On Linear Complexity of Binary Sequences Generated Using Matrix Recurrence Relation Defined Over Z_4

Ramesh S [1], K N Haribhat [2], R Murali [3]

[1]Research Scholar, Dr MGR University, Chennai, India
Faculty, Department of Electronics & Communication Engineering
Dr. Ambedkar Institute of Technology, Bangalore, 560056, India
Email:rameshs_1@yahoo.co.in

[2]Dean Academic and Head, Department of Electronics & Communication Engineering
Nagarjuna College of Engineering & Technology Bangalore, India,
Email:knhari.bhat@gmail.com

[3]Professor, Department of Mathematics
Dr. Ambedkar Institute of Technology, Bangalore- 560056, India
Email:dr_muralir@yahoo.co.in

Abstract:

This paper discusses the linear complexity property of binary sequences generated using matrix recurrence relation defined over Z_4. Generally algorithm to generate random number is based on recursion with seed value/values. In this paper a linear recursion sequence of matrices or vectors over Z_4 is generated from which random binary sequence is obtained. It is shown that such sequences have large linear complexity.

Keywords

Matrix Recurrence Relation, Random Sequence, Linear Complexity, Z_4 Sequence, Random numbers

1. INTRODUCTION

Today's modern cryptography needs to deal with many aspects of security issues.In addition to providing confidentiality, it has to provide data integrity, authentication and non repudiation [1].Ever increasing computing power of potential attackers endangers even well-researched encryption algorithms [2].

Cryptography systems can be broadly classified into symmetric-key systems such as DES, AES and RC4 that use a single key that both the sender and recipient have to encrypt and decrypt respectively. Public-key or asymmetric systems such as RSA, ElGamal and Elliptic curve cryptograchy that use two keys, a public key known to everyone and a private key that only the recipient of messages uses [3], [4, [5]. Symmetric cryptosystem is usually divided into block ciphers and stream ciphers. Block ciphers operate with a fixed transformation on large blocks of plain-text data; stream ciphers operate with a time-varying transformation on individual plain-text digits [1], [5].This classification is not absolute, and any block cipher can be used as a

DOI : 10.5121/ijdps.2010.1207

stream cipher by using certain modes of operation. Cipher Feedback mode (CFB), Output Feedback mode and Counter mode operation on block cipher system can be used to turn a block cipher into a stream cipher. They can be proven secure under the assumption that the block cipher is secure [1], [3], [4].

In block cipher data is devided into blocks and encrypt and decrypt block wise, where in a stream cipher encryption or decryption is bit-by-bit or character by character. Both block ciphers and stream ciphers are in common use today. Generally in stream encryption system, a binary message is encrypted by adding bit by bit modulo 2 a binary random sequence called key sequence. Stream cipher system has the advantage that both encryption and decryption occurs at real time. Stream ciphers are especially prevalent in business and military applications [2], [3], [4] [5] [6], [7].

Security of stream cipher system depends on the randomness properties of the sequence called key sequence. Therefore key sequence generator is very important building block for stream cipher system. A random bit generator can be used to generate binary random bit sequences with desirable statistical properties which are important in cryptographic applications.The need for design of efficient and secure pseudorandom sequence generators remains an ongoing challenge and an important field in cryptographic research up to the present day.

A method of generation of random binary sequences using matrix recurrence relation defined over Z_4 is discussed in [8].There are standard tests like FIPS-1 and NIST-SP-800-2 revision 1 [9] test suites to test the randomness properties of binary sequences. It is shown in [8] that sequences generated using matrix recurrence relation pass these test suites. It is also found that such sequences exhibit good autocorrelation and cross correlation properties [10].

In this correspondence, we discuss the generation of random sequence defined over Z_4 using matrix recurrence relation and corresponding binary sequence is derived from it. Linear complexity of the binary sequences so generated is determined using Massey - Berlekamp algorithm [11] and results are analyzed. It is shown that such sequences exhibit large linear complexity which is desirable characteristics of random sequences required for key sequences in stream cipher systems.

Organization of the rest of the paper:

In Section 2 we introduce random sequence generator. Section 3 introduces a method of generation of random sequence using matrix recurrence relation defined over Z_4. Section 4 discusses the results. Section 5 contains concluding remarks.

2. RANDOM SEQUENCE GENERATORS

Random binary sequences are used as running key sequence in stream cipher system. Here message is in the form of binary sequence is encrypted by adding bit by bit modulo 2 a binary random sequence called key sequence and decrypted at the receiver using the same random key sequence generated at the receiver.

Linear Feedback Shift Registers (LFSRs) are important building blocks for generating key sequences. Maximum length sequences called m- sequences generated by an n stage Linear Feedback Shift Registers have very good randomness properties such as long period, (2^n-1), balance, ideal autocorrelation and good statistical properties which are desirable characteristic[6]. Also LFSR can be easily implemented both in hardware and software.However m-sequences have low linear complexity. For an m-sequence of length 2^n-1 the linear complexity is n. In this case only 2n consecutive bits are required to determine the feedback polynomial of the LFSR and hence the entire sequence. In the practical stream cipher designs, a large linear complexity of the key stream is obtained by a nonlinear transformation of the LFSR output sequence or nonlinear feedback shift register.

The use of non linear feedback mechanism to produce pseudorandom sequence is discussed in [3], [4]. Some of the methods of transformation define three general design categories: combination generators; filter generators and clock controlled generators [3].

The linearity complexity properties of LSFR's can be improved by feeding the outputs of several parallel LFSRs into a non-linear Boolean function to form a combination generator. The various properties of such a combining function are critical for ensuring the security of the resultant scheme, like avoiding correlation attacks.

Clock controlled generators [17] serve the function of introducing non-linearity in LFSRs. This non-linearity is achieved by having LFSR clocked irregularly by being driven by the output of some other LFSR. Several generators based on this principle have been proposed like stop-and-go, alternating step generator [12] and the shrinking generator [3], [4], [13], [14][15][16]

Another approach of improving the linearity complexity of sequence generated by LFSR is to use Filtering Boolean functions [5]. Although, not sufficient to be resistant enough against several attacks, certain characteristics are supposed to be necessary in stream ciphers with this structure. These characteristics include: high non-linearity, balance, and algebraic immunity [17].

