Secure Data communication and Cryptography based on DNA based Message Encoding

Snehal Javheri
Sinhgad Institute of Technology,
Department of Computer Engineering,
Lonavala Pune, University of Pune, India

Rahul Kulkarni
Sinhgad Institute of Technology,
Department of Computer Engineering,
Lonavala Pune, University of Pune, India

ABSTRACT
Information flows throughout the network that may be of local or of global scope. It is mandatory to secure that information to prevent from unauthorized access of it by any node in the path. There are various users and organizations who want to prevent their crucial data from attackers and hackers. Also we need to ensure privacy, integrity and confidentiality about data in the network for it to be a reliable. Thus to achieve security it is very necessary to encode the data before sending it through the various unreliable communication channels available to make it unreadable. This is where the Cryptography comes into picture. Various cryptographic systems were developed in the past year but now the latest development on this field is DNA Cryptography. This concept has emerged after the disclosure of computational ability of Deoxyribo Nucleic Acid (DNA). In this field of DNA Cryptography many research work is going on to make the computational process more complex to the unauthorized user. Well, presently it is in the development phase and requires a lot of work and research to reach an established stage. In this paper; a proposal is given where the concept of DNA is being used in encryption and decryption process. The theoretical analysis shows this method to be efficient in computation, storage and transmission; and it is very powerful in certain attacks. This paper also proposes a secured symmetric key generation scheme which generates primary cipher and this primary cipher is then converted into final cipher using DNA sequences, so as to make it again more complicated in reading. Finally, the implementation methodology and experimental results are presented. And then conclusion with future work is described in the last section.

Keywords
Security, Encryption; Decryption; Key generation; Cipher text; DNA cryptography.

1. INTRODUCTION
The security to a system is essential nowadays ! with the growth of the Information Technology and with the emergence of new techniques, the number of threats a user is supposed to deal with grew exponentially. It doesn't matter if we talk about bank accounts, social security numbers or a simple telephone call. It is important that the information is known only by the intended persons, usually the sender and the receiver. This is where the cryptography comes into picture. Cryptography is the basis of security of all the information.

2. BACKGROUND

2.1 Cryptography
Cryptography is the art and science of achieving security by encoding the simple message to make it unreadable [3] [20]. The typical scenario in cryptographic as shown in Fig. 1 is that Bob (sender) wants to send some messages secretly to Alice (intended receiver) and the Eve is the third person who is trying to read the message but could not succeed, as the message is secured to some extent, by using cryptographic algorithm. The message to send is in simple or ordinary language understood by all, it is called a plaintext. The process of converting plaintext into a form which cannot be understood without having special information is called encryption. This unreadable form is called cipher text and this special information is called encryption key. The conversion of cipher text again into plaintext with a special knowledge is called decryption, whereas special knowledge for decryption is called decryption key. Only the receiver has this special knowledge and only receiver can decrypt a cipher text with this knowledge called decryption key.

There are basically two types of cryptography based on the techniques for converting plaintext to cipher and vice versa which are namely called as symmetric and asymmetric cryptography. In symmetric cryptography sender and receiver use the same key for encryption and decryption of test whereas in asymmetric cryptography systems two keys namely public and private keys are used for encryption and decryption process. By keeping the private key safe, you can assure that the data remain safe. But the disadvantage of asymmetric algorithm is that they are computationally intensive. Therefore, in this proposed system symmetric key cryptography is used with the intension of less computation but high data security.

Cryptography mechanisms are depending on the degree of randomness and uncertainty in the generation of the cipher text from the plain text. Hence depend on the phenomenon of nature there are various types of cryptography such as: Modern Cryptography is based on the difficult mathematical problems such as prime factorization, matrix manipulation. Elliptical Cryptography, make use of elliptical curve problems. Quantum Cryptography uses the randomness of states of electron inside an atom. Moreover DNA Cryptography depends on the difficult biological process concerning to the field of DNA technology [13] [21].

