Integrating DNA Computing in International Data Encryption Algorithm (IDEA)

Pankaj Rakheja
Deep institute of Engineering
& Tech (DIET)
Department of EECE
Rithoda, Haryana

ABSTRACT
DNA cryptography is a field which is being explored worldwide and even in being in its primitive stages it has revolutionized the area of network security. Here we are applying DNA computing on well known IDEA algorithm for making it more secure. We are adding a layer of DNA cipher over the basic IDEA algorithm. The cipher now will be in form of DNA sequence which will even hide very existence of the underlying IDEA algorithm. Key space has been extended a bit to make it more immune to cryptanalytic attacks. It can transmit highly confidential data efficiently and securely. The Matlab has been used for implementation.

Keywords
Cipher, Data, IDEA

1. INTRODUCTION
Information flows throughout the network which may be local or of global scope. It is mandatory to secure that information to prevent unauthorized access of it by any node in path. We need to ensure a right security infrastructure which opens up just enough doors that are mandatory protecting everything else. We need to ensure privacy, integrity and confidentiality in the network for it to be reliable and dependable for information exchange for that we encode the data before sending it through various encoding mechanisms available to make it unreadable. This is where the cryptography comes into picture.

Cryptography[6] [7] [13] [20] is the art and science of achieving security by encoding the simple message to make it unreadable .There are basically two types of cryptography that is a techniques for converting plaintext to cipher text and vice versa which are namely symmetric and asymmetric cryptography. In symmetric cryptography sender and receiver use the same key for encryption and decryption of text whereas in asymmetric cryptographic systems two different keys namely public and private keys are used for encryption and decryption process.

Cryptography mechanisms rely on the degree of randomness and uncertainty in the generation of the cipher from the plaintext for which many phenomenon of nature are and have been used for instance quantum cryptography utilizes the randomness of states of electron inside an atom, Elliptical cryptography makes use of algebraic structure of elliptic curves over finite fields and moreover DNA cryptography exploits the extreme complexity and randomness in the DNA structure for coding and decoding.

DNA computing and cryptography came into picture in 1990s. DNA computing was initialized by L Adleman [4] in 1990 where he solved a directed Hamiltonian path problem , it indicated the feasibility of a molecular approach to solve combinatorial problems, W Li combined recombinant DNA technology in 1994, Lipton R J extended the adleman approach to solve another NP problem in 1995, Boneh D, Dunworth C, Lipton R broke DES using molecular computations in 1995, Adleman L extended DNA computers to RNA used for breaking DES in 1996, C. T. Clelland, V. Risca, C. Bancroft cryptography has been shown to be one of the new applications of DNA computing in 1999, Gehani A, LaBean T H, Reif J H used discrete mathematics for designing a DNA cryptographic mechanism in 2000, P. L. Cox J find that the vast parallelism, exceptional energy efficiency and extraordinary information density are inherent in DNA molecules in 2001, Jie Chen proposed a novel design of DNA-based, molecular Cryptography design [12] Carbon nanotube-based message transformation, and DNA-based cryptosystem an proposed. To demonstrate the performance, we present an interesting example lo encode and decode images wing the proposed scheme in 2003 ,Kartalopoulos S.V. initialized DNA cryptography in optics [17] in 2005 , Xing Wang, Qiang Zhang* use a new way to show how cryptography works with DNA computing, it can transmit message securely and effectively. The RSA algorithm belongs to asymmetric key cryptography, it is used in this paper connecting with DNA computing technique to encrypt message in 2009

We have integrated DNA computing in well known IDEA [2] [8] [9] algorithm to make it more efficient and effective from cryptanalytic point of view. That is it becomes more immune to general attacks which a cryptographic system encounters in day to day scenarios. Many researchers have integrated new mechanisms in IDEA [3][5][16][22] like chaotic series, modular arithmetic and VLSI implementations etc.

IDEA is one of the strongest cryptographic algorithms. It was launched in 1990 and underwent certain changes in name and Capabilities as shown in table below.