Some of the widely used random sequence generators are generator based on recurrence modulo 2 [11],[18], linear feedback shift register generator (LFSR) [4], [20],[21],[22], feedback with carry shift registers (FCSR) [23],[24], nonlinear combination generators [3],[4], non linear feedback shift registers (NLFBSR) [3],[4], Marsaglia random number generators [25]-[26], and elliptic curve based pseudorandom random sequence generator [28] etc. Examples of cryptographically secure pseudorandom bit generators are RSA pseudorandom bit generator [29], Micali-Schnorr pseudorandom bit generator [30, 31], Blum-Blum-Shub pseudorandom bit generator (x^2 mod N generator) [32], [33]. Generally in all these schemes a random sequence is generated using seed key.Generally stream cipher systems are fast and easy to implement both in hardware and software.

The proposed scheme describes a method of obtaining binary sequences with large linear complexity.

3. PROPOSED RANDOM SEQUENCE GENRATOR USING MATRIX RECURRENCE RELATION

The proposed random sequence generator defined over Z_4 is based on Matrix Recurrence Relation defined by,

$$U_j = \sum_{i=0}^{n-1} A_i U_{j-i-1}, \quad j \geq n, \text{ arithmetic modulo } 4 \qquad (1)$$

Where A_is , i = 0,1,…………..,n-1are k×k matrices called co-efficient matrices , U_is , i = 0,1,……………,n-1are k×k initial matrices called seed matrices over Z_4 .With $U_0, U_1, U_2, \ldots U_{n-1}$ known, $U_n, U_{n+1}, \ldots$ satisfies the recurrence relation (1), where n is called order of the recurrence relation.

A Vector recurrence relation can also be defined as follows

$$V_j = \sum_{i=0}^{n-1} A_i V_{j-i-1}, \quad j \geq n \text{ , arithmetic modulo } 4 \qquad (2)$$

Where, as in equation (1) A_i s, i = 0,1,…,n-1 are k×k matrices called co-efficient matrices . V_i s , i = 0,1,…,n-1 are k×1 vector called seed values .With $V_0, V_1, V_2, \ldots V_{n-1}$ known, $V_n, V_{n+1}, \ldots$ satisfies the recurrence relation (2) .

Based on recurrence relation (1),with n arbitrary initial k×k seed matrices $U_0, U_1, U_2, \ldots U_{n-1}$, random sequence of k×k matrices $U_0, U_1, U_2, \ldots U_j$ over Z_4 is generated as follows

$$U_n = A_0 U_{n-1} + A_1 U_{n-2}, \ldots + A_{n-1} U_0 \quad \text{modulo } 4, \qquad (3)$$

$$U_{n+1} = A_0 U_n + A_1 U_{n-1}, \ldots + A_{n-1} U_1, \quad \text{modulo } 4 \qquad (4)$$

. . . .

. . . .

In general for $j \geq n$

$$U_j = A_0 U_{j-1} + A_1 U_{j-2}, \ldots + A_{n-1} U_{j-n}, \text{ modulo } 4 \qquad (5)$$

It can be shown that the sequence generated is strictly periodic if k× k coefficient matrix A_{n-1} is nonsingular [8].

3.1. General Case of Sequence over Z_4

The method of generation of sequences of matrices and sequences of vectors over Z_4 is described below.

Case I: ***Generation of Sequences of Matrices over Z_4***

In this case the n coefficients A_i s, i = 0, 1……….. n-1are k×k matrices over Z_4 which are arbitrarily chosen with A_{n-1} nonsingular to get strictly periodic sequence. The n initial matrices U_i s, i = 0, 1… ……..n-1are also k×k matrices over Z_4. Random sequence of k×k matrices is generated using equation (1) by randomly selecting the coefficient matrices and initial matrices over Z_4. Different sequences are generated for different coefficients and initial matrices. The properties of the sequences generated depend on coefficient matrices, initial seed matrices and n, order of recurrence relation given in expression (1).

To obtain Binary Sequence from Matrix Sequence over Z_4

Coefficient matrices U_is are over Z_4, the elements are from the set {0, 1, 2, 3}. Sequence over Z_4 can be derived from the sequence of k×k matrices by concatenation of the rows of U_i s .Then a sequence of N matrices each having k^2 elements from Z_4 gives rise to a sequence of Nk^2 elements over Z_4. By representing the elements 0, 1, 2, and 3 in binary as 00, 01, 10, and 11 respectively the length of the corresponding binary sequence is $2Nk^2$.

Consider k×k seed matrices, each having k^2 elements.

The number of possible seed matrices over Z_4 for each stage = 4^{k**2} (6)

For k=2, number of possible matrices = 4^4=256 (7)

For n stages the number of possible seed matrices over Z_4 is $4^{n\ k**2}$. If all the possible states are in one cycle, then the corresponding length of sequence is 4^{nk**2}, which is the maximum possible length. Corresponding maximum possible length of binary sequence is equal to $2(4^{nk**2})$. However the actual length of the sequence generated is always less than the maximum possible value and depends on the coefficient matrices and seed matrices.

Case 2: ***Generation of Sequence of Vectors over Z_4***

By having k×k coefficient matrices as above and with the initial values from a set of k×1vectors over Z_4 instead of k×k matrices, the sequences of k×1 vectors over Z_4 are generated. The number of possible seed vectors for one stage is 4^k and for n stages it is 4^{kn}.

To obtain Binary Sequence from Vector Sequence over Z_4

The recurrence relation given by equation (2) is used to generate sequence of k×1 vectors {V_i}, over Z_4.The seed values V_0,V_1,V_2, V_{n-1} are k×1 vectors over Z_4.Seed vectors along with known n coefficient matrices are used to obtain random vector sequence {V_i}. As mentioned earlier with $A_{n-1,}$ a k×k nonsingular coefficient matrix, the sequence generated is strictly periodic. Sequence over Z_4 is generated by concatenating the transpose of vectors in sequence {V_i}.

Consider vector sequence of length N.The corresponding sequence over Z_4 is then of length Nk. The corresponding binary sequence is of length 2Nk. For a vector sequence of size k×1, with n number of stages the maximum possible length of the sequence will be equal to 4^{nk} and the corresponding maximum possible length of binary sequence is equal to $2(4^{nk})$.

