Key Generation for Text Encryption in Cellular Networks using Multi-point Crossover Function

Y.V. Srinivasa Murthy
Assistant Professor
Department of CSE
ANITS, VSKP

Dr. S. C. Satapathy
Professor & HOD
Department of CSE
ANITS, VSKP

P. Srinivasu
Associate Professor
Dept. of CSE
ANITS, VSKP

A.A.S. Saranya
4/4 B.Tech
Dept. of CSE
ANITS, VSKP

ABSTRACT
This paper is mainly concerned with providing security for messages in cellular networks. Encryption of data in cellular networks is mandatory since it is sensitive to eaves dropping. This project focuses on encrypting the data sent between the mobile stations and base stations using a stream cipher method. However, the keys for encryption are generated using an evolutionary computation approach termed genetic algorithm. This genetic algorithm technique gives the best or optimal key for encryption. Before we single point cross over technique is used in generating optimal key for encryption but this paper emphasizes on genetic algorithm technique for different sizes of population and different number of iterations considering multi point crossover. The plain text which is to be encrypted along with the key are encoded using the arithmetic coding technique. Encryption is done to convert the plain text into cipher text. And the comparison with the existing system is explained in detail.

General Terms
Algorithms, Security, Ciphers, Networks, Encoding.

Keywords:
Encryption; Eavesdropping; Vernam technique; Random Key Generation algorithm; Genetic algorithm; Stream cipher; Multi-point Crossover; arithmetic algorithm;

1. INTRODUCTION
Now the world being a mobile ruled world SMS has become a popular means of communication. Even the confidential matters are also sometimes dealt through messages. Exchange of sensitive data has become more common through this medium. But in cellular networks these are vulnerable to the eavesdroppers and hackers. So there is a necessity to encrypt the message in order to safeguard the information. To prevent the data from eaves dropping stream cipher method are applied. This paper deals with the encryption of data by means of multipoint cross over technique followed by a data compression technique called arithmetic coding. [3]Then obtained information is passed over for converting into a cipher text by means of vernam cipher and finally it is thus transferred to the receiver to keep the sent information in a secured manner.

1.1 Vernam Cipher
The Vernam cipher is also called as a one-time pad. The Vernam cipher has been the best ever technology found so far which mainly emphasizes on the computational security i.e. The ease of cracking the key generated by this method becomes a tough task for current technologies. In fact it is proved mathematically as the best in providing the security. In Vernam cipher the message is represented as a binary string (a sequence of 0's and 1's using a coding mechanism such as ASCII coding. [8]The key is of the binary format containing only 0s and 1s. The encryption is done by adding the key to the message exclusive or modulo 2, bit by bit. This process is often called as exclusive or and is denoted by XOR and is represented by the symbol⊕.

1.2 Huffman Coding
Huffman coding is static in nature. It acts like a interpreter which works on each and every symbol one after the other but not all at once. Huffman coding generates codes of different lengths. Grater the length is lowers the occurrence of the symbol in the given text. Loss of data in this compression technique will be less. Entire message will be stipulated into individual symbols and codes are assigned based on their position in the Huffman tree. [6]

1.3 Cryptography
Cryptography is similar to that of encrypting technique. Its main Moto is to convert a meaning data into merely nonsense data which is understood alone by the receiver when decrypted. Basically two types of keys are employed in cryptography. [13] They are symmetric and asymmetric keys. When same key is used is both at sender and receiver for encrypting and decrypting respectively it comes under symmetric key. When two keys are employed namely public key for encryption and private key for decryption such a technique comes under Asymmetric key technique.