Table 1: Brief history of IDEA

.

Year	Name	Description
1990	Proposed Encryption Standard (PES)	Developed at the Swiss Federal Institute of Technology
1991	Improved Proposed Encryption Standard (IPES)	Improvements made to make it secure to cryptanalytic attacks
1992	International Data Encryption Algorithm(IDEA)	No major changes, Simply remained

Although its quite strong but not that popular as DES and AES because it is patented and it does not have good track record available. It is a block cipher which works on 64 bit plaintext blocks with 128 bit key to encode the data. It is also a reversible algorithm that is same algorithm can be used for encryption and decryption process. It employ both confusion and diffusion techniques for encryption. The 64-bit plaintext block is partitioned into four 16-bit sub-blocks six 16-bit key are generated from the 128-bit key. Since a further four 16-bit key-sub-blocks are required for the subsequent output transformation, a total of 52 (= 8 x 6 + 4) different 16-bit sub-blocks have to be generated from the 128-bit key.

First, the 128-bit key [14] is partitioned into eight 16-bit sub-blocks which are then directly used as the first eight key sub-blocks. The 128-bit key is then cyclically shifted to the left by 25 positions, after which the resulting 128-bit block is again partitioned into eight 16-bit sub-blocks to be directly used as the next eight key sub-blocks. The cyclic shift procedure described above is repeated until all of the required 52 16-bit key sub-blocks have been generated

The first four 16-bit key sub-blocks are combined with two of the 16-bit plaintext blocks using addition modulo 2^{16}, and with the other two plaintext blocks using multiplication modulo 2^{16} + 1. At the end of the first encryption round four 16-bit values are produced which are used as input to the second encryption round .The process is repeated in each of the subsequent 7 encryption rounds. The four 16-bit values produced at the end of the 8th encryption round are combined with the last four of the 52 key sub-blocks using addition modulo 2^{16} and multiplication modulo 2^{16} + 1 to form the resulting four 16-bit cipher text blocks .

The computational process used for decryption of the cipher text is essentially the same as that used for encryption. The only difference is that each of the 52 16-bit key sub-blocks used for decryption is the inverse of the key sub-block used during encryption. In addition, the key sub-blocks must be used in the reverse order during decryption in order to reverse the encryption process

IDEA supports all modes of operation such as:

- Electronic Code Book (ECB) mode
- Cipher Block Chaining (CBC)
- Cipher Feedback (CFB)
- Output Feedback (OFB) modes

We have worked on IDEA as Today, there are hundreds of IDEA-based security solutions available in many market areas, ranging from Financial Services, and Broadcasting to Government .The IDEA algorithm can easily be embedded in any encryption software. Data encryption can be used to protect data transmission and storage. Typical fields are: Audio and video data for cable TV pay TV, video conferencing, distance learning; Sensitive financial and commercial data; Email via public networks and Smart cards

It relies on extension of key to be used for each round using circular left shift operations. And then getting sub keys from the new key generated at each round. We have worked on the plaintext given to the IDEA algorithm which is originally in simple alphabets which undergoes DNA encryption to give a 64 bit usual input to the IDEA which further works on it with an additional round to convert the 64 bit cipher to DNA sequence to give it a new shape then the cipher undergoes decryption at two steps first through IDEA then DNA. Thus it provides a protective layering over basic IDEA along with changing the cipher generated by IDEA adding more uncertainty to the cryptographic mechanism used.

2. OVERVIEW

DNA [4][11][15][18][19] is found in the nucleus of every human cell the information in DNA: guides the cell (along with RNA) in making new proteins that determine all of our biological traits and gets passed (copied) from one generation to the next. Thus, it carries design information between generations, and thus accounts for inherited biological traits. These molecules contain the designs for all the material components that a living organism needs for growth, development, and daily living. . It is the major source of the genetic information of any living organism in the biosphere and is composed of two long strands of nucleotides bases arranged in a helix sort of structure as shown in figure 1 below.