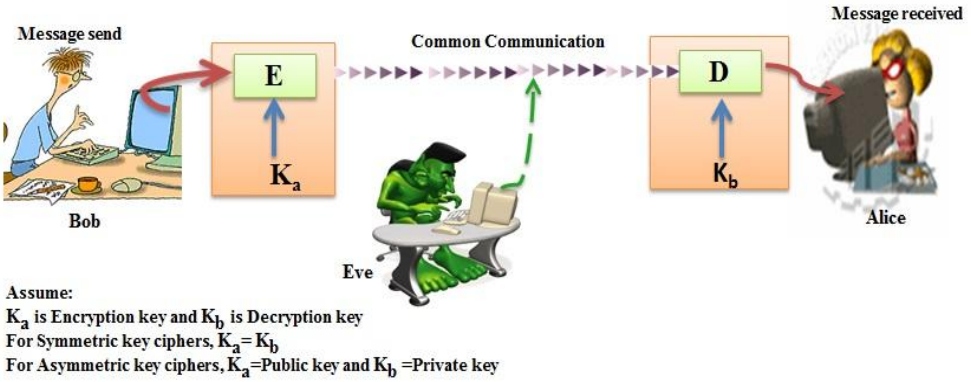


Fig 1: Cryptographic Scenario

DNA computing and cryptography came into picture in 1990[15]. DNA computing was initiated by L. Adleman [14]-[16] in 1990, where he solved a direct Hamiltonian path problem and set the foundation of the research in the field of Bio Computing. In the field of cryptography, Ashish Gehani et al introduces the first algorithm of DNA based cryptography [1] [2] [11].

3. DNA BIO-LOGICAL THEOREM

Deoxyribo Nucleic Acid (DNA) is a biological material or bio-molecule which is present in almost all living things [21]. DNA is located in the cell nucleus but a small amount of DNA may also be found in the mitochondria. DNA contains genetic information of a living thing. It contains genetic instructions which help in constructing other cells [12]. In 1953, James Watson discovered the structure of DNA. A DNA molecule is composed of two single strands which form a double helix structure as shown in Fig. 2:

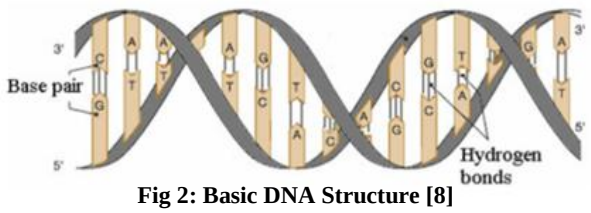


Fig 2: Basic DNA Structure [8]

The DNA sequences consists only four alphabets: Adenine (A) Cytosine (C) Guanine (G) Thymine (T). Each alphabet is related to a nucleotide. Watson-Crick proposed a complimentary rule for DNA sequences that: "A only joint with T through double bound (A = T)" and C only joint with G through triple bound (C ≡ G)" [12]. DNA provides the major support of genetic information for all kind of organism in biosphere. It composed of two long strands of nucleotides, a deoxyribose sugar and a phosphate group. DNA sequences are responsible for transfer of complex information. 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 [12]. Thus, DNA provide information to message RNA (mRNA) through the process transcription, then mRNA transfer the information to Protein, this process is known as translation. These two processes play an important role for information transfer from one age group to another age group.

4. DNA CRYPTOGRAPHY AND RELATED WORKS

DNA cryptography is based on DNA computing where, message is encrypted in the form of DNA nucleotide sequence. DNA computing can be used as conceptual platform for data encryption and decryption by using symmetric or asymmetric key. In current scenario it is not much effective than traditional cryptography but it can provide a hybrid security by combining traditional cryptography with it [21]. However DNA logic can be implemented with traditional cryptography [13]-[21].The ultimate target is to scramble data in the way that the person who doesn't know the key, can't read or modify data.

In the recent year few works on qualitative and quantitative analysis on DNA based Cryptography as well as many new Cryptographic techniques are proposed by the researchers. Bibhash Roy et al [5] [6] [7] proposed a DNA sequencing based encryption and decryption process. This paper also proposes a unique cipher text generation procedure as well as a new key generation procedure. But the experimental result shows that the encryption process requires high time complexity. Tushar Mandge et al [21] designed a DNA encryption technique based on 4*4 matrix manipulations and using a key generation scheme which makes data much secure. Miki Hirabayashi and Akio Nishikawa [18] have proposed theoretical and empirical based analysis on application of DNA cryptography. Pankaj Rakheja [19] designed a new method by integrating DNA computing in IDEA. Such conceptual works can be useful in the development of this new born technology of cryptography to fulfill the future security requirements.

5. PROPOSED METHOD

In actual scenario, DNA cryptography is far away from realization because in current time it can be performed only in labs using chemical operations. In order to provide better security and reliable data transmission an effective method of DNA based cryptography is proposed here. In this method the mixture of mathematical and biological concepts are used to get the encrypted data in the form DNA sequences. The benefit of this scheme is that it makes difficult to read and guess about data (plain text).The proposed algorithm has two phases in consequence: these are Primary Cipher text generation using substitution method followed by Final Cipher text generation using DNA digital coding.

