International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Study on Linear Complexity of Sequences Generated Using Modified A5/1 Algorithm

Anil Kumar K[1], Dr. Ramesh S [2]

[1, 2]Dr. Ambedkar Institute of Technology, Bengaluru

Abstract: ***GSM technology is widely used to provide voice communication for the mobile users. A5/1 is the stream cipher used for encryption in GSM communication system. Initially A5 algorithm was kept secret to ensure security but when algorithm was disclosed many cryptanalytic attacks were proposed and proved that A5 algorithm is cryptographically weak. In this paper the modification in A5/1 is proposed, major improvement in clocking unit and addition of non linear combining function for the output to improve the Linear Complexity of the output bit sequence generated. The Linear Complexity (LC) of binary sequence so generated are computed and result are discussed using Berleykamp-Massey algorithm.***

Keywords: *GSM*, Stream Cipher, A5/1, LFSR, Linear Complexity.

1. Introduction

Cryptography is a mechanism by which security and authentication is provided to the authorized user. A5/x are the encryption algorithms incorporated in GSM communication system to deliver voice encryption and decryption used in mobile phones [1]. This technique makes GSM the most secured mobile communication standard currently available. The Encrypted voice and data communications between the mobile station and the network is accomplished through use of the ciphering algorithm.A5/1 is a symmetric stream cipher that generates pseudo-random binary sequences which are used to encrypt the message signals.

Generally encryption of message is carried out by XORing the message sequence with secrete key sequence and decryption is made by XORing the received encrypted message with the same secret key sequence. The strength and security of these ciphers depends upon the characteristics of bit sequences produced by the stream generation algorithm.

Research studies and analysis has shown that A5/1 has some weaknesses which lead to cryptographic attacks. One of the drawbacks of A5/1 algorithm is the weak clocking mechanism that depends upon majority rule [2]. Majority rule uses three clocking bits c_1, c_2 and $c_3$to determine the value of majority m using m = maj (c_1, c_2, c_3). It defines the majority among these bits, if two or more are 1 then the value of majority m is 1. In this work our objective is to replace weak clocking mechanism with improved clock rule called the M-rule. The study of LC profile for the obtained sequence is made by determining the LC of the sequence using Berleykamp-Massey algorithm.

2. Related Work

Many versions of A5/1 are used in more countries. A5/2 is a weaker version used in countries where export restrictions apply. A5/3 encryption algorithm is used for GSM, Enhanced Circuit Switched Data (ECSD), GPRS Encryption Algorithm 3(GEA3) and General Packet Radio Service (GPRS) [1].A5/1 has some drawbacks in clocking mechanisms and fixed feedback polynomial of linear feedback shift registers. Most of the attacks against A5/1 are known as plain text attacks.

To secure communication from the risk of theft, some modification in feedback shift registers are made that improves structure of A5 algorithm. By using unit delay the strength of the key stream generator is increased along with randomness [3].Two modifications are made in A5/1 and A5/2 ciphers by using tapping mechanism and increasing the number of LFSR from three to five [4].

The Berleykamp-Massey algorithm identifies the shortest LFSR that can be used to generate finite binary sequence. For finite random sequence, the Berleykamp-Massey algorithm is used to calculate the shortest LFSR which is LC [5]. The Maximum LC obtained for a sequence of ‘k’ bit is approximately ‘k/2’. The high LC indicates that longer shift register is needed to generate sequence [6].The random sequence used for key stream cipher system, it is important to have large Linear Complexity [7].

3. GSM A5/1 Stream Cipher

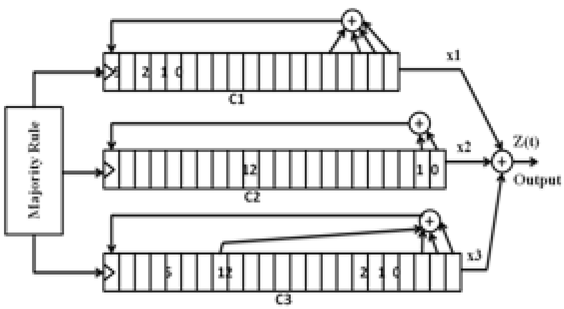


Figure 3.1: A5/1 Stream Cipher

In the general algorithm, the clock clocking unit has three bits c_1, c_2, c_3 which provides majority output given by the Equation 1.

