.86	2	6.01	29.42	3	7.84
	0.99	2	6.01	18.39	2	6.01
	0.42	2	6.01	51.60	2	6.01
	2.78	2	6.01	34.47	1	3.81
	1.04	3	7.84	48.51	2	6.01
	3.30	2	6.01	58.17	2	6.01
A.C. 500 shift	fail values 0.0≤T≤5.4 0.06%	1	3.81	fail values 4.04≤T≤16.6 44.5%	1	3.81

In the same comparison study we can show that the Random Number Generator (RNG) which found by Mitchell [15], it's a digital generator (m=10 only) with good random sequence, but it has low complexity, with period less or equal q-1 for some primes. We expect that the choices of the primes will drop to 35% in order to gain period equal q-1, while the choices of MCGS still open to all primes. Table 6 shows the period of some primes for RNG system with frequencies of the sequence digits.

Lets now take q=997 to generate 996-digits sequence (using α_1=7, α_2=885 and k=1) to compare with same q for RNG in table 7 calculating the Standard Deviation, of the frequency of sequence digits, from the average.

Table 6: periods of RNG primes and frequencies of sequences digits.

Primes	Period	Frequencies				
		0	**1**	**2**	**3**	**4**
991	495	55	54	59	45	45
997	166	29	14	18	15	11
1003	464	52	36	52	60	36
1013	253	26	29	29	22	20
1019	1018	103	102	102	102	102
Primes	**Period**	**5**	**6**	**7**	**8**	**9**
991	495	52	52	39	46	48
997	166	11	14	18	13	23
1003	464	53	32	39	55	49
1013	253	26	29	26	25	31
1019	1018	102	101	101	102	101

Table 7: Standard deviation of the sequences digits from the average.

Gen.	Per. (L)	Frequencies					
		0	**1**	**2**	**3**	**4**	
MCGS	996	99	100	100	100	99	
RNG	166	29	14	18	15	11	
Gen.	**Per. (L)**	**5**	**6**	**7**	**8**	**9**	**SD**
MCGS	996	100	99	100	100	100	0.5
RNG	166	11	14	18	13	23	5.7

9. Conclusions and Future Works

This work concludes the following aspects:

1. It's obvious that we can use the MDSRT to estimate the randomness of binary sequence.
2. In digital auto correlation test, we can applied another method to estimate T(τ), the new method applied by adding (mod (m)) for the shifted sequence by τ with the origin sequence S, we get a new sequence S′ with length L-τ and the expected mean of occurrence of digit i is $E^*_F=\frac{L-\tau}{m}$, then applying the digital frequency test using the following equation:

$$T(\tau)=\sum_{i=0}^{m-1}\frac{(N_i(\tau)-\frac{L-\tau}{m})^2}{\frac{L-\tau}{m}}=\frac{m}{L-\tau}\sum_{i=0}^{m-1}N_i^2(\tau)-(L-\tau)$$

   Where $N_i(\tau)$ is the frequency of the digit I in the sequence S′.
3. We have to generalize more randomness tests, like serial, Poker,…etc. to be applied on digital tests, in order to estimate the real randomness of the sequence.

References

1. Schneier, B. **1996**. *Applied Cryptography: Protocols, Algorithms, and Source Code in C*, John Wiley & Sons, New York, 2nd edition.
2. Motwani, R. and Raghavan, P. **1995**. *Randomized Algorithms*, Cambridge University Press.
3. Menezes, A. P.; van Oorschot, P. and Vanstone, S. **1996**. *Handbook of Applied Cryptography*, CRC Press.
4. Gilbert, W. J. **2002**, *Modern Algebra with Applications*, Wiley-Interscince.
5. Golomb, S.W. **1967**. Reprinted by Aegean Park Press in 1982, *Shift Register Sequences* San Francisco: Holden Day.
6. Selmer, E. S. **1966**. *Linear Recurrence over Finite Field*, University of Bergen, Norway.
7. Schrift, A.W. and Shamir, A. **1993**. Universal Tests for Nonuniform Distributions, *Journal of Cryptology*, **6**:119–133.
8. Gustafson, H.; Dawson, E.; Nielsen, L. and Caelli, W. **1994**. *A Computer Package for Measuring the Strength of Encryption Algorithms*, Computers & Security, **13:**111-129.
9. Gustafson, H. **1996**. Statistical Analysis of Symmetric Ciphers, Ph.D. Thesis, Queensland University of Technology.
10. Gustafson, H.; Dawson, E. and Golić, J. **1996**, *Randomness Measures Related to Subset Occurrence*, Cryptography: Policy and Algorithms, International Conference, Brisbane, Queensland, Australia, (LNCS 1029), pp.132–143.
11. Yan, S. Y. **2000**. *Number Theory for Computing*, Springer - Verlag Berlin Heidel-berg, New York.
12. Bennett, D. J. **1999**. *Randomness*, Harvard University Press.
13. Martinez, W. L. and Martinez, A. R. **2002**. *Computational Statistics Handbook with MATLAB*, Chapman & Hall/CRC, Library of Congress Cataloging-in-Publication Data.
14. Al-Azawi, F. H. **2006**. Use the Multiplicative Cyclic Group to Generate Pseudo Random Digital Sequences, *Journal of Al-Rafidain University College for Sciences*, **20**:31-42.
15. Mitchell, D. W. **1993**. *A Nonlinear Random Number Generator with Known, Long Cycle Length*, Dept. of Economics, West Virginia University, Morgantown USA.