Figure 1. Basic DNA shape

The designs are known as *genes.* Some genes play a role in regulating other genes, and some design *ribonucleic acid*, a close relative of DNA. But mostly, the designs in DNA are for the class of chemicals called proteins. The human body contains

tens of thousands of kinds of proteins, which do all the body's work. Interactions among those proteins and interactions between them and environmental factors account for the processes and structures of the body. Those processes and structures are manifested as inherited traits. DNA is comprised of chains of chemical subunits called *nucleotides*, each of which contains one nitrogenous base: *adenine* (*A*), *thymine* (*T*), *cytosine* (*C*), or *guanine* (*G*). The design instructions in DNA are spelled out as particular sequences of these four bases.. In the case of genes, there are only four letters in the alphabet. Hundreds of nucleotides are linked in a DNA chain in a sequence that spells out instructions for a single gene.

There are two complementary chains in the structure of DNA. Each nucleotide in DNA has a sugar component joined to a phosphate group at one point on the sugar, and to a nitrogen-containing base attached at another point as shown in figure1 below

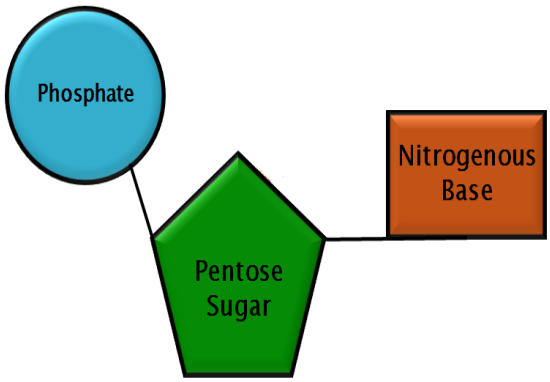

Figure 2. Nucleotide structure

The chains in DNA have the phosphate of one nucleotide linked to the sugar of the next nucleotide to form a strand of alternating sugars and phosphates with dangling nitrogenous bases as shown below in figure2

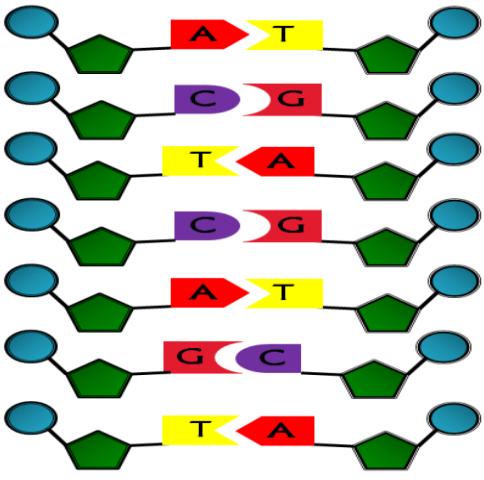

Figure 3. Combination of nucleotide bases in strands

DNA contains two such chains, twisted around each other to form a double-stranded helix with the bases on the inside. Every *A* on one chain forms weak bonds with a *T* on the other strand, and every *C* on a strand bonds weakly to a *G* on the opposite chain. The two strands, held together weakly by the pairing of *A* with *T*, and *G* with *C*, are thus complementary, and the sequence in one can be deduced from the other's sequence. The basic DNA structure is shown below

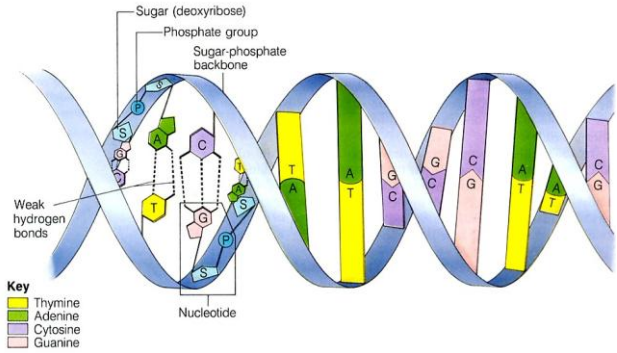

Figure 4. Basic DNA structure

Design information is transmitted as new DNA to new cells during development and growth. The complementarily of the two DNA strands allows their information to be copied. Each old strand is used as a template in synthesizing a new complementary one.