$$m = maj(c_1, c_2, c_3) \quad (1)$$

The linear feedback shift registers (LFSRs) R_1, R_2, $R_3$are of lengths 19, 22, 23 bits respectively. Each LFSR is clocked depending on the output m for example, let (b_1, b_2, b_3)= (1, 1,

0) then according to majority rule m=1 [11]. So $b_1=b_2=m$ and for $b_1=1$, register R_1 is clocked and similarly for $b_2=1$, R_2 is clocked. At each clock the individual LFSR generates one bit $x_i$and output z(t) is given as $z(t) = x1 \oplus x2 \oplus x3$ to produce one bit at the output keystream z(t). The architecture of the A5/1 stream cipher is shown in Figure 3.1.

4. Modified Proposed Stream Cipher

The major enrichments are made in the Clock-Controlling unit and LFSR initializations. The proposed modified A5/1stream cipher system is shown in the Figure 4.1. The clock controlling unit with advanced clocking mechanism consists of six input bits b_1, b_2, b_3, c_1, c_2, c_3 and finds two majority functions

$$m_1 = maj(b_1, b_2, b_3) \quad (2)$$
$$m_2 = maj(c_1, c_2, c_3) \quad (3)$$

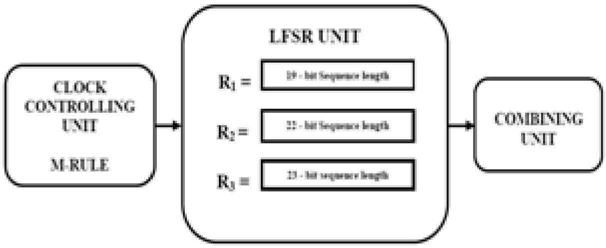


Figure 4.1: Modified Proposed A5/1 Stream Cipher

The six input bit positions are fixed and they are chosen from LSFR respectively. For bit b_1 the value from register position $R_1[5]$ is chosen, where $R_1[5]$ represent fifth position of shift register R_1. For bit $c_1$the value from register position $R_1[11]$ is taken, similarly for bit b_2 it is $R_2[16]$, for bit $c_2$it is $R_2[9]$, for bit $b_3$it is $R_3[6]$, and for bit $c_3$it is $R_3[14]$ the values are chosen which is shown in the Figure 4.2. The logic used to clock the registers is M-rule.

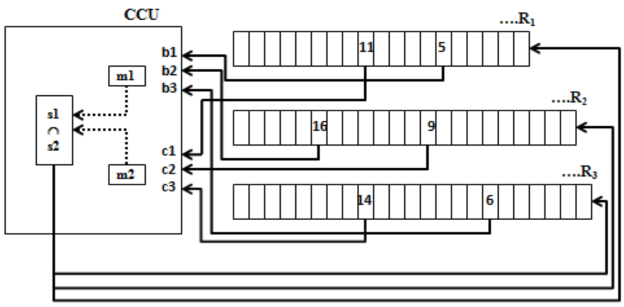


Figure 4.2: Clock Controlling Unit

M-rule considers the two majority functions m_1 and m_2 defined in (2) and (3) respectively. Since binary data is loaded in the shift register, the input bits can take 1 or 0. For an Example if $(b_1, b_2, b_3) = (1, 1, 0)$, then according to Equation (2), $m_1 = 1$. Since bits b_1 and b_2 are 1, the registers R_1, R_2which correspond to bits b_1, b_2, are stored in a set s1{}, that is s1{R_1, R_2}. If $(c_1, c_2, c_3) = (0, 1, 1)$, then from Equation (3) $m_2 = 1$and registers R_2, R_3corresponding to c_1, $c_2$are stored in set s2{}, that is s2{R_2, R_3}. By intersecting the two sets s1{R_1, R_2} and s2{R_2, R_3}, we get R_2 as common and hence R_2 is shifted by one position. Since any one of the register shifts, the values stored also stored, the chances of input bits in clock also changes which is random in nature.