1.4 Genetic Algorithms
In brief genetic algorithm (GA) is a technique that implements the concept of biological evolution as a programming methodology. For a given problem the set of input values will be given which will be evaluated on few techniques mainly aiming at meeting the needs of Fitness Function. The values that are obtained as the solution may be the existing solution or even a new set of values generated in random Genetic Algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions of optimization problems using techniques inspired by natural evolution such as inheritance, mutation, selection and cross over. [1]

The mobile user sends the message and we need to encrypt that message. We generate a number of keys randomly by using a Random Key Generation algorithm. We calculate the fitness function for the generated keys and sort them in the ascending

order of the fitness values. We select a specified number of keys which has maximum fitness values. Using Evolutionary technique we apply crossover to the selected keys and obtain the new generation of keys. Repeat the process until the number of iterations specified. As a result of last iteration we get the best key among generated keys. For the crossover operations we take the multipoint cross over which gives the best optimal key.

Next step is to encode the received plain text and the key generated in the previous step by using arithmetic coding technique. In this technique, we calculate the frequency of each character and then calculate the probability of each character. We assign the probability values and specify the interval for each character which gives the maximum compressed data.

In the final step we encrypt the message using a stream cipher called vernam cipher. In Vernam technique we apply the XOR operation between the plain text and key. Finally we get the cipher text for the given plain text.

2. CELLULAR NETWORKS

Cellular communication is seeing an explosive growth due to increased usage. But it is prone to a threat called eaves dropping which interrupts the privacy of the use. It is therefore essential that the data traffic across the cellular communication network is encrypted. A Cellular network consists of mobile stations attached to a base station (BS). A cluster of BS's which is fixed, attaches to a mobile telephone switching office which is connected to the public switched telephone network (PSTN). [11] Cryptographic schemes are developed for protecting alphanumeric data since the emerging wired and wireless IP networks are vulnerable to eavesdropping. Thus in the case of the cellular network, the messages sent between the Sender and the receiver is encrypted using a stream cipher method. The keys are generated randomly i.e. Initial population and the new population is obtained by using Evolutionary technique known as Genetic algorithm.

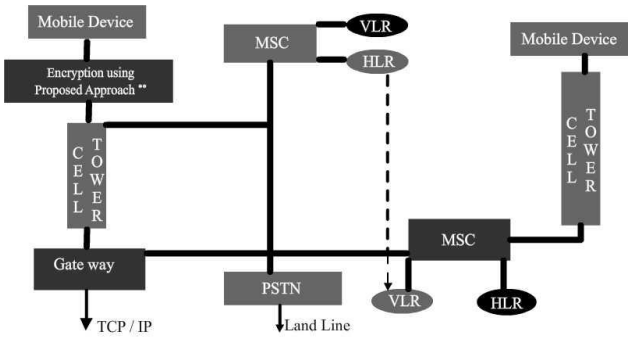

Fig 1: Model of Cellular Network with proposed approach

A mobile switching Center is the electronic field office of a cellular carrier, which automatically coordinates and switches calls between mobile phones in a given service area. Each cell in a cellular network is controlled by an MSC, which constantly monitors each caller's signal strength and arranges cellular handoffs. When a signal begins to fade, the MSC locates another MSC, better positioned to manage the call and re – routes it to maintain the communication link. [15]

MSCs are connected to base stations by T1 landlines or microwave channels, and by landlines to the Public Service Telephone Network (PSTN).

MSCs maintain individual subscriber records, current status of subscribers, and information on call routing and billing in two subscriber databases: the Home Location Register (HLR) and the Visitor Location Register (VLR).

MSCs maintain individual subscriber records, current status of subscribers, and information on call routing and billing in two subscriber databases: the Home Location Register (HLR) and the Visitor Location Register (VLR).

The Home Location Register (HLR) is a database for permanent storage of subscriber data and service profiles. The Visitor Location Register (VLR) is a database that contains temporary information about subscribers. The MSC uses this information to keep track of and serve visiting subscribers – in other words, roamers.

When a cellular phone roams outside its home MSC, the local carrier communicates with the home provider to obtain device and subscriber data, which it loads into its VLR. This information is maintained by the local provider as long as the mobile device roams within the local paging area.

The above diagram shows the various components that make up the cellular network. Notice that gateways provide mobile devices access to the Internet across TCP/IP lines.