These complimentary strands have codons as fundamental building blocks. Codons are basically triplets of nucleotide bases. Figure below shows codons forming DNA sequences in two complimentary strands

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

Figure 5. Complimentary DNA strands

As can be seen DNA nucleotide bases are existing in form of codons and are complimentary to each other that is A-T,G- C are complimentary to each other.

We can use these codons for encoding and decoding of the data.

3. MECHANISM DESIGNED

Here we are proposing a variation in the basic IDEA algorithm structure to make it more secure and efficient. We introduce DNA computing in IDEA. The plaintext to be encoded is first given to DNA encryption algorithm [23] which first generates a unique decimal number in 0-255 range corresponding to each letter of the plaintext of length 8. Then the cipher of it is converted to binary making it 64 bit usual plaintext of IDEA algorithm which encodes it using 128 bit key. And the final cipher of 64 bit is first represented in hexadecimal form which then mapped to a secret lookup table mapping hexadecimal value to codon pair. Then the final cipher in DNA sequence form is sent. That's what happens at sender side as shown in figure4 below that is the flow chart implemented at sender side.

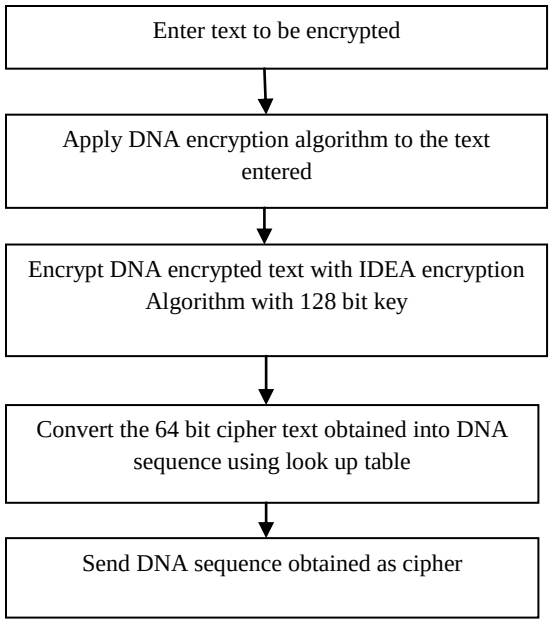

Figure 6. Encryption process at sender side

At the receiver side we first get the 64 bit cipher to be used by IDEA decryption algorithm from the DNA sequence received using a lookup table mapping codons to their hexadecimal equivalents. Then the IDEA decrypts the message which still remains unreadable so needs to be decrypted using DNA decryption algorithm [23] which returns the original plaintext. Figure5 shows the flowchart implemented at receiver side.

