International Journal of Science Engineering and Advance Technology

Cryptography Based on DNA Using Random key Generation Scheme

P.Surendra Varma, K.Govinda Raju
Department of Computer Science and Engineering
Srinivasa Institute of Engineering and Technology, Cheyyeru
suren548@gmail.com, govindarajukynm@gmail.com

Abstract— With the growth of technological advancements, the threats dealt by a user grow exponentially. The 21st century is a period of information explosion in which information has become a very important strategic resource, and so the task of information security has become increasingly important in data storage and transmission. As traditional cryptographic systems are now vulnerable to attacks, the concept of using DNA Cryptography has been identified as a possible technology that brings forward a new hope for unbreakable algorithms. A new field of cryptography is emerging based on DNA computing due to high storage capacity, vast parallelism and exceptional energy efficiency of biological DNA. This field is in initial stage so a lot of research has to be done yet. This paper analyzes the different approach on DNA Cryptography based on matrix manipulation and secure key generation scheme.

Keywords— DNA, DNA computing, DNA cryptography, Matrix manipulation, Key generation.

I. INTRODUCTION

Cryptography is one of the ways of improving security of information by scrambling the data in a way that message becomes non-readable(Cipher text) to an intruder. So that the data can't be read or modified by third party. More generally, it is about constructing and analyzing protocols that overcome the influence of adversaries and which are related to various aspects in information security such as data confidentiality, data integrity, authentication, and non-repudiation. Modern cryptography intersects the disciplines of mathematics and computer science. cryptography referred almost exclusively to encryption, which is the process of converting ordinary information (called plaintext) into unintelligible text (called ciphertext). Decryption is the reverse, in other words, moving from the unintelligible ciphertext back to plaintext. A cipher (or cypher) is a pair of algorithms that create the encryption and the reversing decryption. The detailed operation of a cipher is controlled both by the algorithm and in each instance by a "key". This is a secret (ideally known only to the communicants), usually a short string of characters, which is needed to decrypt the ciphertext. A "cryptosystem" is the ordered list of elements of finite possible plaintexts, finite possible cyphertexts, finite possible keys, and the encryption and decryption algorithms which correspond to each key. Keys are important, as ciphers without variable keys can be trivially broken with only the knowledge of the cipher used and are therefore useless for most purposes. Historically, ciphers were often used directly for encryption or decryption without additional procedures such as authentication.

DNA computing is a new field which is growing in the modern days. The journey of DNA cryptography started with the development of DNA computing. DNA computing was introduced by L. Adleman [2] in the year of 1994 to solve the complex computational problem. In his study he found that DNA has high storage and computational capability. Using DNA computation, he solved a searching problem named directed Hamiltonian path problem with seven vertices where he assumed molecules as vertices and encoded them in a molecule sequence and performed computations by chemical operations in lab.

This is similar to traveling salesman path problem where large possible solutions generated to find better paths to reach from source to destination. An image encryption algorithm based on DNA sequence addition operation is presented by Wang et. al. A DNA sequence matrix is obtained by encoding the original image and it is divided into some equal blocks and two logistic maps, DNA complementarity and DNA sequence addition operation are utilized

to add these blocks. A DNA sequence matrix is decoded to get the encrypted image DNA computing also called as Biomolecular computing. The reason behind taking DNA in to account as a computation medium is its high storage capacity, vast parallelism and high energy efficiency. Due to these capabilities DNA can act as processor which can process large amount of data and can perform computations to get several possible solutions.

A gram of DNA contains 1021 DNA bases which is nearly equal to 108 tera-bytes of data, so it can be clearly observed that it vastly exceeds the capacity of traditional storage media such as electronic, optical, magnetic media etc. DNA's high storage capacity and its vast parallelism proved it a new medium of information and its computing power explored new ideas for solving complex mathematical problems. Guangzhao Cui et. al. encryption scheme is designed by using the technologies of DNA synthesis, PCR amplification, DNA digital coding and the theory of traditional cryptography [5].

Biological difficult issues and cryptography computing difficulties provide a double security safeguards for the scheme. Souhila Sadeg et. al. proposed a symmetric key block cipher algorithm including that simulation ideas from the processes of transcription (from DNA to mRNA) and translation (from mRNA into amino acids) [13]. This algorithm is believed to be efficient in computation and very secure.

