A Novel method to secure data using DNA sequence and Armstrong Number

Mr. S Pratap Singh
Asst. Professor
SPs IOKCOE, Pune, MH, India
pratap.singh.s@gmail.com

Dr. M. Ekambaram Naidu
Professor and Principal
SRK Institute of Technology, Vijayawada,AP, India
menaidu2005@yahoomail.co.in

***Abstract*—DNA-based cryptography is a new developing interdisciplinary area which combines various disciplines like computer science, cryptography, mathematical modeling, biochemistry and molecular biology. This paper firstly proposes a user authentication method using DNA sequence and secondly a novel method to secure data using DNA sequence and Armstrong number.**

***Keywords*— DNA Cryptography, Armstrong number, cryptosystem, RNA, DNA Sequence.**

I. INTRODUCTION

Information plays an important role in our day to day life. Providing security to information is again a big task. Information security is very important in today's world. Information security main goals are defined interms of CIA (Confidentiality, Integrity, and Availability). To implement the information security various techniques are proposed like cryptography and Steganography. Cryptography is a Greek word means “secret writing”. Cryptography is a science and art of transforming messages to make them secure and immune for attacks. Steganography is also a Greek word mean “covered writing”. In this techniques encryption and decryption process is used to hide simple data from unauthorized users by converting it into unreadable form and again retrieve it in original form. If the same key is used during the encryption process and decryption process, it is called as symmetric key cryptography, and if different key is used during encryption process one key and during decryption different then it is called as Asymmetric key cryptography [1].

To secure the data from unauthorized access and modification, we need data security. As the technology is upgrading very fast, even the DES and RSA algorithms have been cracked there is need to secure data which is transmitted over the network. The unsecured networks can be easily vulnerable to various kinds of attacks. To secure the data various traditional methods and techniques have been introduced, but still there is need of techniques that can stand for long time. Now a day, it's very important that new approach to data security is needed, if organization wants to stay ahead of attackers and more effectively secure their intellectual properties, data, employee or customers information.[2]

A. RGB representation

Any color is the mixture of three colors RGB (Red, Green and Blue) in preset quantities. This is nothing but a RGB representation. Here values for Red, Green and Blue represent each pixel. So any color can be individually represented with the help of three dimensional RGB cube. RGB model uses 24 bits, 8 bits for each color[2]. RGB color space or RGB color system, constructs all the colors from the combination of the Red, Green and Blue colors. The red, green and blue use 8 bits each, which have integer values from 0 to 255. This makes 256*256*256=16777216 possible colors[3][6] .

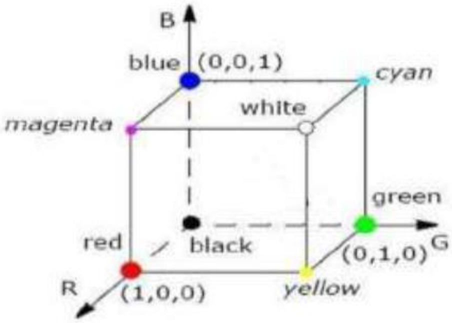


Fig:1 RBG Color image

B. DNA CRYPTOGRAPHY

DNA Cryptography is one of the rapid emerging technology which works on concept of DNA Computing. A new technique for securing data was introduced using the biological structure of DNA called DNA Computing. It was invented by Leonard Max Adleman in the year 1994 for solving the complex problems such as directed Hamilton path problem which is similar to travel salesman problem. Adleman is one of the inventors of RSA algorithm which have been named as RSA based on their names. DNA can be used to store and transmit data. The concept of using DNA computing in the fields of cryptography and Steganography has been identified as a possible technology that may bring forward a new hope for unbreakable algorithms.

Central dogma of molecular biology describes the flow of genetic information in cells from DNA to messenger RNA (mRNA) to protein. It states that genes specify the sequence of mRNA molecules, which in turn specify the sequence of proteins. DNA encodes RNA and RNA encodes proteins.

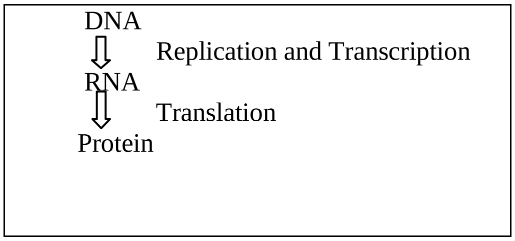


Fig.2. Central Dogma of Molecular biology [4]

DNA contains the complete genetic information that defines the structure and function of an organism. Proteins are formed using the genetic code of the DNA. Three different processes are responsible for the inheritance of genetic information and for its conversion from one form to another [4].