In this case also the actual length of sequences depends on number of stages n, choice of coefficient matrices and seed values and is less than the maximum possible value.

3.2. Linear Complexity Measure

One of the most useful concepts in the study of key sequences used in stream ciphers is that of linear complexity. The linear complexity of a binary sequence is defined as the length of the shortest linear feedback shift register that generates it. If a sequence has small linear complexity, then the synthesis of a linear equivalent of the sequence generator becomes computationally feasible. The linear complexity of a finite sequence is determined using Massey –Berlekamp algorithm [11]. The algorithm is briefly described below.

Algorithm:

INPUT: A binary sequence $s_n = s_0, s1, s2,\ldots, s_{n-1}$ of length n.

OUTPUT: Linear complexity $L(s^n)$ of s_n, $0 \le i \le n$.

Initialization: C (D) = 1= 0, *l*= 0, m= −1, B(D) =1, N= 0.

2. While (N < n) do the following:

2.1 Compute the next discrepancy d.

$$d = (s_N + \sum_{i=1}^{L} c_i s_{N-1}) \bmod 2,$$

2.2 If d = 1 then do the following:

T (D) = C (D), C (D) =C (D) + B (D) D^{N-m}.

If L ≤ N/2 then L =N + 1 − L, m= N, B (D) =T (D).

2.3 N= N + 1.

3. Return (L).

Where C (D) is connection polynomial, L is linear complexity.

It is desirable that random sequence which can be used as key sequence in stream cipher systems to have a large linear complexity. The necessary (but not sufficient) condition to be secure in running key stream cipher is to have large linear complexity [11], [34], [35]..

Even for a given order n, for different arbitrarily chosen n coefficient matrices and seed values, the generated sequence of length m, need not necessarily have same linear complexity. Further just by knowing n coefficient matrices and initial values, linear complexity cannot be predicted .Hence linear complexity can be treated as a random variable. Thus to study the statistical behavior of the linear complexity, number of sequences are generated and their linear complexity is computed. From these data we compute Mean, μ_x Variance, σ_x^2, and Standard deviation, σ_x of linear complexity. The definition of Mean, Variance, and Standard deviation are available in [37].The definitions are repeated here.

Mean

Let $X_1, X_2, X_3 \ldots X_N$ be N linear complexity values of N sequence of same length.The mean value, μ_x or expected value of linear complexity is defined as

$$\text{Mean}=\mu_x=\frac{1}{N}\sum_{i=1}^{N} X_i \qquad (8)$$

Variance and standard deviation, σ_x^2

Variance of random variable X is a measure of how far the value of random variable deviates from its mean value. It is defined as

$$\sigma_x^2=\frac{\sum_{i=1}^{N}(X_i-\mu_x)^2}{N} \qquad (9)$$

Very small value of variance indicates that X takes values almost equal to μ_x. The standard deviation σ_x is the positive square root of its variance σ_x^2. Standard deviation is a widely used measure of the variability or dispersion from mean value. It shows how much variation there is from the "average" (mean or expected) value. A low standard deviation indicates that the values tend to be very close to the mean, whereas high standard deviation indicates that the values are spread out over a large range of values, around the mean value.

Through simulation it is seen that by proper choice of size, k of matrix and number of stages, n it is possible to generate random binary sequences of large linear complexity.

4. SIMULATION RESULTS AND DISCUSSIONS

In general the binary sequence generated will have its linear complexity which depends on

i) n , the order of recurrence relation
ii) choice of n coefficient matrices and
iii) the n seed value

4.1. CASE 1: Matrix Sequences and Corresponding Binary Sequences

In the following studies k = 2 is chosen. From equation (6), the number of 2×2 matrices over Z_4 is 256.As n increases the linear complexity of the sequence also increases. This is verified by means of computer simulation. For n=5,10,15,20,25 and 30, the corresponding number of 2×2 coefficient matrices are arbitrarily chosen from a set of 256 , 2×2 matrices. Likewise the n seed values are also chosen arbitrarily. In each case, 10 different sequences of length m = 5000 bits are considered. Linear complexities of these 10 sequences are computed using Berlekamp-Massey algorithm [11].The statistical behavior of linear complexity X is found by computing its mean μ_x , variance, σ_x^2 and standard deviation ,σ_x after the linear complexities of 10 sequences are computed. The results are tabulated in Table 1.

Columns of Table 1shows the computed linear complexity of generated sequences using recurrence relation of order n =5, n=10, 15, 20, 25 and 30 respectively. It is observed from the Table 1 that the mean linear complexity μ_x for sequences of length 5000 bits for n = 5 are found to be 556. Similarly the mean linear complexities μ_x, for n=10, 15, 20, 25, and 30 are found to be 1875.8, 2495.8, 2493, 2498.3, 2499.3 respectively. The corresponding variance σ_x^2 is found to be 125.7777, 126.4, 130.4889, 52, 11.3444 and 0.9111 respectively. Likewise the corresponding standard deviation σ_x is found to be11.2150, 11.2427, 11.4231, 7.2111, 3.3681 and 0.6749 respectively. From these results it is evident that the linear complexity value is increasing with the order n of recurrence relation and approaches m/2 for n≥ 15.Correspondingly variance, and standard deviation decreases .This implies that the sequences generated have mean value of linear complexity μ_x almost equal to m/2 with high probability for m=5000 bits and n≥15.

Similarly Table 2 and Table 3, list the linear complexities X , mean μ_x, variance σ_x^2 and standard deviation σ_x for sequences of length 10000 bits and 20000 bits respectively , for different n = 5,10,15,20, 25 and 30. It is seen from Table 2 and Table 3 that for m= 10000 and 20000 and for n ≥20 the linear complexity approaches m/2 which is desirable for random sequences [36].

The variation of mean μ_x, variance σ_x^2 and standard deviation σ_x with different m and n are depicted in Figure.1, Figure.2 and Figure.3.respectively. For different length of sequence m =5000, 10000 and 20000 it is also seen that the values of variance and standard deviation are decreasing as n increases.

From the above results it is seen that for n ≥ 15 the mean value of linear complexity μ_x is almost equal to m/2 and hence the generated sequences have linear complexity m/2 with high probability for m=5000 bits. It is also seen that variance σ_x^2 and standard deviation σ_x are small. The same behavior of linear complexity, variance and standard deviation of linear complexity are observed for m = 10000 and 20000 bits. These are depicted in Figure.1, Figure.2 and Figure.3 respectively.