From Table 1, consider the third row where the input bits are stored with values (0, 1, 0, 1, 0, 1) where majority functions m_1 and m_2 are defined. Using M-rule s1{R_1, R_3}, s2{R_1, R_3} are assigned and by comparing the sets, the common registers R_1, R_3 are selected for clocking. Similarly few possible combinations for selecting registers to be clocked are shown.

Table 1: Majority Table According to M-Rule

(b_1, b_2, b_3)	(c_1, c_2, c_3)	s1{}	s2{}	Registers clocked
0 0 0	1 1 1	R_1, R_2,R_3	R_1, R_2,R_3	R_1, R_2,R_3
0 0 1	1 0 0	R_1, R_2	R_2, R_3	R_2
0 1 0	1 0 1	R_1, R_3	R_1, R_3	R_1, R_3
0 1 1	1 1 0	R_2, R_3	R_2, R_1	R_2
1 0 0	0 0 1	R_2, R_3	R_1, R_2	R_2
1 0 1	0 1 0	R_1, R_3	R_1, R_3	R_1, R_3
1 1 0	0 1 1	R_1, R_2	R_2, R_3	R_2
1 1 1	0 0 0	R_1,R_2, R_3	R_1,R_2, R_3	R_1,R_2, R_3

By using M-rule the probability of individual LFSR being clocked is improved.

The Linear Feedback Shift Registers used are R_1 of 19 bit, R_2 of 22 bit and R_3 of 23 bit length. Totally 64 bits are initialized by 0 or 1 before clocking.

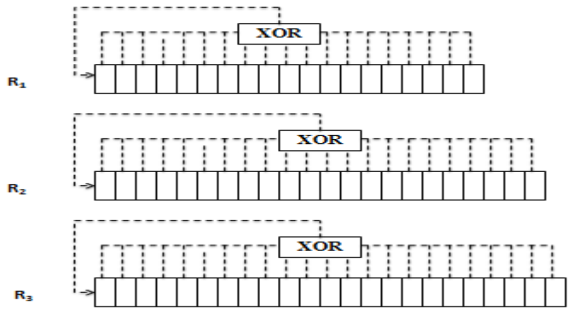


Figure 4.3: LFSR with Feedback Initialization

The initialization is done using primitive polynomials [8] shown in Table 2. For each register, one polynomial is defined for Example the register R_1 having the polynomial $x^{19}+x^{15}+x^{14}+x^8+x^7+x^3+x^2+x^1$, the register is loaded as [1000110000011000111] and the feedback is connected with XOR gate whose inputs are the bit values stored with 1 and the output for MSB of the same register.

Two different initializations are obtained by dynamically choosing feedback polynomials shown in Table 2. For the first case polynomials p_1, p_2, p_3 are chosen for R_1, R_2, R_3 and for the second case polynomialsp_4, p_5, p_6 are chosen for R_1, R_2, R_3.

Table 2: Feedback tapping for LFSR

LFSR length	Feedback
R_1 = 19 bit	$p_{1} = x19+x^{15}+x^{14}+x^{8}+x^{7}+x^{3}+x^{2}+x^{1}$ $p_{4} = x^{19}+x^{18}+x^{15}+x^{12}+x^{1}$
R_2 = 22 bit	$p_{2} = x^{22}+x^{20}+x^{12}+x^{11}+x^{9}+x^{7}+x^{6}+x^{4}+x^{3}+x^{2}$ $p_{5} = x^{22}+x^{21}+x^{10}+x^{9}+x^{1}$
R_3 = 23 bit	$p_{3} = x^{19}+x^{17}+x^{9}+x^{7}+x^{3}+x^{2}$ $p_{6} = x^{23}+x^{18}+x^{1}$

Combining Function

The Combining functions [2] f_1 and f_2 are given by the Equations 4 and 5.

$$f_1=R_1[1]\oplus R_3[2]\oplus(R_2[1]*R_3[1])\oplus(R_1[1]*R_3[1]) \quad (4)$$
$$f_2=R_2[1]\oplus R_3[1]\oplus(R_1[1]*R_3[5])\oplus(R_1[1]\oplus R_3[1]) \quad (5)$$

where,
$\oplus$ --- represents XOR function
* --- represents AND function
$R_1[1]$ --- represents binary value in register R_1 at position 1. Similarly for other registers it is defined