1. Replication: a double stranded nucleic acid is duplicated to give identical copies. This process perpetuates the genetic information.

2. Transcription: a DNA segment that constitutes a gene is read and transcribed into a single stranded sequence of RNA. The RNA moves from the nucleus into the cytoplasm.

3. Translation: the RNA sequence is translated into a sequence of amino acids as the protein

DNA and RNA are totally different. DNA, stands for Deoxyribo Nucleic Acid, is a long polymer with a deoxyribose and phosphate backbone and four different bases: adenine, guanine, cytosine and thymine (A,G,C,T), while RNA,stands for Ribo Nucleic Acid, is a polymer with a ribose and phosphate backbone and four different bases: adenine, guanine, cytosine and uracil (A,G,C,U). DNA can only be found in nucleus, while RNA can either be found in cytoplasm or nucleus. Generally DNA is a double-stranded molecule with a long chain of nucleotides, while RNA is a single-stranded molecule with a short chain of nucleotides. DNA is a medium of long-term storage and is used in transmission of genetic information. RNA is not. The nucleotide adenine will make a pair with thymine (A-T) and cytosine always makes pair with guanine (G-C). DNA Cryptography can be defined as a science of hiding data in the form of DNA sequence. This research paper firstly focus on various arithmetic methods based on DNA cryptography. Secondly, a novel method to secure data using DNA sequence and Armstrong number is proposed.

II. LITERATURE SURVEY

The various operation used by the researcher for operation on DNA sequence are Addition, subtraction, complementary, and XOR operation. Apart from this method of operation, other operation Apart from this method of operation, other operation also used likes substitution and transposition. The 0 and 1 are the binary numbers and 00 and 11 is a pair and similarly 10 and 01 is also pair, based on this the operation are performed. In this research paper A, G, C, T are replaced with 00, 01, 10, and 11 respectively. These are illustrated in the following tables[11], TABLE I, TABLE II.