In the Primary Cipher text generation phase, the encryption algorithm uses OTP (one-time-pad) key generation scheme, since almost one key for one piece of information is sufficient to provide lots of strength in encoding technique. The proposed method uses randomly generated symmetric key of 8 bits size by the intended receiver and provided to the sender. Thus the sender will have a partial knowledge of the private key only and then it generates the rest part of the keys (Private Keys: Level 1 and 2) to encode the information. The Byte values are extracted from the input file or message. The further encryption process works on unsigned byte values of the input file or text called as plain text. These byte values are replaced by combination of alphabets and special symbols using substitution method. And then this substitution values are converted into its binary value. In order to embed more security extra bits are padded at both the ends of the primary cipher text. These extra bits are nothing but the file size information, which is provided to the receiver through Level 2 key. Thus the secret key, the information of primer pairs are shared between sender and receiver through the secret channel.

In the DNA digital coding phase, the Final Cipher text is generated from Primary Cipher text using DNA digital encoding technique. From a computational point of view, we cannot process the DNA molecules as in form of alphabets, so the DNA sequence encoding is used in this method through which the binary data is converted into DNA format and its vice versa. The four subunits of DNA molecule called as nucleotide bases: A: adenine; G: Guanine; C: Cytosine and T: Thymine are converted into 2 bit binary as A: 0(00), T: 1(01), C: 2(10), G: 3(11). Obviously, there are 4! =24 possible coding patterns by this encoding format. However, according to the Watson-Crick complementarily rule, in double helix DNA structure, the two DNA strands are held together complementary in terms of sequence, i.e. A to T and C to G.

Taking DNA digital coding into account, it is required to reflect the biological characteristics of 4 nucleotides DNA bases, the complementary rule that (~0=1) and (~1=0) is proposed in [17]. As per the rules, 3(11) is complement of 0(00) and 2(10) of 1(01). So among 24 patterns, only 8 kinds of patterns (0123/CTAG, 0123/CATG, 0123/GTAC, 0123/GATC, 0123/TCGA, 0123/TGCA, 0123/ACGT AND 0123/AGCT) are fit as per complimentary rule of the nucleotide bases. It is suggested that the coding pattern 0123/CTAG is the best for nucleotide bases [17]. Thus A and T are corresponds to '00' and '11'respectively and C and G to '01' and '10' respectively. So substitution rule is A=00, T=11, C=01 and G=10 as illustrate in Table 1.

Table 1. DNA Digital Coding

DNA nucleotide	*Decimal*	*Binary*
A	0	00
C	1	01
G	2	10
T	3	11

6. ALGORITHMIC PRESENTATION

The encryption scheme proposed in this paper has following phases:

Abbreviations used are: PT- Plain Text, PK: Public Key, PK1- Private key Level 1, PK2- Private Key Level 2 and SPM- Starting Primer; EPM-Ending Primer, PCT- Primary Cipher text, FC-Final Cipher text.

6.1 Format of Cipher Text

From plain text (PT) the Primary cipher text (PCT) is obtained by using the encryption algorithm and Level-1 Private key (PK1).

STEP TO OBTAIN THE FINAL CIPHER TEXT

Begin

Step 1: Encrypt the plain text with Private key Level 1(PK1).

Step 2: Calculate Private Key Level 2(PK2) and attach the SPM (Starting Primers) and EPM (Ending Primers) to the Primary cipher text (PCT) so obtained from Step 1.

Step 3: Apply DNA digital encoding technique to get Final cipher text (FCT). (Fig. 3)

End

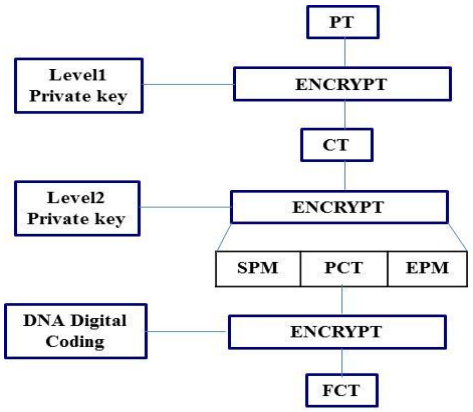


Fig 3: Final Cipher Format

6.2 Procedure for Level1 Private Key Generation at both sides

Procedure: *Sender's side Computation*

Begin

Step 1: First the receiver will send a number as public key (PK) through channel (private or public). This key should be any positive number between the ranges 1 to 255.

Step 2: Sender will generate one random number (R).