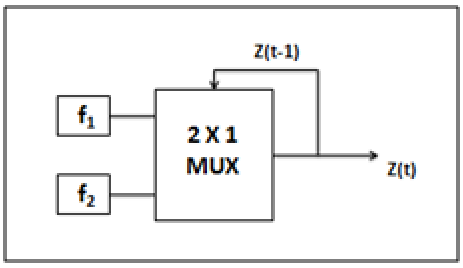


Figure 4.4: Combining Function for Output z(t)

The bit values in each register change randomly, accordingly f_1, f_2 also changes. By obtaining f_1 and f_2, a data selector is taken for generating output bit z(t) where, f_1 and f_2 are taken as inputs with select line z(t-1) and output is z(t) as shown in the Figure 4.4.The select line z(t-1) is taken from the output which stores previous bit. Depending on this value the output is generated.

5. Implementation Methodology

From the modified cipher system, the clock unit, LFSR initialization and combining function are set up. The cipher is made to run for one time where clock with M-rule perform set of operations and controls LFSR clocking. The combining functions f_1, f_2 are obtained from which one bit z(t)is obtained finally. In this way if the process is continued for 50000 runs, totally 5000 bits are generated at the output z(t) like (10001010100001101100…….1010100001). Since LFSR is having two different feedback polynomial combinations, for each initialization the z(t) of sequence length 50000 bits are obtained.

To compute Linear Complexity, The BM algorithm is implemented for two different random binary sequences obtained. They are shown in case 1 and case 2 in the following section. For any cryptographic applications it is necessary to have sequence of larger LC for the algorithm to be robust and strong towards any attack.

6. Results And Discussion

For the generated set of sequences, LC values are shown in Table 3.Generation of binary sequence is discussed for two cases, LC is obtained for binary sequence and also it is determined for different sub-sequences.

Case 1:
The LFSR is initialized for the registers R_1, R_2 and R_3 of length 19bits, 22bits and 23bits respectively.

R_1 --- 0 1 0 0 1 1 0 0 1 0 1 1 0 0 0 0 1 1 1
R_2 --- 1 0 1 0 0 0 0 0 0 0 1 1 0 1 0 1 1 0 1 1 1 0
R_3 --- 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 1 1 0

Let S_1......................$S_{49999}$is the binary sequence generated with 50000 bits. Table 3 shows the computed values of LC for sub-sequences of different lengths which are randomly selected from sequence S_1 to $S_{49999.}$

The Table 3 consists of 9 rows and 9 columns with a total of 81 LC values for different sub-sequences. Each column has fixed length starting from 8 bit, 32 bit, upto 20480 bit.

The row consists of sub-sequences with initial bit positions (IBP) S_1, S_8, S_{16}, S_{32}, till S_{512}. From the generated sequence, the sub-sequences are selected where S_{32} starts from 32^{nd} bit of generated sequence, S_{16} start from 16^{th} bit of generated sequence and similarly other sub-sequences are chosen. The table has LC for each sub-sequence with known sequence range and lengths respectively.

Table 3: LC for sequences starting from length 8 to 20K bit

IBP	Length of binary sub-sequence in bits								
	8	32	64	128	512	1K	5K	10K	20K
S_1	4	17	32	64	256	512	2560	5121	10238
S_8	4	17	32	64	256	512	2560	5121	10238
S_{16}	3	16	31	63	255	511	2558	5121	10241
S_{24}	3	16	31	63	254	511	2559	5120	10240
S_{32}	3	16	31	63	255	511	2558	5120	10240
S_{64}	4	17	32	64	256	512	2558	5121	10235
S_{100}	4	16	32	64	256	512	2560	5121	10240
S_{256}	4	16	32	64	256	512	2560	5121	10240
S_{512}	4	16	32	64	256	512	2560	5121	10240

By observing Table 3, the values in the 5^{th} column give the LC value of binary sequences of length 512 bits. For Example the value 256 in 2^{nd} row 5^{th} column is the value of LC of sub sequence taken from binary sequence S_1......................S_{49999}, selecting from S_8……S_{520}.The LC values in 4^{th}columncorrespond to sequences of length 128 bit. For Example the value in 4^{th} row 4^{th} column, the LC value 63 is for sequence S_{24}……S_{152}.The LC values in 9^{th} column correspond to sequences of length 20480 bit.