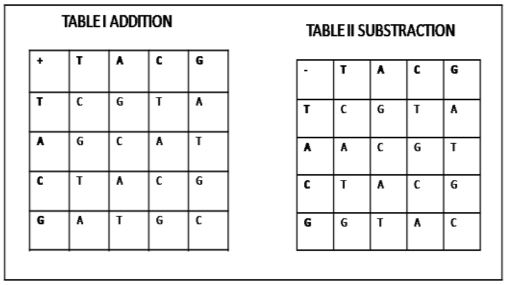


TABLE I ADDITION

+	T	A	C	G
T	C	G	T	A
A	G	C	A	T
C	T	A	C	G
G	A	T	G	C

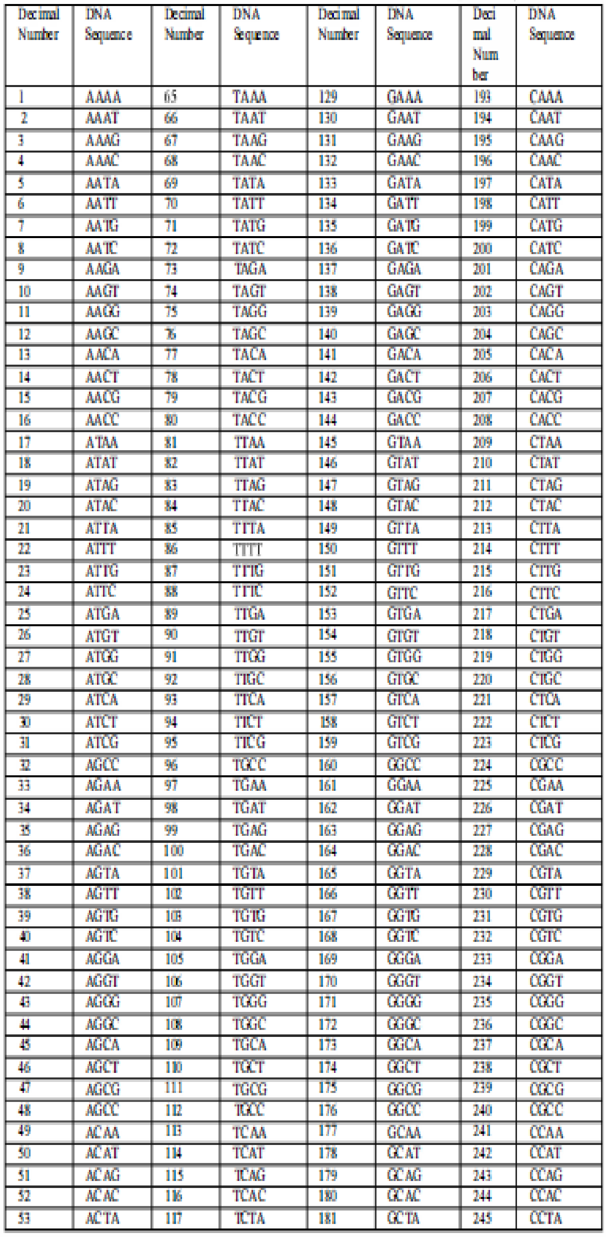
TABLE II SUBSTRACTION

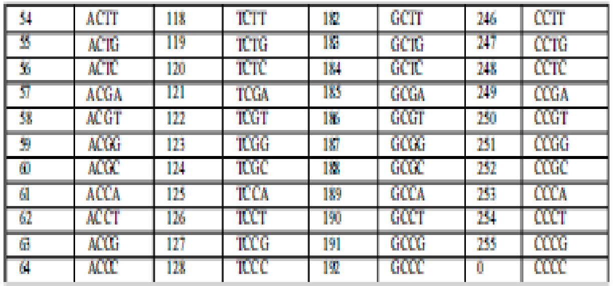
-	T	A	C	G
T	C	G	T	A
A	A	C	G	T
C	T	A	C	G
G	G	T	A	C

The various papers [6][11-15] had shown the data security using colors and Armstrong numbers.

The authors (Shipra Jain and Vishal Bhatnagar)[9] has prepared the table of 256 decimal numbers and their corresponding DNA sequence of length. There are possible 256 combinations of DNA nucleotides of 4 lengths. These 256 DNA sequence may vary from person to person. The author has chosen the length of DNA sequence of 4 to increase the key domain. These 256 decimal numbers and their corresponding DNA sequence will acts as a key between sender and receiver of data. The 256 decimal numbers and their corresponding DNA sequence are shown in table [3] below

TABLE III
DNA SEQUENCE DICTIONARY [10]