Table 1: List of Linear Complexity of Binary Sequences Generated Using Matrix Recurrence Relation Defined by equation (1), CASE 1 with k = 2
Length of Sequence = 5000 bits

Stage , n	5	10	15	20	25	30
LC of Sequence 1	544	1897	2500	2500	2499	2500
LC of Sequence 2	545	1871	2500	2489	2500	2500
LC of Sequence 3	567	1881	2498	2499	2500	2499
LC of Sequence 4	567	1855	2500	2489	2499	2499
LC of Sequence 5	565	1881	2500	2500	2500	2500
LC of Sequence 6	546	1877	2497	2488	2500	2498
LC of Sequence 7	539	1881	2500	2479	2498	2499
LC of Sequence 8	561	1879	2488	2489	2489	2499
LC of Sequence 9	567	1865	2488	2497	2498	2499
LC of Sequence 10	559	1871	2487	2500	2500	2500
Mean, μ_x	556	1875.8	2495.8	2493	2498.3	2499.3
Variance, σ_x^2	125.7777	126.4	130.4889	52	11.3444	0.9111
Standard Deviation, σ_x	11.2150	11.2427	11.4231	7.21110	3.3681	0.6749

Table 2: List of Linear Complexity of Binary Sequences Generated Using Matrix Recurrence Relation Defined by equation (1), CASE 1 with k = 2
Length of Sequence = 10000 bits

Stage , n	5	10	15	20	25	30
LC of Sequence 1	544	1897	3650	5000	5000	5000
LC of Sequence 2	545	1871	3662	4990	5000	5000
LC of Sequence 3	567	1881	3665	5000	5000	5000
LC of Sequence 4	567	1855	3664	5000	5000	5000
LC of Sequence 5	565	1881	3653	4997	4999	4999
LC of Sequence 6	546	1877	3650	5000	5000	5000
LC of Sequence 7	539	1881	3660	5000	4999	5000
LC of Sequence 8	561	1879	3655	4994	4998	4999
LC of Sequence 9	567	1865	3654	5000	5000	5000
LC of Sequence 10	559	1871	3643	5000	5000	5000
Mean, μ_x	556	1875.8	3655.6	4998.1	4999.6	4999.8
Variance, σ_x^2	125.7778	126.4	50.0444	12.1	0.4888	0.1777
Standard Deviation, σ_x	11.2150	11.2427	6.71118	3.4785	0.6992	0.4216

Table 3: List of Linear Complexity of Binary Sequences Generated Using Matrix Recurrence Relation Defined by equation (1), CASE 1 with k = 2
Length of Sequence = 20000 bits

Stage , n	5	10	15	20	25	30
LC of Sequence 1	544	1897	3650	9989	9998	10000
LC of Sequence 2	545	1871	3662	9989	9998	10000
LC of Sequence 3	567	1881	3665	9988	9998	9998
LC of Sequence 4	567	1855	3664	9989	9997	9998
LC of Sequence 5	565	1881	3653	10000	10000	10000
LC of Sequence 6	546	1877	3650	9989	9999	10000
LC of Sequence 7	539	1881	3660	9987	9987	9997
LC of Sequence 8	561	1879	3655	9988	9989	9998
LC of Sequence 9	567	1865	3654	9999	9999	10000

LC of Sequence 10	559	1871	3643	9989	10000	10000
Mean, μ_x	556	1875.8	3655.6	9990.7	9996.5	9999.1
Variance, σ_x^2	125.7778	126.4	50.0444	22.0111	21.1666	1.4333
Standard Deviation, σ_x	11.2150	11.2427	7.07420	4.6916	4.6007	1.1972

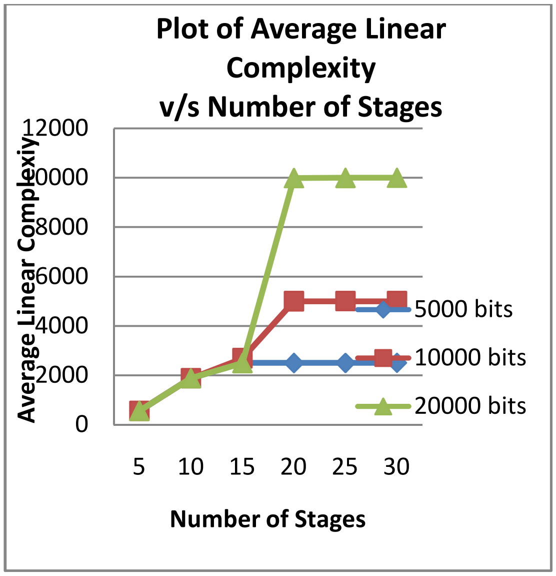

Figure 1: Plot of Mean Linear Complexity, μ_x v/s Number of Stages,n
Case 1 Matrix sequence with k=2

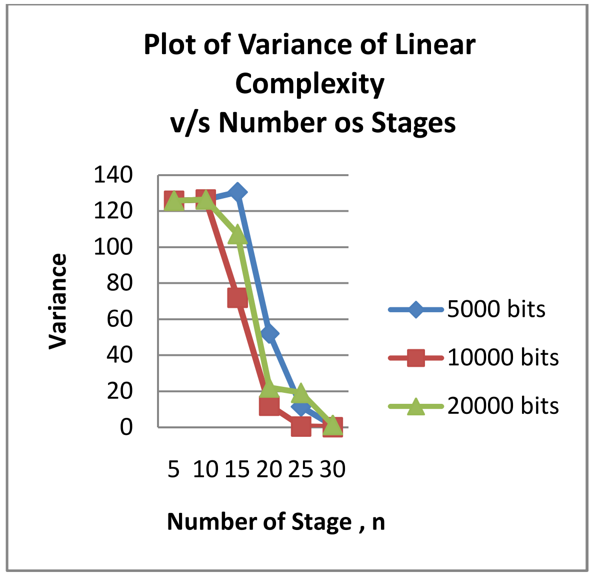

Figure 2: Plot of Variance of Linear Complexity, σ_x^2 v/s Number of Stages,n
Case 1 Matrix sequence with k=2