II. BIOLOGICAL STUDY

DNA stands for Deoxyribo Nucleic Acid. DNA represents the genetic blueprint of living creatures. DNA contains "instructions" for assembling cells. Every cell in human body has a complete set of DNA. DNA is unique for each individual. DNA) is a molecule that encodes the genetic instructions used in the development and functioning of all known living organisms and many viruses. DNA is a nucleic acid; alongside proteins and carbohydrates, nucleic acids compose the three major macromolecules essential for all known forms of life. [8]. In 1953, James Watson discovered the structure of DNA. Most DNA molecules consist of two biopolymer strands coiled around each other to form a double helix as depicted in fig.1 [9].

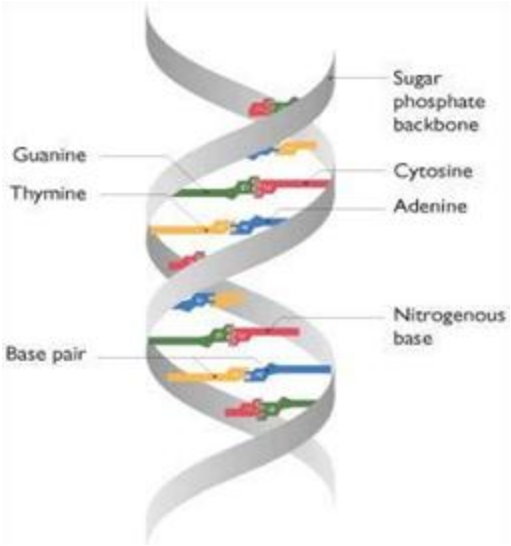

Figure 1. DNA Structure [10]

The two DNA strands are known as polynucleotides since they are composed of simpler units called nucleotides. Each nucleotide is composed of a nitrogen-containing nucleobase either guanine (G), adenine (A), thymine (T), or cytosine (C) as well as a monosaccharide sugar called deoxyribose and a phosphate group. According to base pairing rules (A with T and C with G), hydrogen bonds bind the nitrogenous bases of the two separate polynucleotide strands to make double-stranded DNA. [11]. The sequence of these bases determines the information available for building or forming an organism, similar to the way in which letters of the alphabet appear in a certain order to form words and sentences [8] as an example shown in fig.2,

DNA strand made of letters (DNA bases):

ATACTTGAATATATGTCAATTAGT

Letters make words (codons):

ATA CTT GAA TAT ATG TCA ATT AGT

Words make sentences (Genes):

ATA -CTT -GAA -TAT ATG -TCA -ATT –AGT

Figure 2. DNA sequence terminology

Vast majority of living organisms encode their genes in long strands of DNA (deoxyribonucleic acid). DNA consists of a chain made from four types of nucleotide subunits, each composed of: a five-carbon sugar (2'-deoxyribose), a phosphate group,

and one of the four bases adenine, cytosine, guanine, and thymine. The most common form of DNA in a cell is in a double helix structure, in which two individual DNA strands twist around each other in a right-handed spiral. In this structure, the base pairing rules specify that guanine pairs with cytosine and adenine pairs with thymine. The base pairing between guanine and cytosine forms three hydrogen bonds, whereas the base pairing between adenine and thymine forms two hydrogen bonds. The two strands in a double helix must therefore be complementary, that is, their bases must align such that the adenines of one strand are paired with the thymines of the other strand, and so on. DNA strands have chemical polarity of 5' and 3' at top and bottom, which binds two single stranded DNAs in anti -parallel way see fig.3. This complementary DNA structure was presented by Watson and Crick.

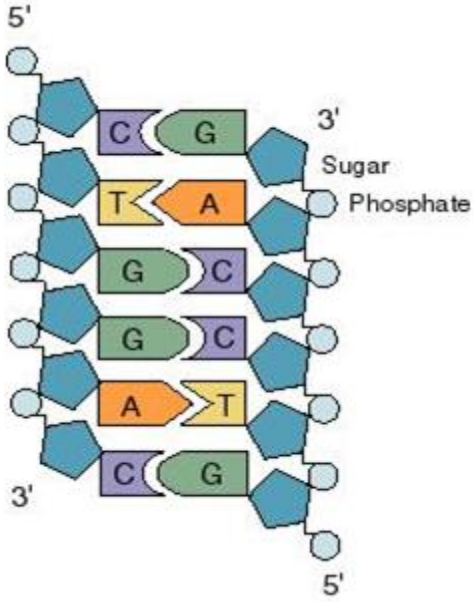

Figure 3. Nucleotide base pairing of strands [13]

DNA strands have directionality. One end of a DNA polymer contains an exposed hydroxyl group on the deoxyribose; this is known as the 3' end of the molecule. The other end contains an exposed phosphate group; this is the 5' end. The directionality of DNA is vitally important to many cellular processes, since double helices are necessarily directional [9].