Decimal Number	DNA Sequence	Decimal Number	DNA Sequence	Decimal Number	DNA Sequence	Deci mal Num ber	DNA Sequence
1	AAAA	65	TAAA	129	GAAA	193	CAAA
2	AAAT	66	TAAT	130	GAAT	194	CAAT
3	AAAG	67	TAAG	131	GAAG	195	CAAG
4	AAAC	68	TAAC	132	GAAC	196	CAAC
5	AATA	69	TATA	133	GATA	197	CATA
6	AATT	70	TATT	134	GATT	198	CATT
7	AATG	71	TATG	135	GATG	199	CATG
8	AATC	72	TATC	136	GATC	200	CATC
9	AAGA	73	TAGA	137	GAGA	201	CAGA
10	AAGT	74	TAGT	138	GAGT	202	CAGT
11	AAGG	75	TAGG	139	GAGG	203	CAGG
12	AAGC	76	TAGC	140	GAGC	204	CAGC
13	AACA	77	TACA	141	GACA	205	CACA
14	AACT	78	TACT	142	GACT	206	CACT
15	AACG	79	TACG	143	GACG	207	CACG
16	AACC	80	TACC	144	GACC	208	CACC
17	ATAA	81	TTAA	145	GTAA	209	CTAA
18	ATAT	82	TTAT	146	GTAT	210	CTAT
19	ATAG	83	TTAG	147	GTAG	211	CTAG
20	ATAC	84	TTAC	148	GTAC	212	CTAC
21	ATTA	85	TTTA	149	GTTA	213	CTTA
22	ATTT	86	TTTT	150	GTTT	214	CTTT
23	ATTG	87	TTTG	151	GTTG	215	CTTG
24	ATTC	88	TTTC	152	GTTC	216	CTTC
25	ATGA	89	TTGA	153	GTGA	217	CTGA
26	ATGT	90	TTGT	154	GTGT	218	CTGT
27	ATGG	91	TTGG	155	GTGG	219	CTGG
28	ATGC	92	TTGC	156	GTGC	220	CTGC
29	ATCA	93	TTCA	157	GTCA	221	CTCA
30	ATCT	94	TTCT	158	GTCT	222	CTCT
31	ATCG	95	TTCG	159	GTCG	223	CTCG
32	AGCC	96	TGCC	160	GGCC	224	CGCC
33	AGAA	97	TGAA	161	GGAA	225	CGAA
34	AGAT	98	TGAT	162	GGAT	226	CGAT
35	AGAG	99	TGAG	163	GGAG	227	CGAG
36	AGAC	100	TGAC	164	GGAC	228	CGAC
37	AGTA	101	TGTA	165	GGTA	229	CGTA
38	AGTT	102	TGTT	166	GGTT	230	CGTT
39	AGTG	103	TGTG	167	GGTG	231	CGTG
40	AGTC	104	TGTC	168	GGTC	232	CGTC
41	AGGA	105	TGGA	169	GGGA	233	CGGA
42	AGGT	106	TGGT	170	GGGT	234	CGGT
43	AGGG	107	TGGG	171	GGGG	235	CGGG
44	AGGC	108	TGGC	172	GGGC	236	CGGC
45	AGCA	109	TGCA	173	GGCA	237	CGCA
46	AGCT	110	TGCT	174	GGCT	238	CGCT
47	AGCG	111	TGCG	175	GGCG	239	CGCG
48	AGCC	112	TGCC	176	GGCC	240	CGCC
49	ACAA	113	TCAA	177	GCAA	241	CCAA
50	ACAT	114	TCAT	178	GCAT	242	CCAT
51	ACAG	115	TCAG	179	GCAG	243	CCAG
52	ACAC	116	TCAC	180	GCAC	244	CCAC
53	ACTA	117	TCTA	181	GCTA	245	CCTA
54	ACTT	118	TCTT	182	GCTT	246	CCTT
55	ACTG	119	TCTG	183	GCTG	247	CCTG
56	ACTC	120	TCTC	184	GCTC	248	CCTC
57	ACGA	121	TCGA	185	GCGA	249	CCGA
58	ACGT	122	TCGT	186	GCGT	250	CCGT
59	ACGG	123	TCGG	187	GCGG	251	CCGG
60	ACGC	124	TCGC	188	GCGC	252	CCGC
61	ACCA	125	TCCA	189	GCCA	253	CCCA
62	ACCT	126	TCCT	190	GCCT	254	CCCT
63	ACCG	127	TCCG	191	GCCG	255	CCCG
64	ACCC	128	TCCC	192	GCCC	0	CCCC

III. PROPOSED METHOD

In the Proposed method, two level of security is provided, in the first level authentication is done and in the second level encryption of message is done. This method uses the DNA sequence Dictionary.
The system architecture is shown below.

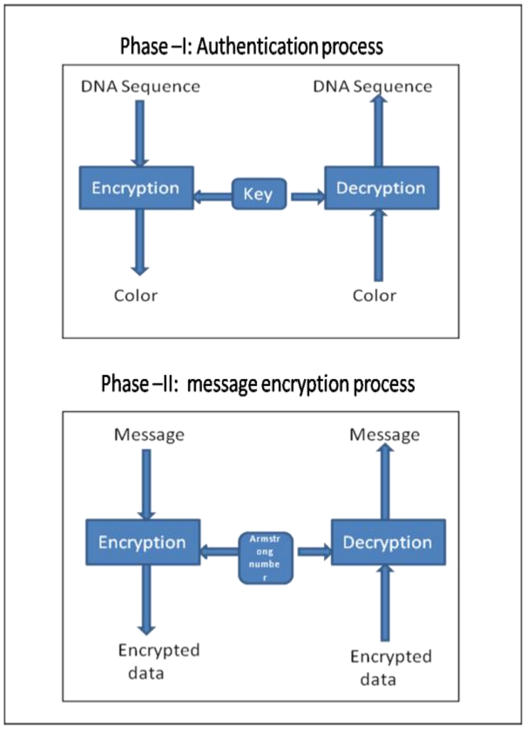


Fig. 3: System architecture

At fisrt level the steps are as follows:

A. Sender side

Step 1: Identify the unique receiver using a DNA sequence.
Step 2: convert the DNA sequence to Decimal numbers
Step 3: Add triplets' values to decimals numbers
Step 4: Then convert this new generated values to RBG values (0-255)
Step 5: This RBG will gives a color, and this color will act as a password for authentication.