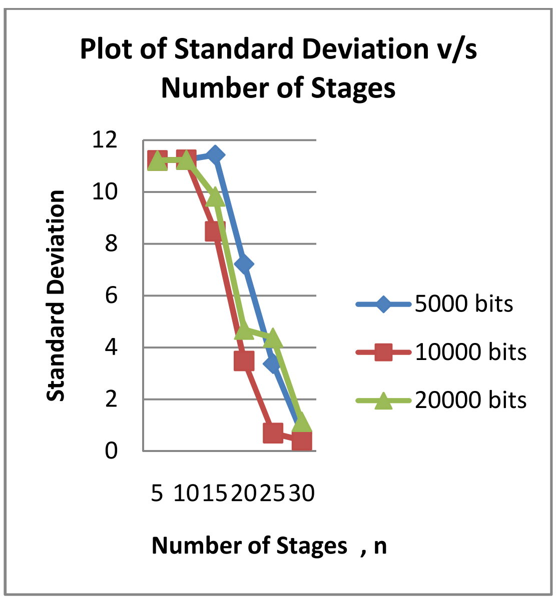

Figure 3: Plot of Standard Deviation of Linear Complexity, σ_x v/s Number of Stages,n
Case 1 Matrix sequence with k=2

4.2. CASE 2: Vector Sequences and Corresponding Binary Sequences.

For this case also k= 2 is chosen.
The simulation study is carried out with seed values of 16 possible 2×1 vectors over Z_4 instead of matrices. For n=5,10,15,20,25 and 30, the corresponding number of 2×2 coefficient matrices are arbitrarily chosen from a set of 256 possible 2×2 matrices over Z_4.

In each case, 10 different sequences of length m = 5000 bits are generated after binary transformation of sequences over Z_4 as discussed in section 3.1. Linear complexities of these 10 sequences are computed using Berlekamp- Massey algorithm [11].The statistical behavior of linear complexity is found by computing its mean value, variance and standard deviation.The results are tabulated in Table 4.Columns of Table 4 shows the computed linear complexity of 10 sequences generated using recurrence relation of order n =5, 10, 15, 20, 25 and 30 respectively. The corresponding mean linear complexity μ_x, variance σ_x^2 and corresponding standard deviation σ_x are also listed. From these results it is evident that the linear complexity value is increasing with the order n of recurrence relation and approaches m/2 for n≥ 15.Correspondingly variance and standard deviation decreases .This implies that the sequence generated have linear complexity almost equal to m/2 with high probability for m=5000 bits and n≥15.

Similarly in Table 5 linear complexities X, mean value of linear complexity μ_x, variance σ_x^2 and corresponding standard deviation σ_x for sequences of length 10000 bits for n = 5, 10,15,20,25 and 30 are listed.

Likewise in Table 6 linear complexities X, mean valueof linear complexity μ_x, variance σ_x^2 and corresponding standard deviation σ_x for sequences of length 20000 bits for n = 5, 10,15,20,25 and 30 are also listed. It is seen from Table 5 and Table 6 that for m= 10000 and 20000 and for n ≥20 the linear complexity approaches m/2.

The variation of mean μ_x, variance σ_x^2 and standard deviation σ_x with different m and n are depicted in Figure.4, Figure.5 and Figure.6.respectively for different length of sequence m =5000, 10000 and 20000. It is seen that the values of variance and standard deviation are decreasing as n increases.

From the above results it is seen that for n ≥ 15 the mean value of linear complexity μ_x is almost equal to m/2 and hence the generated sequences have linear complexity m/2 with high probability for m=5000 bits. It is also seen that variance σ_x^2 and standard deviation σ_x are small. The same behavior of linear complexity, variance and standard deviation of linear complexity are observed for m = 10000 for n ≥ 20 and for m = 20000.For n≥30, linear complexity nearly m/2.

Table 4: List of Linear Complexity of Binary Sequences Generated Using Vector Recurrence Relation Defined by equation (2), CASE 2 with k = 2
Length of Sequence = 5000 bits

Stage , n	5	10	15	20	25	30
LC of Sequence 1	251	789	2497	2500	2500	2500
LC of Sequence 2	254	765	2500	2500	2500	2500
LC of Sequence 3	245	778	2475	2489	2499	2499

LC of Sequence 4	254	754	2498	2489	2499	2499
LC of Sequence 5	257	775	2500	2478	2489	2498
LC of Sequence 6	251	779	2500	2500	2500	2500
LC of Sequence 7	255	765	2500	2500	2500	2500
LC of Sequence 8	243	770	2475	2489	2500	2498
LC of Sequence 9	254	764	2497	2489	2499	2498
LC of Sequence 10	255	775	2500	2500	2500	2500
Mean, μ_x	251.9	771.4	2494.2	2493.4	2498.6	2499.2
Variance, σ_x^2	20.7666	97.6	103.955	59.1555	11.6	0.8444
Standard Deviation, σ_x	4.5570	9.8792	9.6726	7.6912	3.4058	0.9189

Table 5: List of Linear Complexity of Binary Sequences Generated Using Vector Recurrence Relation
Defined by equation (2), CASE 2 with k = 2
Length of Sequence = 10000 bits

Stage , n	5	10	15	20	25	30
LC of Sequence 1	251	789	2680	4989	5000	5000
LC of Sequence 2	254	765	2670	4988	5000	5000
LC of Sequence 3	245	778	2680	4979	4999	4999
LC of Sequence 4	254	754	2689	4998	4999	5000
LC of Sequence 5	257	775	2686	4993	5000	5000
LC of Sequence 6	251	779	2690	4985	4997	4998
LC of Sequence 7	255	765	2697	4999	5000	5000
LC of Sequence 8	243	770	2680	4983	5000	5000
LC of Sequence 9	254	764	2697	4989	5000	5000
LC of Sequence 10	255	775	2690	4989	4999	4998
Mean, μ_x	251.9	771.4	2685.9	4989.2	4999.4	4999.5
Variance, σ_x^2	20.7666	97.6	71.8777	38.8444	0.93333	0.7222
Standard Deviation, σ_x	4.5570	9.8792	8.0430	6.2325	0.9660	0.8498