III. DNA CRYPTOGRAPHY

The relation between cryptography and molecular biology was originally irrelevant, but with the in-depth study of modern biotechnology and DNA computing, begin to work together more closely. DNA cryptography and information science was born after research in the field of DNA computing field by Adleman. Many scholars from all over the world have done a large number of studies on DNA cryptography. In current scenario it is not much effective than traditional cryptography but it can provide a hybrid security by combining traditional cryptography with it [9]. For applying cryptography operations, we use the ASCII table to convert each of the individual letters into a numerical value, for example, (i=105, G=71, etc.), which can then be changed from base-10 to base-4 (105=1221, 71=0113, etc.). Finally, those numbers can be changed into their DNA base equivalents, with 0, 1, 2, and 3 replaced with A, T, C, and G. In the field of information science, the most basic encoding method is binary encoding. This is because everything can be encoded by the two states of 0 and 1. However, for DNA there are four basic units:

1. Adenine (A);
2. Thymine (T);
3. Cytosine (C);
4. Guanine (G).

The easiest way to encode is to represent these four units as four figures:

1. A(0) –00;
2. T(1) –01;
3. C(2)–10;
4. G(3)–11.

Once the raw data is ready, the researchers say a few algorithms can be used to weed out redundant and repetitive information. That doesn't just save a ton of space - lots of repetition in the DNA sequence can actually be biologically harmful to the wellbeing of the DNA and bacteria, so this step rather neatly solves two problems at once. DNA strands aren't long enough to store complicated information like a photograph or a book, so the best available solution is to fragment the data into lots of little pieces and spread it among the different cells. DNA cryptography is a subject of study about how to use DNA as an information carrier and it uses modern biotechnology as a measure to transfer ciphertext into plaintext. Thus, biotechnology plays an important role in the field of DNA cryptography For efficient use of DNA in computing and cryptography silicon chips can be replaced by DNA chips or bio-chips in future. Table I [15] shows a basic comparison between these two.

IV. RELATED WORKS

Research work is being done on DNA Computing either using test tubes (biologically) or simulating the operations of DNA using computers. An image encryption algorithm based on

DNA sequence addition operation is presented by Wang et. al. [10]. A DNA sequence matrix is obtained by encoding the original image and it is divided into some equal blocks and two logistic maps. DNA sequence matrix is decoded to get the encrypted image. Leier et. al. presented two different cryptographic approaches based on DNA binary strands with the idea that a potential interceptor cannot distinguish between dummies and message strand [7]. The first approach hid information in DNA binary strands and the second designed a molecular checksum. The YAEADNA algorithm proposed by Sherif et. al. uses a search technique in order to locate and return the position of quadruple DNA nucleotide sequence representing the binary octets of plain text characters[12]. Plain text character and a random binary file are given as input and the output PTR is a pointer to the location of the found quadruple DNA nucleotide sequence representing the binary octet. The encryption process was tested on images showing random the selection of DNA octet's locations on the encrypting sequence.

TABLE I. BASIC COMPARISON BETWEEN TRADITIONAL AND DNA CRYPTOGRAPHY

	Traditional cryptography	**DNA based cryptography**
Ideal System	Silicon chip based	DNA chip based
Information Storage	Silicon computer chips	DNA strands
Storage Capacity	1 gm silicon chip carries 16 Mega-bytes	1 gm DNA carries 10^8 Tera-bytes
Processing Time	Less	high
Performance Dependency	Implementation and system configuration	Environmental conditions

V. PROPOSED SYSTEM

We designed a DNA encryption technique based on matrix manipulations and using a key generation scheme which makes data much secure. Here text message is converted to ASCII code and placed in a 4*4 matrix. On this matrix, mathematical manipulation and scrambling is performed in cycles and XOR operation is performed with the initial key in each cycle to scramble data properly to make message non-readable. A secure key generation scheme is also used in this encryption system. Using generated key we XOR the result of matrix manipulation to generate mini cipher. The benefit of using this scheme is that it always generates different cipher text for same message text and even for same key. So it does not provide any clue or hint to make guesses about plain text.