B. Receiver side

Step 1: The receiver will receive a color as password for authentication.
Step 2: convert the color in to RBG values.
Step 3: Subtracts triplets' values from the RBG values.
Step 4: then convert this new generated values to DNA equivalent (0-255)

Step 5: This DNA equivalent values forms a DNA sequence.

The second level is data security which uses the Armstrong number

1. *The Encryption process:*

Step1: Assume that the information to be send to the receiver as plaintext (it can be a text, image, audio or video)
Step 2: Convert this plain text to ASCII values
Step 3: Add Armstrong number to ASCII values
Step4: Convert this newly generated values to into 8x8 Matrix
Step 5: Read the matrix either in a row wise or column wise by replacing a row or columns ASCII value with the DNA sequence as in DNA sequence matrix.

II. The Decryption process is the reverse of the encryption process at the receiver side.

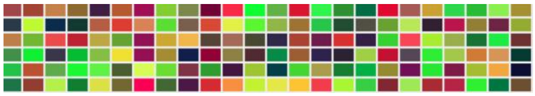
IV. ILLUSTRATION

A. Authentication Process:

To illustrate the technique, consider a DNA sequence as
“ATGGCGCAGGTA TTAGACGTACGCCTAGCTCCATGGACC…..”
Read first four character of the sequence as ATGG, CGCA, GGTA, and so on. Convert these DNA sequence to Decimal equivalent as shown in DNA Sequence Dictionary, then add triplets(5, 10, and 15) to this decimals numbers and it will generate new values (32, 247, and 180)and represent this values to RGB values, this RGB values gives us a color[7]. So this color will act as password.

TABLE IV

27	ATGG
237	CGCA
165	GGTA
83	TTAG
58	ACGT
60	ACGC



Once the color is received by the receiver, he decrypts its RBG values (32, 247,180) and subtracts the triplets (5, 10, and 15) to generate original value (27,237,165) to generate the DNA sequence as
“ATGGCGCAGGTA TTAGACGTACGCCTAGCTCCATGGACC…..”

B. Message encryption and decryption process:

Step 1: Assume the message that need to be transmitted be “SECURITYTECH“
Step: 2 find the ASCII equivalent values of all the characters[8]

S	E	C	U	R	I	T	Y	T	E	C	H
83	69	67	85	82	73	84	89	84	69	67	72

Step 3: Perform the addition of Armstrong number with this numbers as follows

83	69	67	85	82	73	84	89	84	69	67	72
3	7	1	9	49	1	27	343	1	3	7	1
86	76	68	94	131	74	111	432	85	72	74	73

Step 4: Convert these newly generated values as new ASCII values and generate its DNA sequence as: “TTTT TAGC TAAC TTCT GAAG TGCG TATC GCAA”

System Flow chart

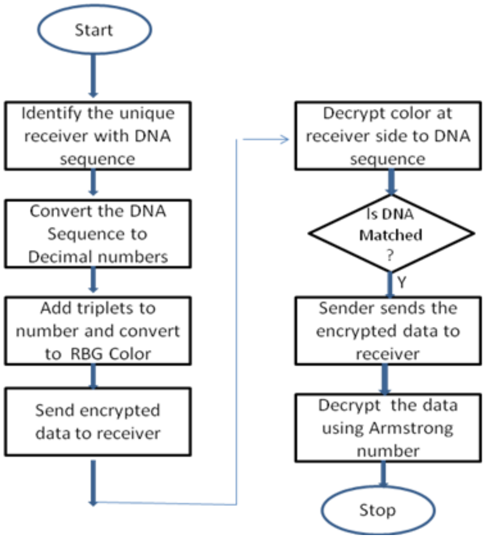

Start
Identify the unique receiver with DNA sequence
Convert the DNA Sequence to Decimal numbers
Add triplets to number and convert to RBG Color
Send encrypted data to receiver
Decrypt color at receiver side to DNA sequence
Is DNA Matched ?
Y
Sender sends the encrypted data to receiver
Decrypt the data using Armstrong number
Stop

Fig: 4 System flow chart

V. EXPERIMENTAL RESULSTS

The implementation of the proposed method is done in java language. In the paper, the message is a plaintext file, image, audio and video. The author had used online tools to generate the results. The tools like image2Ascii, ascgen to convert the image into ASCII image.