For Example the value in 6^{th} row 9^{th} column, the LC value 10240 is for sequence S_{32}……S_{20512}. From the LC values computed for different lengths considering 8 bit, 16 bit,………,20480 bits and randomly choosing the sub sequences, it is seen that the LC values is found to be approximately equal to N/2 where N is the sequence length.

Case 2:
LFSR is initialized for the second time with R_1=19bits, R_2=22 bits and R_3=23 bits.

R_1 --- 1 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1
R_2 --- 1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1
R_3 --- 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1

The LC values for sub-sequences of different lengths which are randomly selected from S_1 to S_{49999} areshown in the Table 4.It consists of 9 rows and 9 columns with a total of 81 LC values for different sub-sequences. Each column has fixed length starting from 8 bit, 32 bit, upto 20480 bit.

The row consists of sub-sequences with initial bit positions (IBP) S_1, S_8, S_{16}, S_{75}, till S_{1000}. From the generated sequence, the sub-sequences are selected where S_{75} starts from 75^{th}bit of generated sequence, S_{16} start from 16^{th} bit of generated sequence and similarly other sub-sequences are chosen. The table has LC for each sub-sequence with known sequence range and lengths respectively.

Table 4: LC for sequences starting from length 8 to 20K bit

IBP	Length of binary sub-sequence in bits								
	8	16	64	128	256	1K	5K	10K	20K
S_1	4	8	32	64	128	512	2560	5120	10241
S_8	4	8	32	64	128	512	2560	5120	10241
S_{16}	4	8	32	64	128	512	2560	5120	10241
S_{75}	3	9	33	64	128	512	2560	5120	10241
S_{100}	4	8	32	64	128	512	2560	5120	10241
S_{500}	3	7	31	63	127	511	2559	5119	10240
S_{600}	4	8	32	64	128	512	2560	5121	10240
S_{750}	4	8	32	64	128	512	2560	5121	10240
S_{1000}	4	8	32	64	128	512	2560	5121	10240

By observing Table 4, the values in the 5^{th} column give the LC value of binary sequences of length 256 bits. For Example the value 128 in 2^{nd} row 5^{th} column is the value of LC of sub sequence taken from binary sequence S_1.......................S_{49999}, selecting from S_8……S_{264}. The LC values in 2^{nd}column correspond to sequences of length 16 bit. For Example the value in 4^{th} row 2^{nd} column, the LC value 9 is for sequence S_{75}……S_{91}. The LC values in 9^{th} column correspond to sequences of length 20480 bit.

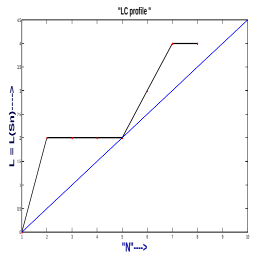
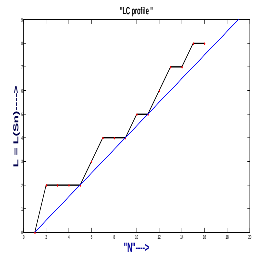
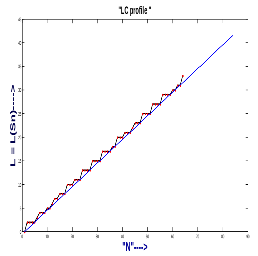
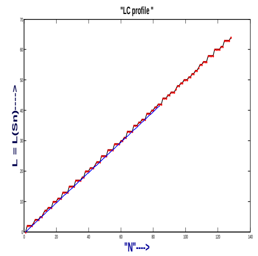
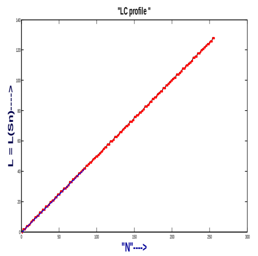
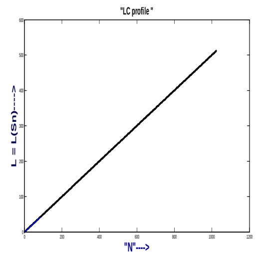
For Example the value in 7^{th} row 9^{th} column, the LC value 10240 is for sequence S_{600}……S_{20980}. From the LC values computed for different lengths considering 8 bit, 16 bit ,………, 20480 bits and randomly choosing the sub sequences, it is seen that the LC values is found to be approximately equal to N/2 where N is the sequence length.