DNA digital coding is performed on the mini cipher result to generate DNA nucleotide based codes which are in the form of A, C, T and G [14]. Primer pairs are used as keys to change the nucleotide sequence. Amino acid sequence is generated using DNA codes as a final cipher text. This sequence somewhat helps in hiding the existence of DNA coding usage from attacker. In this way this design provides a simple and secure system based on matrix computations. The use of three keys like initial key, generated key based on our scheme and third one is pair of primers makes encryption process much efficient. This new mechanism is based on the mixture of mathematical and biological operations and concepts.

A. Encryption DNA cryptography is based on the concepts of DNA computing. Actual DNA cryptography is far away from realization because in current time it can be performed in labs using chemical operations. DNA chips can be used in computers in place of silicon chips [15]. At that time it will be possible to use real DNA as a computational medium and also as data to process. In the current scenario we can implement a DNA encryption system which is based on the composition of mathematical computations and DNA concepts which results data in the form of biological DNA sequences. It makes difficult to read and guess about data.

In our proposed system there are two parts of our mechanism. First part belongs to mathematical manipulations of data matrix whereas second part belongs to DNA encryption process where DNA digital coding and DNA sequence modification is

performed to make data secure with the help of primer pairs.

In first part of encryption process, up to 128 bit data and 128 bit key can be processed. As shown in fig.4, in the first step original plaintext is converted to ASCII codes, after that data is placed in a 4 X 4 data matrix. Matrix manipulation operation is performed in cycles where first of all row shifting is performed as operated in AES algorithm [1]. First row of matrix is kept same, second row is shifted by one element, third row by two and fourth row is shifted by three elements. Second step of manipulation cycle is flipping operation where matrix is flipped from left to right. This operation can be called as mirror operation as it works like a mirror which displays a flipped view of any object. Third one is up to down flipping operation which performs a vertical flipping. In this way a matrix is scrambled in proper way.

In each cycle after these three steps a XOR operation is performed between the results of first three steps of cycle with the initial key. This cycle is based on the length of initial key. If the length of initial key is n then the number of cycles will be equivalent to 2*n. The output of matrix manipulation cycle is XORed with result of first part of our algorithm. Minicipher will be always different for same plaintext and same key because of our secure key generation scheme. This feature makes data safe because it does not give any hint due to its different outputs.

In the second part of the encryption as shown in fig.5, base- 4 conversion is performed which is a sequence of 0, 1, 2, 3. On this data reshaping operation is performed to modify the data sequence. DNA digital coding is performed on reshaped data where 0, 1, 2, 3 are replaced by DNA bases A, C, T, G. In this way we get a nucleotide sequence which is like a biological information sequence present in DNA. Biological DNA consists of genetic information in the form of these nucleotide bases. So in this step, our message takes a new face which relates to biological environment. On this nucleotide sequence several biological DNA operations can be performed to make data much complicated to read. Primer pairs are used as keys to change the sequence. In last step the result can be converted to amino acid sequence using biological tool.

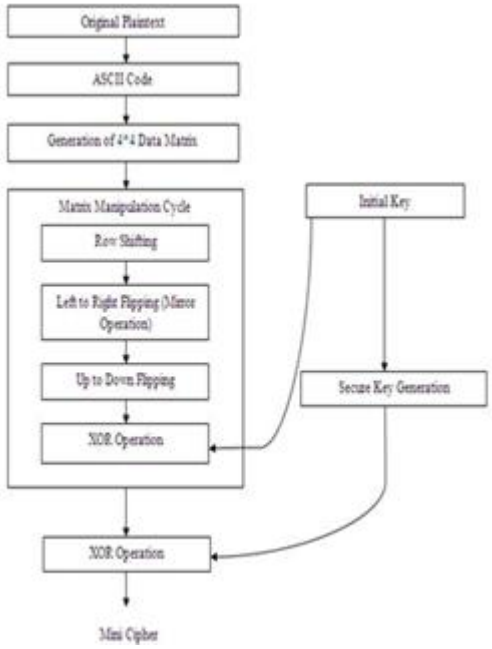

Figure 4. Mini cipher generation

the new key generated by the secure key generation scheme using initial key. This results in a minicipher, which is the