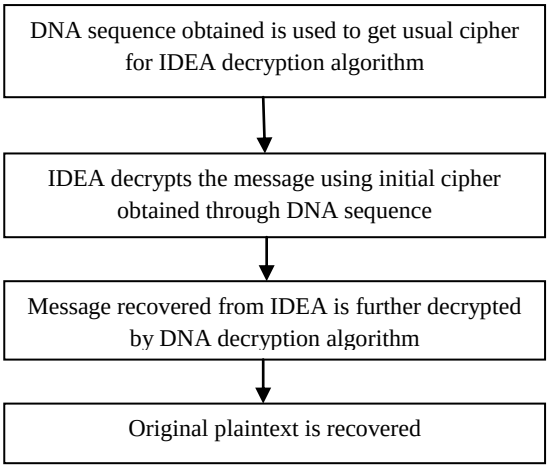

Figure 7. Decryption process at the receiver side

The DNA encryption algorithm takes 8 letter text to be encrypted and generates a 6*1 key which is extended using the key extension process employed in AES [7][10][20] key extension. Then each letter is replaced by codon corresponding to it using lookup table 2 [21]. Then numeral value corresponding to each nucleotide base is obtained using lookup table 3 [21]. Then corresponding matrix of alphabet obtained is multiplied with the key corresponding to it to get a unique decimal value which is made to be in range 0 to 255 using mod 255 operations. Figure6 shows the DNA encryption process which repeats for each letter of the plaintext entered that is the flow chart of the DNA encryption algorithm which is then followed by table 2 and table 3 which form the basis of introducing DNA sequence in the cipher mechanism designed here we actually use the simple substitution of the plaintext in alphabetic form to codons that is base nucleotide triplets then these codons are converted to numerals using table 3 which has the numeral corresponding to each base in the codon than we apply matrix based operations with the matrix of the numeral obtained and the key for that corresponding letter of the text to get a unique encoded symbol corresponding to it.

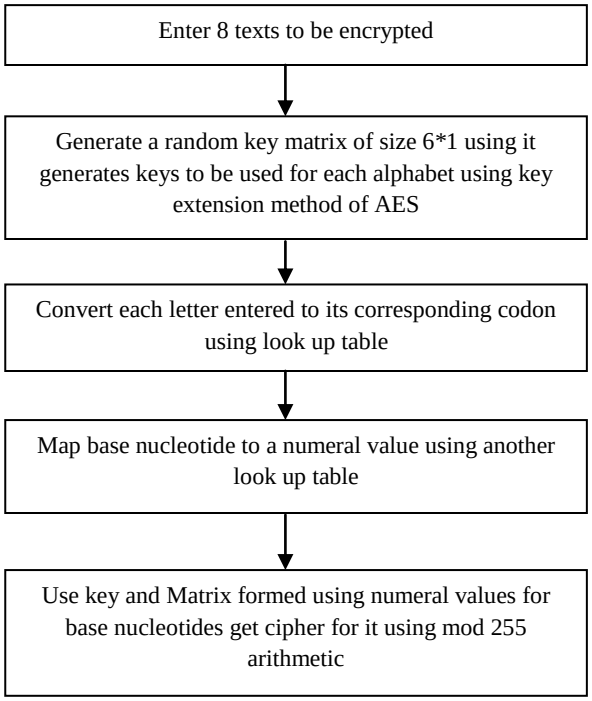

Figure 8. DNA encryption process

Table 2:Alphabet to codon mapping table

S.No	Alphabet	codon	S.No	Alphabet	Codon
1	A	CCA	14	N	TCT
2	B	GTT	15	O	CGG
3	C	TTG	16	P	ACA
4	D	GGT	17	Q	CAA
5	E	TTT	18	R	ACT
6	F	TCG	19	S	GCA
7	G	CGC	20	T	CTT
8	H	ATG	21	U	GTC
9	I	AGT	22	V	TCC
10	J	CGA	23	W	GCC
11	K	GAA	24	X	ATC
12	L	CGT	25	Y	AAA
13	M	CCT	26	Z	TCA

Table 3: Mapping nucleotide base to numerical value

S.No	Base Nucleotide	Numeric value
1	A	01
2	C	03
3	G	07
4	T	20

DNA decryption process obtains message decrypted from IDEA then decrypts it using the extended key retrieved from the original key used through key sharing mechanism. Then the corresponding numeral value matrix obtained using the cipher and the key. Then that is mapped to corresponding base using lookup table 4 and then the lookup table 5 maps the codons obtained to corresponding alphabets figure7 shows the DNA decryption process.