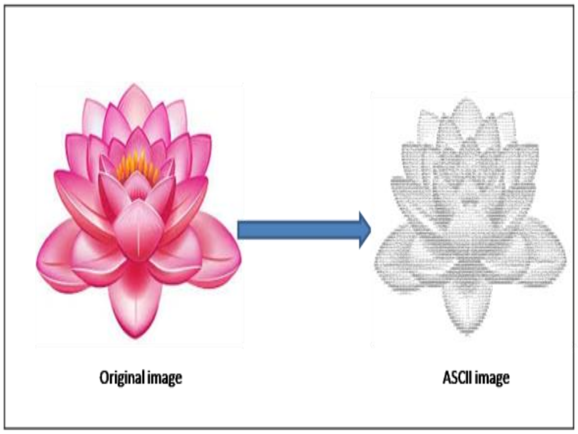


Fig: 5 Original images to ASCII Image [9]

This text file is converted to ASCII values and a Add Armstrong number to ASCII values, Convert this newly generated values to into 8x8 Matrix form Then Read the matrix either in a row wise or column wise by replacing a row or columns ASCII value with the DNA sequence as in DNA sequence matrix. The proposed system accepts input files as text, image, audio and videos. The output will be s given below sequence.

"TTTT TAGC TAAC TTCT GAAG TGCG TATC GCAA"

VI. CONCLUSION

Thus the authors used DNA sequence and Triplets as key for authentication and generated the color as password to the receiver.

The receiver decrypts the color to generate the DNA sequence and if the DNA sequenced is matched then he is authenticated to decrypt the message. The second phase of the proposed system an encryption process is done for data security using Armstrong number and again new DNA sequence is generated as Cipher text. Thus this proposed system provides two level of security. This system can be used at some confidential areas were security is given more importance, as DNA sequence Colors, key values and Armstrong numbers which are three set of keys in this technique makes sure that there is secured message or data transmission and is available to authorized person only.

VII. REFERENCES

1. Cryptography and Network Security by Behrouz A. Forouzan. The MC Graw - Hills companies.
2. International Journal of Science, Technology & Management www.ijstm.com Volume No.04, Issue No. 02, February 2015 ISSN (online): 2394-1537.
3. securityaffairs.co/wordpress/33879/security/dna-cryptography.html
4. http://users.ugent.be/~avierstr/pdf/principles.pdf "The Central Dogma: DNA Encodes RNA and RNA Encodes Protein." Boundless Biology. Boundless, 26 May. 2016. Retrieved 23 Oct. 2016 from https://www.boundless.com/biology/textbooks/boundless-biology-textbook/genes-and-proteins-15/the-genetic-code-106/the-central-dogma-dna-encodes-rna-and-rna-encodes-protein-441-11665/
5. S. Pavithra Deepa,S. Kannimuthu, V. Keerthika., "Security Using Colors and Armstrong Numbers", Proceedings of the National Conference on Innovations in Emerging Technology-2011. India.17 & 18 February, 2011.pp.157-160.
6. http://www.colorcodepicker.com/
7. http://www.rapidtables.com/web/color/RGB_Color.htm
8. Ascgen software accessed from uri http://ascgendotnet.jmsoftware.co.uk/ on 2nd feb, 2014
9. "Shipr Jain and Dr. Vishal Bhatnagar" A Novel DNA Sequence Dictionary method for Securing Data in DNA using Spiral Approach and Framework 0/ DNA Cryptography. IEEE International Conference on Advances in Engineering &Technology Research (ICAETR - 2014), August 01-02.
10. Fasil K. A and Deepthy Antony, " A Multiphase Cryptosystem with Secure Key Encapsulation Scheme Based on Principles of DNA Computing" Fourth International Conference on Advances in Computing and Communications 2014.
11. Gordon L. Miller and Mary T. Whalen," Armstrong Numbers ",University of Wisconsin, Stevens Point, WI 54481 (Submitted October 1990).
12. S.Belose, M.Malekar , G.Dharmawat,"Data Security Using Armstrong Numbers", International Journal of Emerging Technology and Advanced Engineering.Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 4, April 2012).
13. M.F.Armstrong, "A brief introduction to Armstrong Numbers"
14. Chavan Satish, Lokhande Yogesh, Shinde Pravin, Yewale Sandeep,Sardeshpande S. A, "Secure Email using Colors and Armstrong Numbers over web services", International Journal Of Research In Computer Engineering And Information Technology VOLUME 1 No. 2.
15. M.Renuga Devi, S.Christobel Diana, "Enhancing Security in Message Passing Between Sender and Receiver Using Colors and Armstrong Numbers", International Conference on Computing and Control Engineering(ICCCE 2012), 12 & 13 April, 2012