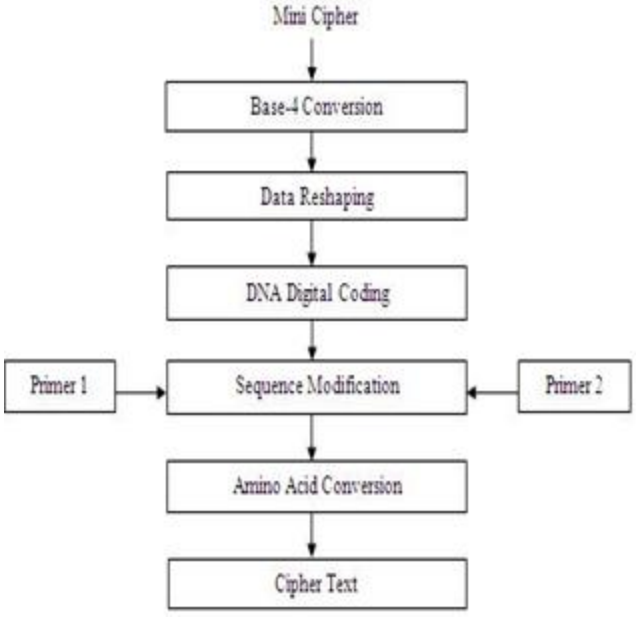

Figure 5. Cipher text generation

B. Secure key generation scheme

This scheme takes initial key as input and generates a new key. This newly generated key consists of a random number, set of remainders and set of quotients. In key generation process first of all a

random number is generated. This random number is divided by each character of initial key to generate remainders and quotients. As we know the ASCII range for characters is from 0 to 255, so random number should be greater than 255 and must be a prime number. The reason behind taking prime number is that, we can get remainders only when we use a prime number otherwise in some cases we may not get values of remainder if number is fully divided by ASCII values of initial key characters.

- Initial key ASCII Code.
- Random number generation (Prime Random number > 255).

- Remainders=Random number / ASCII value [1 to end].

- Quotients= Random number/ASCII value [1 to end].

Generated key= {Random number, Remainders, Quotients}

We described an example which shows how the values are calculated for key, suppose initial key is "DEMO" and random number generated is "331" as shown in Table II. Based on calculations the newly generated key will be as follows,

Generated Key= {331, 59, 55, 23, 15, 4, 4, 4, 4}

Only the generated key is shared with the receiver not the initial key because only the person knowing this scheme can generate the initial key for message recovery at the receiver end, which makes encryption process safe and secure. As we use random number to generate values, each time different key is generated even if the same initial key used. The remainder part of generated key is used for XOR operation with the result occurred by matrix manipulation cycle. But whole key is required to generate initial key. To get initial key from generated key in decryption process we simply apply the operation on key as given below,

- Initial key= (Random number – Remainder [1 to end]) / Quotient [1 to end].

C. Decryption

Message recovery of original plaintext is a reverse process of encryption.
Two ke

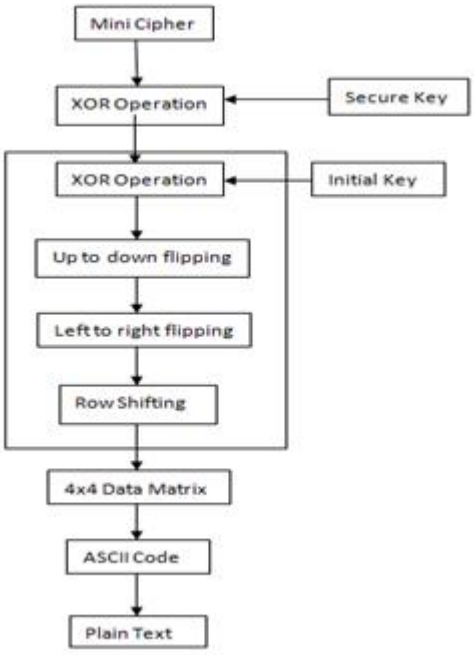

Figure 6. Plain Text Generation

Two keys are shared with the receiver; first key is generated by the secure key generation scheme and second is the primer key pair. In first step amino acid sequence is converted to DNA nucleotide sequence then primer key pairs are used to get encrypted original nucleotide sequence. After that DNA digital coding is applied to replace A, C, T and G with 0, 1, 2, 3. This generated sequence is reshaped in appropriate form so we get a base-4 data sequence which is a minicipher. Now minicipher is XORed with the secure key and then reverse matrix manipulation cycles are performed. Initial key is used in each cycle to perform XOR operation with results got in each cycle. Using first key receiver can generate initial key to apply further decryption process by the use of receiver side key generation scheme.