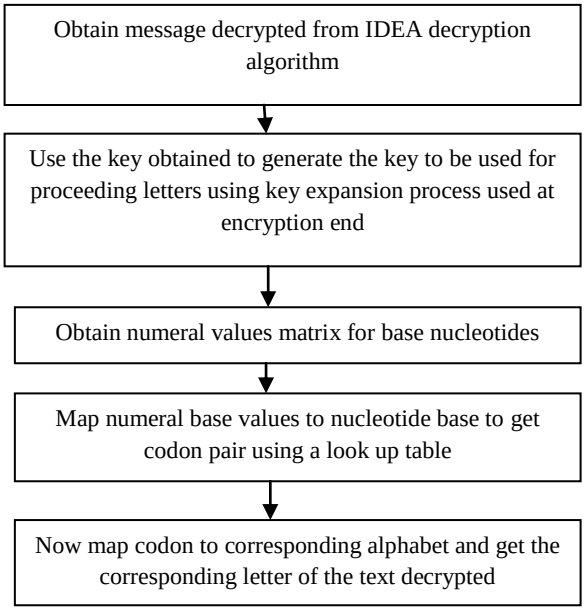

Figure 9. DNA decryption process

Table 4: Lookup table for mapping numeric value to base nucleotide

S.No	Numeric value	Base Nucleotide
1	01	A
2	03	C
3	07	G
4	20	T

Table5: Lookup table for mapping codon to Corresponding Alphabet

S.No	codon	Alphabet	S.No	Codon	Alphabet
1	CCA	A	14	TCT	N
2	GTT	B	15	CGG	O
3	TTG	C	16	ACA	P
4	GGT	D	17	CAA	Q
5	TTT	E	18	ACT	R
6	TCG	F	19	GCA	S
7	CGC	G	20	CTT	T
8	ATG	H	21	GTC	U
9	AGT	I	22	TCC	V
10	CGA	J	23	GCC	W
11	GAA	K	24	ATC	X
12	CGT	L	25	AAA	Y
13	CCT	M	26	TCA	Z

4. RESULTS

For carrying out implementation of the cryptographic mechanism designed we have used Matlab, which is a matrix-oriented programming language, perfectly suited for the Matrix based data structure and modular operations in IDEA and even DNA cipher designed here also relies on Look up tables and Matrix based computations. The plaintext after undergoing encryption and decryption process employed here has successfully been recovered. We have encrypted the plaintext and run the cipher. Here in the snapshot we first have the plaintext then cipher generated by DNA encryption algorithm, then the plaintext generated for IDEA then the cipher generated by idea in char form then final DNA sequence cipher which will be sent to receiver then it shows the IDEA decrypted message followed by original message retrieved by decoding through DNA decryption algorithm.

```
Enter text to be encrypted in CAPS AAAASSSS
Cipher=
234  213  192   171  55  15  230  190
Plaintext=
11101010
11010101
11000000
10101011
00110111
00001111
11100110
10111110
CipherIDEA =
Óp□□–□□ő
DNAcipher=
GGGCTACCCTTCGCGTTATTACCCCTCGGGTTCCT
AAAGCCCGGAGTA
RecoveredtextIDEA =
234  213  192   171  55  15  230  190
Plaintext_recovered =
AAAASSSS
```

It shows the original message entered then cipher obtained through DNA encryption algorithm which is converted to 64 bit plaintext for IDEA which then decodes it to get the cipher for DNA decryption algorithm which then retrieves the original message sent

5. CONCLUSION AND FUTURE SCOPE

DNA cryptography is the future of the information security. Its complexity and randomness provides a great uncertainty which makes encoding of data in DNA format better than other mechanism of cryptography. And on integrating it with a well known symmetric cryptographic mechanism that is IDEA makes it very difficult to decode the data without precise knowledge of the key. It even hides the presence of IDEA as the final cipher is in form DNA sequence. And the key space of IDEA has also been extended now the receiver needs to know the 48 bit key used for DNA encryption along with the 128 bit IDEA key and it also engulfs the key extension mechanism employed by the AES algorithm. So it has great computational complexity along with the usage of mechanisms of good cryptographic algorithms like AES, IDEA etc.

Future work comprises of analyzing its performance to basic cryptanalytic attacks and comparing it with IDEA to knowing exactly how much improvement is achieved.