When a text gets transferred from the mobile device it is encrypted using the proposed approach which follows a sequence of steps namely random key generation, selection of the best key, encoding followed by the encryption thus finally simulating the cipher text. There by the encrypted message gets transferred through the cell towers which in turn follow the guidelines of GSM architecture finally delivering that to the receiver's cell tower where decryption will be done and is transferred in safe to the mobile device at the end point.

2.1 GSM

GSM will be under the category of Cellular networks where the mobile searching will be placed in the station or a cell that is immediately to the current system. It falls under second generation mobile phone technology. GSM has overcome the many disadvantages of the previous versions in the aspects of roaming, portability and even to the network operators. The well known feature short message service (SMS), otherwise known to be text messaging is implemented at a reasonable cost. [12]

Various features of GSM can be mentioned as follows:

2.1.1 Full length Access to the services

This provides one of the user friendly features where a user can have the total access to the network irrespective of his location.

2.1.2 Effective Capacity utilization and frequency Allocation

GSM overcomes the disadvantage of analog based networks where the capacity becomes in an issue in the cities. Proper utilization of giving frequency bandwidth and even small cells

makes GSM to overcome that problem. Various techniques are also implemented in modulation as well.

2.1.3 Lock to the personal information

Security is the major challenge ahead for any system in the current world. GSM provides proper methods for securing data that is sent. User's secrecy can be maintained by means of encryption and even frequency hopping and is implemented by using digital systems and signaling alone.

2.1.4 Always at the door

Various services that are provided by GSM are: transfer of data, voice, messages additional features such as call forwarding is also allowed in it.

2.2 Architecture of GSM

GSM consists of Mobile Station that is a phone which interacts with the base station system (BSS). Base transceiver system (BTS) is the radio interface to the mobile. Base Station controller (BSC) supports the functions and physical links between MSC and BTS. [10]

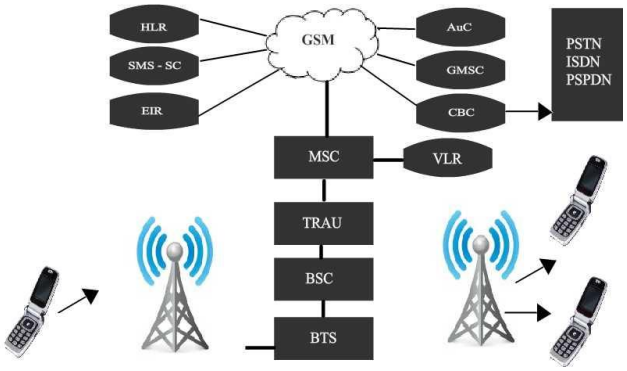

Fig 2: Architecture of GSM

HLR is the permanent database of the subscriber's information for a cellular network. VLR is a database that contains information about the subscribers roaming within a mobile switching center's (MSC) location area. The primary role of the VLR is to minimize the number of queries that MSCs have to make to the home location register (HLR), which holds permanent data regarding the cellular network's subscribers.The GSM equipment identity register (EIR) database contains information on the identity of mobile equipment to prevent calls from stolen, unauthorized or defective mobile stations.GSM network operators maintain three lists of international mobile equipment identities (IMEI) in their equipment identity register (EIR):

grey - GSM mobile phones to be tracked

black - barred GSM mobile phones

white - valid GSM mobile phones

A short message service center (SMSC) is the portion of a wireless network that handles SMS operations, such as routing, forwarding and storing incoming text messages on their way to desired endpoints. Wireless network operators connect SMSCs through SMS gateways.

The authentication center (AUC) provides authentication and encryption parameters that verify the user's identity and ensure the confidentiality of each call. The authentication center (AUC) also protects network operators from fraud.Transcoder and Rate Adaptation Unit, or TRAU, performs transcoding function for speech channels and RA (Rate Adaptation) for data channels in the GSM network.