VI. RESULTS

To implement our encryption technique we used Matlab language which is a matrix based language and it is much suitable to perform our designed mechanism. Matlab also provides bioinformatics toolbox and methods [18] for biological computing. As shown in the fig.6, we used plaintext "SECRET MESSAGE" and initial key "AX085769*12" and performed encryption. Initial key generates a new key using key generation algorithm.

ISSN 2321-6905

TABLE II. KEY GENERATION

Character	ASCII	Division	Remainder	Quotient
D	68	331/68	59	4
E	69	331/69	55	4
M	77	331/77	23	4
O	79	331/79	15	4

We get mini-cipher after applying matrix manipulations. This mini-cipher is processed further by applying DNA encryption nucleotide sequence and applying amino acid conversion final cipher text is occurred. We also recovered original message successfully in decryption process.

Figure 7. DNA encryption results

VII. CONCLUSION

In this paper, we presented a new DNA encryption technique which is based on mathematical matrix manipulation where we used a secure generation algorithm to generate new key for encryption process. The benefit of this key generation scheme is that we always get a new cipher data for same plaintext and same key. So it provides a good security layer which does not give any hint about plaintext. DNA binary strands support feasibility and applicability of DNA-based Cryptography. The security and the performance of the DNA based cryptographic algorithms are satisfactory for multi-level security applications of today's network. Certain DNA algorithms can resist exhaustive attack, statistical attack and differential attack.The field of DNA computing is still in its infancy and the applications for this technology have not yet been fully understood. DNA computing is viable and DNA authentication methods have shown great promise in the marketplace of today and it is hoped that its applications will continue to expand. DNA Cipher is the beneficial supplement to the existing mathematical cipher. If the molecular word can be controlled at will, it may be possible to achieve vastly better performance for information storage and security.

REFERENCES

[1] Atul Kahate, Cryptography and network security (New Delhi: Tata McGraw Hill, 2012).

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

[3] R. J. Lipton, "Using DNA to Solve NP-Complete problems", Science, vol. 268, 1995, 542-545.

[4] Boneh, C. Dunworth, and R. Lipton, "Breaking DES using a molecular computer", Proceedings of DIMACS workshop on DNA computing, 1995, 37–65.

[5] Taylor Clelland, "Hiding messages in DNA Microdots", Nature Magazine vol.399, June 1999.

[6] Gehani, T. LaBean, and J. Reif, "DNA-Based Cryptography", Lecture Notes in Computer Science, Springer, 2004.

[7] G. Cui, L. Qin , Y. Wang , X. Zhang, "An Encryption Scheme Using DNA Technology", IEEE, 2008.

[8] Genetic home reference, a service of the U.S. National Library of Medicine, http://ghr.nlm.nih.gov/ handbook/basics/dna, 2012.

[9] S. Jeevidha, Dr. M. S. Saleem Basha and Dr. P. Dhavachelvan, "Analysis on DNA based Cryptography to Secure Data Transmission", IJCA, Volume 29– No.8, September 2011.

[10] DNA Structure, http://ijarovic.wordpress com, 2012.

[11] Monica Borda and Olga Tornea, "DNA secret writing Techniques", IEEE conference, 2010.

[12] Learn Genetics, University of Utah, http:// learn.genetics.utah.edu/content/begin/tour, 2012.

[13] Nucleotide base pairing of strands, http://dedunn.edublogs.org, 2012.

[14] D.Prabhu and M.Adimoolam, Bi-serial DNA Encryption Algorithm (BDEA), Cornell university library, http://arxiv.org/abs/1101.2577, 2011.

[15] B. Anam, K. Sakib, Md. A. Hossain, K. Dahal, Review on the Advancements of DNA cryptography, 2010.

[16] M. Borda , O. Tornea and T. Hodorogea, "secret writing by DNA hybridization", acta technica napocensis Electronics and Telecommunications, Volume 50, Number 2, 2009.

[17] Pankaj Rakheja, "Integrating DNA Computing in International Data Encryption Algorithm", IJCA, Volume 26– No.3, July 2011.