Table 6: List of Linear Complexity of Binary Sequences Generated Using Vector Recurrence Relation
Defined by equation (2), CASE 2 with k = 2
Length of Sequence = 20000 bits

Stage , n	5	10	15	20	25	30
LC of Sequence 1	251	789	3650	4989	6600	7560
LC of Sequence 2	254	765	3662	4988	6600	7568
LC of Sequence 3	245	778	3665	4979	6612	7569
LC of Sequence 4	254	754	3664	4999	6604	7565
LC of Sequence 5	257	775	3653	4993	6602	7565
LC of Sequence 6	251	779	3650	4985	6603	7564
LC of Sequence 7	255	765	3660	4999	6603	7560
LC of Sequence 8	243	770	3655	4983	6603	7564
LC of Sequence 9	254	764	3654	4989	6602	7562
LC of Sequence 10	255	775	3643	4989	6601	7560
Mean, μ_x	251.9	771.4	3655.6	4989.3	6603	7563.7
Variance, σ_x^2	20.7666	97.6	50.0444	40.9	11.7777	10.4555
Standard Deviation, σ_x	4.5570	9.8792	7.0742	6.3953	3.4318	3.2335

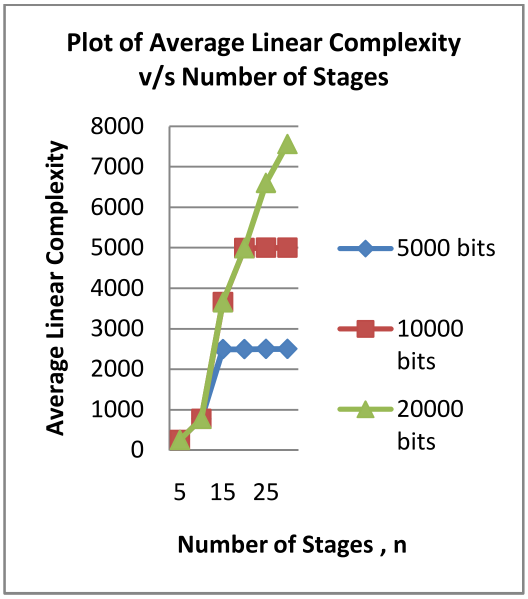

Figure 4:Plot of Mean Linear Complexity, μ_x v/s Number of Stages,n
Case I Matrix sequence with k=2

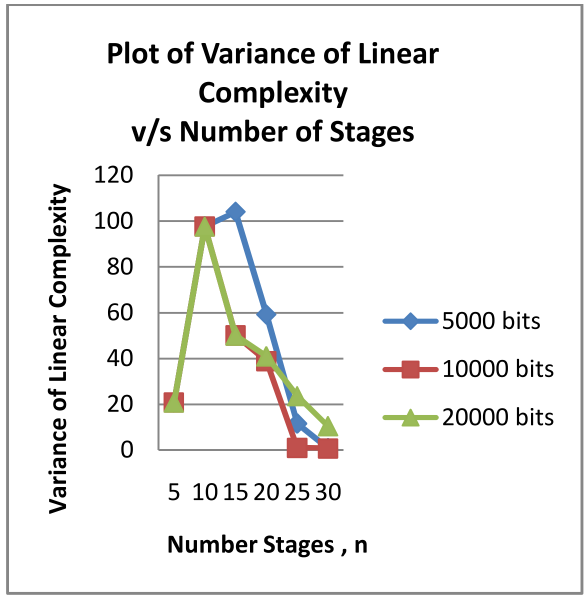

Figure 5: Plot of Variance of Linear Complexity, σ_x^2 v/s Number of Stages,n
Case I Matrix sequence with k=2

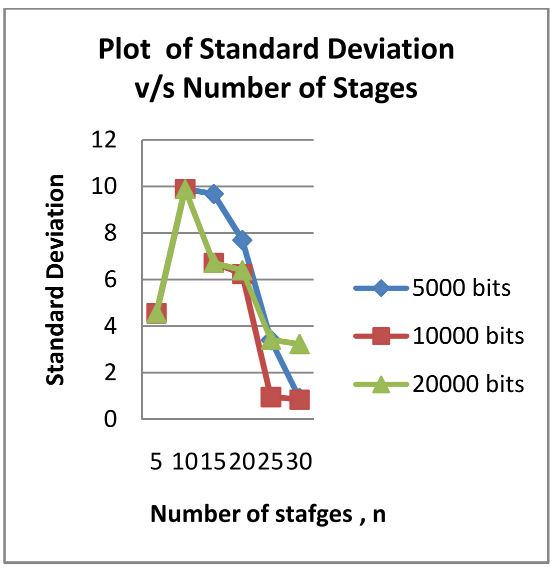

Figure 6: Plot of Standard Deviation of Linear Complexity, σ_x^2 v/s Number of Stages,n
Case I Matrix sequence with k=2

The statistical results μ_x, σ_x^2 based on 10 trials in each case are used to get the bound on probability of linear complexity X lying within the range $\mu_x-25 \leq X \leq \mu_x+25$ using Chebychev inequality [20]. First Chebychev inequality as defined in [38] is given below.

$$P (| X- \mu_x| \geq \delta) \leq \sigma_x^2/ \delta^2 ; \delta > 0 \quad (10)$$

Where $P (| X- \mu_x| \geq \delta)$ is the probability that X lies outside the range $(\mu_x - \delta) \leq X \leq (\mu_x + \delta)$. This probability is always less than or equal to σ_x^2/ δ^2.

Chebychev inequality is generally expressed as in equation (8). It can also be interpreted as,

$$P (| X- \mu_x| \leq \delta) \geq 1- \sigma_x^2/ \delta^2 ; \delta> \sigma_x \quad (11)$$

where $P (| X- \mu_x| \leq \delta)$ is the probability that X lies inside the range $(\mu_x - \delta) \leq X \leq (\mu_x + \delta)$.This probability is always greater than or equal to $1- \sigma_x^2/ \delta^2$

Equation.11 is used to observe the change in probability bound with increase in n from 15.The probability bound is computed for n = 15,20,25 and 30 using Table 1 to table 6. In all these cases δ is taken as 25.The results are tabulated in Table 7, 8, 9, 10, 11 and 12, corresponding to matrix and vector recurrence relations.