3. ENCRYPTION USING GENETIC ALGORITHMS

3.1 Existing System using Ant Colony Key Generation

The existing system provides security by encrypting the message which involves the generation of key by using an Ant Colony Key Generation algorithm. In Ant Colony Key Generation Algorithm, key is generated randomly and it will calculate the energy value for that key. It is accepted if it is greater than the threshold value which is specified by the user otherwise generate another key. Here we are not comparing the strength of keys with one another. We are accepting the key if it is just greater than the threshold value.

3.2 Proposed System

Our proposed system overcomes the limitations in the existing system. We generate the best key using an evolutionary technique called genetic algorithm. In this technique, we generate a set of keys randomness and calculate fitness function for each key. Select the best keys which have maximum fitness value and apply crossover to generate a new set of keys. Now calculate the fitness value for new keys generated and select best keys among them. Again apply crossover and generate new keys. Repeat this process for specified number of iterations. Then select the best key which has maximum fitness value. Here the generation of best key is done by selecting a key among a huge number of keys and we are comparing the fitness values of all keys. Then encryption process is done by using Vernam Cipher

4. GENETIC ALGORITHM AS A ROOT FOR KEY GENERATION

Genetic algorithms work with candidate keys in a large amount. The search space of this algorithm may not truly direct it towards the solution but even make it a failure. For the current paper we consider a data that is to be transmitted. Based on the population size and iteration size mentioned by the user we generate the key set using various techniques. Entire encryption of data can be done by means of using a single point crossover. But this paper emphasizes on the multi point cross over technique to get better optimal key for encryption. [5]

4.1 Multipoint cross over

Crossover results in the formation of new off springs. It is simply an action of cute-swap-splice of the "genetic material" of

the parents from the randomly chosen crossover point and forming new off springs. In Single-point crossover only one position is opted for cross over. *k*_ [1,2, ...,*Nvar*-1], *Nvar*: number of variables of an individual, is selected uniformly at random and the variables exchanged between the individuals about this point, then two new offspring is produced. Whereas in case of multipoint cross over many of the variations are done on the crossover operators. That is those crossover operators many will not be on the adjacent strings. [7]

4.2 Roulette Wheel Selection

Roulette wheel selection is applied even in genetic algorithms. Selection of parent chromosomes is based on their fitness function. Better the value of fitness function better the chances of selecting the parent chromosome followed by the better opportunity in the roulette wheel. Selection of chromosome is directly proportional to its fitness function. [2]

Algorithm to implement the roulette wheel:

Step1: find the fitness of entire population which is obtained by summation of individual fitness functions. Let the sum be V. this step is performed only once for the entire population.

Step2: generate a number in random within the interval (0-V) let it be S.

Step3: search in the population and total fitness from 0 to V. when V is greater than S stop the process and return the current chromosome.

4.3Procedure

The process of key generation is explained by the following steps:

Step 1: Initially generate a set of keys depending on the population size randomly of the specified key size.
Step 2: Retrieve the message i.e. Plain text to be encrypted from the GUI.
Step 3: Calculate the fitness function for each chromosome i.e. the key using the strategy of the following function:

$$KE_r = \frac{(K_{rc} \epsilon T_r)}{Size\ (K)}$$

Where

KE_r – Key Energy of r^{th} key
K_{rc} – character at r^{th} row and c^{th} column
T_r is the plain text at r^{th} row.
K is the Key

Step 4: Now sort the chromosomes according to the fitness function.
Step 5: Read the crossover probability (cp) from the GUI and perform crossover

$$N = \left(\frac{cp}{100}\right) * ip$$

Where

N – No. of Chromosomes
cp – Crossover Probability
ip – Initial Population Size

Step 6: Generate new set of chromosomes by multipoint crossover, now we get N new chromosomes
Step 7: Calculate the fitness function for these new chromosomes and add them to the keys in the previous iteration.
Step 8: Sort the chromosomes and select the best set of population for the next iteration.
Step 9: Repeat the steps from 3 to 8 till certain specified number of iterations.
Step 10: Choose the top most key from the last iteration as the key to be used for encryption.