6. REFERENCES

[1] Meier, W., On the Security of the IDEA block cipher, Advances in Cryptology,1990

[2] Lai, Xuejia, and Massey, James L., A Proposal for a New Block Encryption Standard, Advancesin Cryptology - EUROCRYPT '90, Lecture Notes in Computer Science, Springer-Verlag, 1991:389-404.

[3] Lai, X., Massey, J., and Murphy, S., Markov Ciphers and Differential Cryptanalysis, Advances in Cryptology – EUROCRYPT '91, Lecture Notes in Computer Science, Springer-Verlag, 1991:17-38.

[4] L. Adleman, "Molecular computation of solutions to combinatorial problems," *Science, JSTOR*, vol. 266, pp. 1021–1025, 1994.

[5] Wolter, S.; Matz, H.; Schubert, A.; Laur, R.; "On the VLSI implementation of the international data encryption algorithm IDEA" , IEEE International Symposium on Circuits and Systems, 1995. ISCAS '95,

[6] Piper," Basic principles of cryptography" , IEEE Colloquium on Public Uses of Cryptography, 1996

[7] B. Schneier, "Applied Cryptography: Protocols, Algorithms, and SourceCode in C", John Wiley & Sons, Inc, 1996

[8] Menezes, A., van Oorschot, P., and Vanstone, S. 1996. Handbook of App

[9] Mediacrypt AG, The IDEA Block Cipher, submission to the NESSIE Project,http://cryptonessie.org,2000

[10] Sanchez-Avila, C. Sanchez-Reillol, "The Rijndael block cipher (AES proposal) : a comparison with DES" 2001 IEEE 35th International Carnahan Conference on Security Technology

[11] J. Chen, "A DNA-based, biomolecular cryptography design," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2003, pp. 822–825

[12] Jie Chen ,"A DNA-based, biomolecular cryptography design ",Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS 03

[13] Sakalli, M.T. Bulus, E. Buyuksaracoglu, F," Cryptography education for students", ITHET 2004 Proceedings of the Fifth International Conference on information Technology Based Higher Education and Training, 2004.

[14] "IDEA International Data Encryption Alrorithm" CS-627-1 fall 2004 by How shen Chang

[15] P. Rothemund, N. Papadakis, and E. Winfree, "Algorithmic self-assembly of DNA sierpinski triangles," *PLoS Biology*, vol. 2, no. 12, pp. 2041– 2053, 2004.

[16] Gwo-Ruey Yu,Secure communication using H∞ chaotic synchronization and international data encryption algorithm , American Control Conference, 2004. Proceedings of the 2004

[17] Kartalopoulos, S.V.;," DNA-inspired cryptographic method in optical communications, authentication and data mimicking ",Military Communications Conference, 2005. MILCOM 2005. IEEE

[18] R. Barish, P. Rothemund, and E. Winfree, "Two computational primitives for algorithmic self-assembly: copying and counting," *Nano Letters*, vol. 5, no. 12, pp. 2586–2592, 2005.

[19] G. Cui, Y. Liu, and X. Zhang, "New direction of data storage: DNA molecular storage technology," *Computer Engineering and Application*,vol. 42, no. 26, pp. 29–32, 2006.

[20] Atul Kahate ,"Cryptography and Network Security", Tata Macgraw Hill,2009

[21] Xing Wang, Qiang Zhang ,"DNA computing-based cryptography", Fourth International Conference on Bio-Inspired Computing, 2009.

[22] Modugu, R.; Yong-Bin Kim; Minsu Choi, "Design and performance measurement of efficient IDEA (International Data Encryption Algorithm) crypto-hardware using novel modular arithmetic components ",Instrumentation and Measurement Technology Conference (I2MTC), 2010 IEEE

[23] Pankaj Rakheja; Amanpreet kaur "A Unique Cryptographic Mechanism for Encoding Data Using DNA Structure", in International conference on Network Communication and Computers (ICNCC 2011) organized and sponsored by IACSIT, The Institute of Electrical and Electronics Engineers (IEEE), Singapore Institute of Electronics and other organizations.