Table 7: Statistics computed for Matrix Sequences
Length of the sequence: 5000

Number of stage, n	15	20	25	30
Corresponding mean, μ_x	2495.8	2493	2498.3	2499.3
Corresponding variance , σ_x^2	130.4889	52	11.3444	0.9111
$P (\| X- \mu_x\| \leq 25) \geq 1- \sigma_x^2/ 625$	0.7912	0.9168	0.9818	0.9985

Table 8: Statistics computed for Matrix Sequences
Length of the sequence: 10000

Number of stage, n	15	20	25	30
Corresponding mean, μ_x	3655.6	4998.1	4999.6	4999.8
Corresponding variance , σ_x^2	50.04444	12.1	0.488889	0.177778
$P (\| X- \mu_x\| \leq 25) \geq 1- \sigma_x^2/ 625$	0.9199	0.9806	0.9992	0.9997

Table 9: Statistics computed for Matrix Sequences
Length of the sequence: 20000

Number of stage, n	15	20	25	30
Corresponding mean, μ_x	3655.6	9990.7	9996.5	9999.1
Corresponding variance , σ_x^2	50.0444	22.0111	21.1666	1.4333
$P (\| X- \mu_x\| \leq 25) \geq 1- \sigma_x^2/ 625$	0.9193	0.9648	0.9661	0.9977

Table 10: Statistics Computed for Vector Sequences
Length of the sequence: 5000

Number of stage, n	15	20	25	30
Corresponding mean, μ_x	2494.2	2493.4	2498.6	2499.2
Corresponding variance , σ_x^2	103.955	59.1555	11.6	0.8444
$P (\| X- \mu_x\| \leq 25) \geq 1- \sigma_x^2/ 625$	0.8336	0.9053	0.9814	0.9986

Table 11: Statistics Computed for Vector Sequences
Length of the sequence: 10000

Number of stage, n	15	20	25	30
Corresponding mean, μ_x	2685.9	4989.2	4999.4	4999.5
Corresponding variance , σ_x^2	71.8777	38.8444	0.93333	0.7222
$P (\| X- \mu_x\| \leq 25) \geq 1- \sigma_x^2/ 625$	0.8849	0.9378	0.9985	0.9988

Table 12: Statistics Computed for Vector Sequences
Length of the sequence: 20000

Number of stage, n	15	20	25	30
Corresponding mean, μ_x	3655.6	4989.3	6603	7563.7
Corresponding variance , σ_x^2	50.0444	40.9	11.7777	10.4555
$P (\| X- \mu_x\| \leq 25) \geq 1- \sigma_x^2/ 625$	0.9199	0.9358	0.9811	0.9832

5. CONCLUSIONS

Use of matrix recurrence relation (1) and (2) defined over Z_4 for the generation of random binary sequences derived from sequences over Z_4, results in random sequences with large linear complexity determined using Massey- Berlekamp algorithm. To study the statistical properties of the linear complexity, mean μ_x, variance σ_x^2 and standard deviation σ_x are determined by considering randomly chosen sequences of different lengths. Six cases for n=5, 10,15,20,25 and 30 are considered.In each case sequences of length 5000, 10000 and 20000 bits is considered. For each combination of n and m, by randomly chosing different initial matrices and coefficient matrices 10 sequnces are generated and their linear complexity property is investigated.

It is seen from the Tables 7 to12, that for $n \geq 15$ the probability that the linear complexity X differ by mean value μ_x by 25 always increases with n and m. For m = 20000 and n = 30 probability that linear complexity X is within (10000 ± 25) is always greater than or equal to 0.9977 in case of random binary sequence derived from recursion relation (1). Similarly it is greater than 0.9832 for random binary sequence derived from recursion relation (2).

The proposed method of generation of sequences over Z_4 is linear; the corresponding binary sequence with binary conversion 0,1,2,3 to 00,01,10,11 respectively turns out to be nonlinear. Hence the linear complexity of the sequence is increased compared to m- sequence. Algorithm is simple with modulo4 arithmetic. It is possible to implement both in hardware and software.There are large choices for number of stages,n initial content (U_i's or V_i's) of different size and the different coefficient matrices (A_i's) which can be chosen to generate large number of sequences.

REFERENCES

[1] William Stallings,Cryptography and Network Security, Priciples and Practice, 5th Edition, Pearson Education Inc,2006

[2] E. Zenner , " Cryptanalysis of LFSR-Based Pseudorandom Generators – A Survey", University of Mannhoim ,(Germany), 2004

[3] A. Menezes, P, Van Oorschot, and S. Vanstone, Handbook of Applied Cryptography. CRC Press, 1997

[4] Bruce Schneier, Applied Cryptography, Second Edition, John Wiley and Sons, 1996.

[5] Rainer A. Rueppel, "Stream ciphers", The Science of Information Integrity, IEEE press, New York, 1992

[6] S. Golomb, "Shift Register Sequences", Aegean Park Press (1982) reprint, Laguna Hills, California,1967

[7] M.J.B. Robshaw. "Stream Ciphers" , RSA Laboratories Technical Report TR-701,Version 2.0, July 25, 1995

[8] Ramesh S, K.N.Haribhat, Murali R, "Generation of Random Binary Sequences with Large Linear Complexity Using Matrix Recurrence Relation Defined over Z_4 and their Suitability for Cryptographic Applications ", International Journal of Advanced Engineering & Application, Vol.III.pp.343-350, June 2010

[9] Andrew Rukhin et.al, "A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications " , NIST Special Publication 800-22 Revision 1, May 2010

[10] Ramesh S, K.N.Haribhat, and Murali R, "Generation of Sequence of Random Numbers Defined over Z_4 and their Correlation Properties", Proceedings, International Conference on Communication Technology, IEEE –ICCT 2006 ,Gulin, China, Nov 27-30, 2006

[11] J.L.Massey, "Shift-Register Synthesis and BCH Decoding", IEEE Transaction on Information. Theory. Vol. IT-15, pp. 122-127, January 1969

[12] C. G Gunthar, " Alternating Step Generators Controlled by De Bruijn Sequence", Advances in Cryptology, EUROCRYPT'87, LNCS 304, pp.5-14,1988