4.4 Vernam Cipher:

The Verna cipher is based on the technique of merging the plain text and the random text that is the key obtained from the GA. The resultant cipher obtained is free from all the risks that are ahead and can be sent in a safe from sender station. At the receiver's end the separation of key and plain text takes place thus retrieving the information that is sent. [14]

Steps to implement vernal cipher:

Step1: By means of Using the encoding techniques like Arithmetic algorithm in this case generate an equivalent numerical value for the characters..
Step2: Encrypting key is generated by using GA technique.
Step3: XOR operation is performed between the numerical values of the characters of the plain text and the corresponding key value .
Step4: XOR operation on the receiver's side with the key generates the plain text again.
Step5: In order check the safety of the message XOR it with the key which generates the pad content again if it is unbiased.

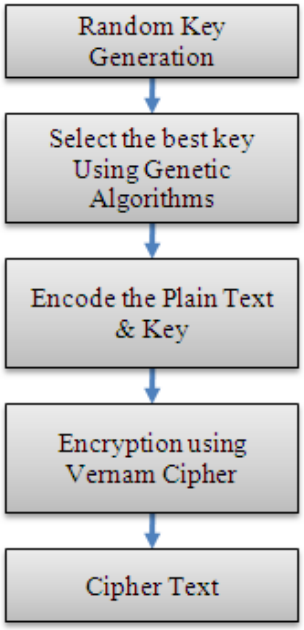

Fig 3: Gist of Paper

5. Huffman Coding Algorithm

Huffman coding is an encoding technique which reduces the repetition of characters in a string. This method is not dynamic but generates codes of different string lengths. [3]Smaller the

length of code grater the frequency of character in the text based on the codes assigned they are placed in the ascending order or the descending order as per the user's requirement and Huffman tree is constructed according to that.

5.1.1 Building Huffman Tree

The format of Huffman tree is represented by means of a binary tree. Each node of the binary tree consists of symbol that are to be encoded and the probability of the occurrence of each character can be represented by means of a binary string which is obtained by top down parsing. The binary equivalent of the string obtained gives the code assigned for the specified symbol. The tree may be constructed as follows:

Here we take the plain text and the key generated from the genetic algorithm as the input:

Step 1: *Calculate the frequencies and probabilities of each character in the key and the plain text.*
Step 2: *Create a parentless node for each symbol. Node is the combination of symbol and the corresponding probability.*
Step 3: *Select the two parentless nodes with the lowest probabilities.*
Step 4: *Create a new node which is the parent of the two lowest probability nodes.*
Step 5: *Assign the new node a probability equal to the sum of its children's probabilities.*
Step 6: *Repeat from Step 2 until there is only one parentless node left.*

The code for each symbol is the value obtained by the top down parsing. The leaf nodes of the tree are the symbols or the characters of the string. For the left sub tree if the value is assigned as 1 then right sub tree will be given 0. if the leaf node is obtained by parsing the tree thrice left and twice right then value will be '11100'. Representation of Huffman tree can be depicted in the following fashion. [4]

*****
**/ **
(0) (1)
**/ **
(10)(11)
**/ **
(110)(111)

After building the Huffman tree codes of symbols can be retrieved by means of following steps:

Step 1: *Remember the lengths of the codes resulting from a Huffman tree generated as per above.*
Step 2: Arrange the symbols according to the length of the codes according to the increasing or decreasing orders..
Step 3: let the current node be assigned a value zero that is longest length code is assigned zero value. Now assign the values from shortest to longest as per the following order:
If the length of the symbol is greater than the next symbol then perform the right shift operation.
Assign the code obtained to the present symbol.
Increase the code to the next code.
Now again sort out the symbols and select the longest code.
Continue the process till all the symbols are assigned codes.