6.1 Linear Complexity Profile (LC Profile)

The LC profile shows a graph of LC values plotted along x-axis for given sequence length (N). The LC value increases with increase in sequence length. The N/2 line is exactly at 45 degree inclined with respect to x-axis for each graph shown in Table 5.

Considering the first row of a sequence s_1 from Table 4, the binary sequence lengths starts from 8 bit to 1024 bit. The LC values are plotted separately in each graph for each sub sequence.

Table 5 : LC Profile for Binary Sequences

LC Profile	
 (a): For 8 bit sequence length	 (b): For 16 bit sequence length
 (c): For 64 bit sequence length	 (d): For 128 bit sequence length
 (e): For 256 bit sequence length	 (f): For 1024 bit sequence length

From Table 5 (a), the LC values obtained are plotted for sequence s1 of length 8 bit. Similarly in (b), the LC plot for sequence of length 16 bit, in (c) LC plot for length 64 bit,(d) LC plot for length 128 bit,(e) LC plot for length 256 bit,(f) LC plot for length 1024 bit is shown. It is seen that LC line closely follows N/2 line for randomly chosen sequence of different lengths considered in this investigating. This nature of LC profile is desirable for sequences to be random.

7. Conclusion

The modified A5 algorithm is proposed for generation of random binary sequences of different lengths. Binary sequences are obtained and their LC and LC profile are studied. It is found that LC values approximately follows N/2 where N is length of binary sequence which is good indicator for randomness of sequence generated.

References

[1] AlpeshR.Sankaliya, V. Mishra and AbilashMandoli, *"Implementation of cryptographic algorithms for GSM cellular standard"*, GANPUT University Journal of engineering and technology, Vol-1, issue-1, Jan-June-2011.

[2] Musheer Ahmed and Izharuddin, *"Enhanced A5/1 Cipher with Improved Linear Complexity"*, impact-2009, 978-1-4244-3604-0/09/, 2009 IEEE.

[3] Sattar B. Sadkhan, NibrasHadiJavad, *"Improvement of A5/1 encryption algorithm based on using Unit Delay",* University of Babylon college of sciences.
[4] NurHafizaZakaria, KamaruzzamanSeman and Ismail Abdullah, *"Modified A5/1 Based Stream Cipher for Secured GSM Communication"*, IJCSNS International journal of computer science and network security, Vol-11, no-2, feb-2011.
[5] A. Menezes, P, Van Oorschot, and S. Vanstone, *"Handbook of Applied Cryptography",* CRC Press, 1997.
[6] M. J. B. Robshaw, "*Stream Ciphers*", RSA laboratory technical report. TR-701, version 2.0- july-25, 1995.
[7] Ramesh S, K N Haribhat, R Murali, *"On Linear Complexity of Binary Sequences Generated Using Matrix Recurrence Relation Defined Over Z4"*, International Journal of Distributed and Parallel Systems (IJDPS) Vol.1, No.2, November 2010.
[8] Nikesh Bajaj, *"Enhancement of A5/1 using variable feedback polynomials of LFSR"*, 978-1-4577-0240- 2/11/, 2011 IEEE.

Author Profile

Anil Kumar K received the B.E degree in Electronics & Communication Engineering from Banglore University, Karnataka, India in 2012 and currently working towards M.Tech Degree in VLSI Design and Embedded Systems from Visvesvaraya Technological University, Belgaum, India (2013-2015). His interested areas include VLSI Design, CMOS circuit design in Cadence and Cryptography.

Dr Ramesh S received the B.E degree in Electronics & Communication Engineering from Gulbarga University, Karnataka, India in 1990, M.Tech Degree in Industrial Electronics from Visvesvaraya Technological University, Belgaum, India in 2001 and Ph.D Degree in 2013 from Dr MGR University, Chennai, India. He is working as Faculty in the Department of Electronics & Communication Engineering, Dr Ambedkar Institute of Technology, Bangalore, India since 23 years; His research areas include Analog Communication, Digital Communication and Cryptography. He has authored more than 25 papers in National/international Conferences and Journals.