[13] Simon R Blackburn, " The Linear Complexity of Self Shrinking Generator",IEEE Transaction on Information Theory,Vol.45,No.6, September 1999

[14] Borislav Stoyanov, "Recent Attacks Against Summation,Shrinking and Self Shrinking Stream Ciphers-Short Suvey, Fourth Scientific Confernce with International Participation, Space,Ecology, Nanotechnology,Safety, SENS 2008, Varna ,Bulgaria ,4-7 June 2008

[15] Zhaneta,Tasheva et.al., "Self Shrinking P-adic Cryptographic Generator, iCST 2005,Serbia and Montenegro, Nis, June 29-July1, 2005

[16] Willi Meier ,Othmar Staffelbach " The Self Shrinking Generator",Proceedings of Advances in Cryptology, EuroCrypt '94, Springer-Verlag, pp.205-214, 1998

[17] R.A. Rueppel, "New Approaches to Stream Ciphers. Ph.D Thesis", Swiss Federal Institute of Technology, Zurich, 1984

[18] Francois Panneton and Pierre L'Ecuyer, "Random Number Generators Based on Linear Recurrences in 2^{w}", available at http://www.iro.umontreal.ca/_lecuyer, 2004

[19] Pierre L'Ecueyer, François Panneton,Makoto Matsumoto, "Improved Long-period Generators Based on Linear Recurrences Modulo 2", ACM Transactions on Mathematical Software, Volume 32 , Issue 1 , pp. 1 - 16, March 2006

[20] Knuth, D.E, The Art of Computer Programming, Vol. 2, Semi Numerical Algorithms, Third Edition. Pearson Education,1998

[21] P. Ekdahl and T. Johansson, "A New Version of the Stream Cipher SNOW, Selected Areas in Cryptography ", LNCS 2595 pp. 47-61,2002

[22] Soloman Wolf Golomb, Guang Gong, Signal Design for Good Correlation for Wireless Communication, Cryptography and Radar, Cambridge University Press, 2005

[23] A. Klapper , M.Goresky "Feedback Shift Registers, 2-Adic Span and Combiners with Memory", Journal of Cryptology Vol. 10, pp. 111-147, 1997

[24] M. Goresky and A. Klapper, Algebraic Shift Register Sequences, 2009

[25] G Marsaglia, "Random Numbers Fall Mainly in the Plains", Proceedings, National Academic of Science 61(1),pp.25-28 ,1968

[26] Richard P. Brent, "Note on Marsaglia's XOR shift Random Number Generators", Journal of Statistical Software, Volume 11, Issue 5, August 2004

[27] George Marsaglia , "Xorshift RNGs", Journal of Statistical Software, Vol. 8, Issue 14, Jul 2003

[28] Sathyanarayana S .V, M.Ashwatha Kumar , K.N Haribhat, "A Study of Elliptic Curve Pseudorandom Sequence Generator", Proceedings ,5th National Workshop on Cryptology, J.N.N.College of Engineering,Shimoga, India,2005

[29] W. Alexi, B. Chor, O. Goldreich and C.P. Schnorr, "RSA and Rabin Functions: Certain Parts are as Hard as the Whole", SIAM Journal of Computing (2) 17, 194-209, 1988

[30] Blum, Micali, "How to Generate Cryptographically Strong Sequences of Pseudo-random bits", SIAM Journal on Computing, Volume 13 , Issue 4 , pp. 850 - 864 , November 1984

[31] S.Micali, C.P Schnorr, "Efficient, Perfect Random Number Generators" Preprint MIT, University of Frankfurt, 1988

[32] L. Blum, M. Blum, M. Shub, "A Simple Unpredictable Pseudo-random Number Generator", SIAM J. Computer, Vol. 15, pp. 364–3, 1986

[33] Christopher Drake, Anthony Nicholson, "A Survey of Pseudorandom Number Generators". University of Michigan, Department of Electrical Engineering and Computer Science,2002

[34] J.L.Massey,"Cryptography and System Theory," Rreprint-Proceedings, Allerton Conference on Communication Control, Computing, pp. 1-3, October. 1986

[35] Harald Niederreiter, " Sequences with Almost Perfect Linear Cmplexity Profile",Proceedings of the 6th Annual International Conference on Theory and Application of Cryptographic Techniques,Spinger Verlag, pp. 37-51,1987

[36] E Dawson,H Gustafson,N Davies, " Black Box Analysis of Stream Ciphers" Australasian Journal of Combinitorials,pp. 59-70,1991

[37] Vijay K Rohatgi, A. K. Md.Ehsanes Saleh, "An Introduction to Probability and Statistics", John Wiley & Sons(ASIA) Pte Ltd,2nd Edition

[38] John G. Proakis "Digital Communication" , McGraw Hill Higher Education, 4th Edition, December 1, 2000

Authors

Ramesh S received the B.E degree in Electronics & Communication Engineering from Gulbarga University, Karnataka, India in 1990, and M.Tech Degree in Industrial Electronics from Visvesvaraya Technological University, Belgaum, India in 2001 and currently working towards Ph.D Degree at Dr MGR University, Chennai, India. He is working as Faculty in the Department of Electronics & Communication Engineering, Dr Ambedkar Institute of Technology, Bangalore, India. His research areas include Analog Communication, Digital Communication, Cryptography and VLSI Design.

Dr K N Haribhat received the B.E Degree with honors from Mysore University in 1966, M.Tech and Ph.D in Electronics & Communication Engineering from Indian Institute of Technology, Kanpur, in 1973 and 1986, respectively. He is currently working as Dean Academic and Head, Department of Electronics & Communication Engineering at Nagarjuna College of Engineering & Technology, Bangalore, India. He was with Karnataka Regional Engineering College, Suratkal, India (Currently known as NIT-K) for more than 30 years. His research areas include Analog Communication, Digital Communication and Cryptography. He has authored more than 25 papers in National/international Conferences and Journals.He has coauthored two books on communication.

Dr Murali R received the M.Sc Degree in Mathematics from Bangalore University, Karnataka, India in 1990, and Ph.D Degree in Mathematics from Bangalore University, India in 1999 and currently working as Professor in the Department of Mathematics, at Dr Ambedkar Institute of Technology, Bangalore, India. His research areas include Graph Theory-Hamiltonian graphs. He has authored more than 12 papers in National/international Journals.