6. EXPERIMENTAL RESULTS

Implementation of giving paper can be done on a given input message as "Genetic Algorithm based Key Generation for Text Encryption in Cellular Networks". We need to accept the plain text, opt for the best key from a set of keys generated and then generate the binary code for each character using Huffman coding. Now cipher text is generated based on the x-or operation performed between key and character binary codes. Thus secured cipher text is transferred from sender to receiver.

6.1 Results after Huffman Coding

Table 1: Huffman Coding results

Character	Code	Decimal	Character	Code	Decimal
G	1011	19	a	011	3
E	011	3	s	10010	18
N	1101	13	d	1110000	112
T	1011	11	K	1110011	115
I	1010	10	y	00000	0
C	100	4	F	1110010	114
	010	2	t	1111101	125
A	1010	10	x	1111100	124
L	11101	29	E	1111111	127
G	01	1	p	011	3
O	1000	8	c	10	2
R	1100	12	u	1111110	126
H	10111	23	N	111101	61
M	10110	22	w	1111001	121
B	1110001	113	K	1111000	120

6.2 Implementation of Vernam Cipher

Table 2: Results after Vernam Cipher

Plain	Code	Key	Code	Cipher	Plain	Code	Key	Code	Cipher
G	19	i	10	25	e	3	a	9	10
E	3	s	18	17	x	124	g	7	123
N	13	C	2	15	t	11	a	9	2
E	3	G	19	16		2	e	9	11
T	11	N	61	54	E	127	i	18	109
I	10	n	13	7	n	13	s	26	23

C	4	P	3	7	c	4	C	10	14
	2	R	12	14	r	12	G	27	23
A	10	a	3	9	y	0	N	69	69
I	29	g	1	28	p	3	n	21	22
G	1	a	3	2	t	11	p	11	0
O	8	e	3	11	i	10	r	20	30
R	12	i	12	0	o	8	a	11	3
I	10	s	20	30	n	13	g	9	4
T	11	C	4	15		2	a	11	9
H	23	G	21	2	I	10	e	11	1
M	22	N	63	41	N	13	i	20	25
	2	n	15	13		2	s	28	30
B	113	p	5	116	C	2	C	12	14
A	3	r	14	13	e	3	G	29	30
S	18	a	5	23	l	29	N	71	90
E	3	g	3	0	l	29	n	23	10
D	112	a	5	117	u	126	p	13	115
	2	e	5	7	l	29	r	22	11
K	115	i	14	125	a	3	a	13	14
E	3	s	22	21	r	12	g	11	7
Y	0	C	6	6		2	a	13	15
	2	G	23	21	N	61	e	13	48
G	19	N	65	82	e	3	i	22	21
E	3	n	17	18	t	11	s	30	21
N	13	P	7	10	w	121	C	14	119
E	3	R	16	19	o	8	G	31	23
R	12	A	7	11	r	12	N	73	69
A	3	G	5	6	k	120	n	25	97
T	11	A	7	12	s	18	p	15	29
I	10	E	7	13	--	--	--	--	--
O	8	I	16	24	--	--	--	--	--
N	13	S	24	21	--	--	--	--	--
	2	C	8	10	--	--	--	--	--
F	114	G	25	107	--	--	--	--	--
O	8	N	67	75	--	--	--	--	--
R	12	N	19	31	--	--	--	--	--
	2	P	9	11	--	--	--	--	--
T	125	R	18	111	--	--	--	--	--

6.3 Efficiency of Algorithm

On Sample text "Genetic Algorithm based Key Generation for Text Encryption in Cellular Networks" we calculated the efficiency by calling this algorithm recurrently for 10 times with some fixed population size. The key which is generated for each eyelet along with energy value represented in the following graph with an average energy value of 0.783 which gave more accuracy than existing system.

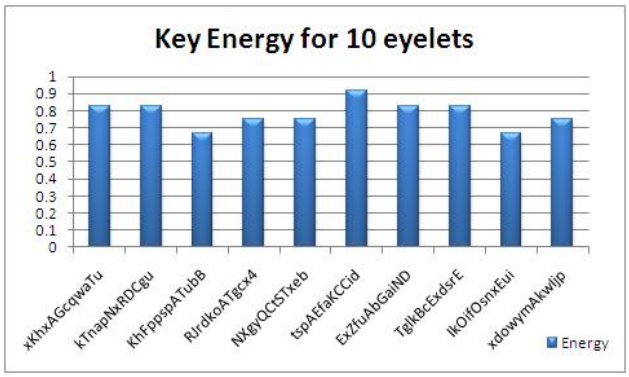

Fig 4: Statistics of proposed Approach

7. ACKNOWLEDGMENTS

We owe our tributes to Dr S.C. Satapathy who inculcated the idea of developing this paper based on Genetic Algorithms. This entire work goes in credit of Y.V.S.Murthy, who put forth his ideas in implementing the ideas and making this paper a successful one.

8. CONCLUSION OF FUTURE WORKS:

Genetic algorithms stands out the best even in the complex scenarios. It is an emerging technology in which several improvements have to be made. One of such improvements in the implementation of Grey code the segment merging during mutation which helps in the yield of better results. The future work of this work can be expected by means of implementing seal cipher instead of stream cipher and the implementation of arithmetic encoding in the place of Huffman encoding. [9]

9. REFERENCES:

[1] [DAVI91] L. Davis, "Handbook of Genetic Algorithms", Von Nostrand Reinhold, New York, 1991.

[2] [GOLD88] D. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley, Red-wood City, CA, 1988.

[3] Charles P. Fleeger, Shari Lawrence P. Fleeger, "Security in computing", Third Edition, Prentice Hall of India, 2003.

[4] Chung-Ping Wu, C.C. Jay Kuo, "Design of Integrated Multimedia Compression and Encryption Systems", IEEE Tran-sactions on Multimedia, Volume - 7, Issue - 5, October, 2005 Page (s): 828 - 839.

[5] David E. Goldberg, Kelsey Miman and Christina Tidd. "Genetic Algorithm: A Bibliography". University of Illinois at Urbana – Champaign 1992.

[6] Donald E Knuth, "Dynamic Huffman Coding, Journal of Algorithms", Volume 6, Issue 2, June 1985, Pages 163 – 180, ISSN: 0196 – 6774, 10.1016/0196-6774 (85) 90036-7.

[7] Gary William Flake, "The Computational Beauty of Nature". MIT Press, Cambridge, MA.

[8] Gustavus J. Simmons. 1979. "Symmetric and Asymmetric Encryption". ACM Comput. Surv. 11,4 (December, 1979), 305-330. DOI=10.1145/356789.356793

[9] Ian H. Witten, Radford M Neal and John G. Cleary, 1987. "Arithmetic Coding for Data Compression". Comm. ACM 30, 6 (June–1987), 520 540. DOI=10.1145/214762.214771.

[10] J. Eberspacher, H. J. Vogel, C. Bettestetter, and C. Hartmann, "GSM, Architectuer, Protocols and Services". 3rd Edition. John Wiley, Ltd., 2001.

[11] Jingyuan Zhang, Ivan Stojmenovic, "Cellular Networks", University of Alabama, University of Ottawa, Canada.

[12] Michel Mouly, Marie-Bernadette Pautet. 1992. "The GSM System for Mobile Communications". Telecom Publishing.

[13] Rivest, Ronald L. (1990), "Cryptography: In Handbook of Wireless Networks and Mobile Computing", (Stojmenovic I Ed), John Wiley & Sons, 27 – 49, (2002)

[14] Scottt Fluhrer, Itsik Mantin and Adi Shamir, "Weaknesses in the key scheduling algorithm of RC4", Lecture Notes in Computer Science, Vol. 2259, Revised Papers from the 8th Annual International Workshop on Selected Areas in Cryptography Pages: 1 - 24, Year of Publication: 2001.

[15] Tom Clements. "Making Sense of Cellular". Article, published in Sun Developer Network(SDN) of Oracle Corporation, July 2002