Step 3: The random number selected is being represented in binary and then its complement is being again converted into decimal which will be used as the Encryption key (E).
(E.g. Let Public Key is PK=7, and the random number R=5.Binary representation of R=0101(4-bit).Complement of R =0010. Therefore, In Decimal R= 2. This 2 will be used as Encryption Key (EK=2)).

Step 4: The sender will compute the level-1 private key (PK1) as follows:
Remainder computation (r): (PK * R) % 16= (7*5) % 16 =3 (Remainder) Hexadecimal Notation = 3.
Quotient computation (c): (PK * R) / 16 = (7*5) / 16 =2, Hexadecimal Notation = 2.
Concatenating these two hexadecimal notations, we get *rc* = 32.

Step 5: Sender will send *rc* as level1 private key (PK1) with level2 private keys (PK2) through private Channel in a progress.

End

Procedure: *Receiver's side Computation*

Begin

Step 1: Receiver will receive *rc* = 32 and separate the numbers *r* and *c* and convert into decimal notation.

Step 2: Receiver will compute the decryption key as follows:
Decimal value computation (X): X= (16 * c) + r (e.g. X= (16 * 2) + 3 = 35)
Intermediate key computation (K1): K1=*(*X / PK) (e.g. K1= (35 / 7) = 5, where PK=7)

Step 3: Convert to binary form and complement it. E.g. Binary of 5=101. In Decimal Notation = 010 =2.

Step 4: Therefore, 2 is Level-1the Private Key (PK1) to be used for decryption.

End

6.3 Procedure for Level2 Private Key Generation

Procedure: *Sender's side Computation*

The Level-2 private key gives the information about the length of the primer. The primers are added at the start and end of the primary cipher text (PCT). The sum of the digits of the sender's file length is taken as the input (PK2) to decide the primer's length. To encoded this primers information use following procedure:

Begin

Step 1: Let P be an array which will hold the secondary level of keys.

Step 2: Take a variable and initialize it with a number which is the file length.

Step 3: Repeat through the following steps for 1 to number of digits in N.

Step 4: Perform digit wise X – OR of N from left to right (i.e. from MSB to LSB) as shown in Fig. 3.

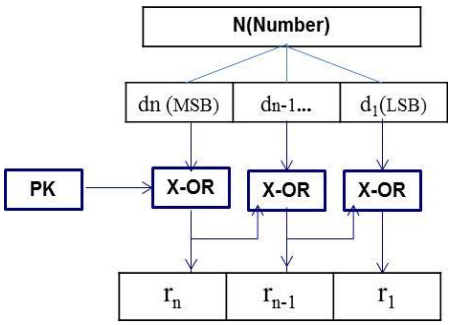


Fig 3: XOR Operation of the PK

Step 5: N = r_n-1 r_n – 2…….r_1.

Step 6: P [i++] = r_n.

Send the array P (level2 private key) to the receiver to get the file length by applying reverse procedure of the above *(Note: The sender will send both Level-1 and Level-2 private keys to the receiver in a digest form).*

6.4 Procedure for Encryption

Procedure: *Sender's side Computation*

Begin

Step 1: Let, Q is an array of size 16 and R is an array of size 16 also. For example, Q[16]={'A','B','C', 'D','E','F','G','H','I','J','K','L','M','N','O','P'}
R [16] = {'!', '#', '$' ,'%' ,'&' ,'(',')' ,'+' , ',' , '-' ,'.' ,';' ,'=' ,'@' ,'[' ,']' };

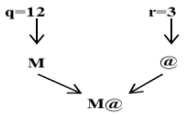
Step 2: Input the file as plain text and Convert the file into its byte codes (the ranges from -128 to +127).

Step 3: In order to get the index of the arrays we have to change the negative value byte codes into positive values by adding +128 to each of the byte values. For example,-120 becomes +8.Thus the range of the byte codes becomes 0 to 255.

Step 4: Each of the byte will be taken in account of calculation as: n1= (byte code / 16) and n2= (byte_code%16).For example, if the byte code=92 then n1= 92 / 16 =5 and n2= 92%16= 12 Therefore, n1=5 and n2=12
(Note: We are using 16 in the calculation as 16 is the size of the array. So as there are two arrays, the range of the byte codes will be 16 x *16= 256.)*

Step 5: Now the key will be added up with the numbers n1 & n2 to get the new indexes q and r as: q= [(n1+k1) %16] and r= [(n2+ k1) % 16], where k1 is the key value. For example, Let the key is k1=7. So, q = (n1+7) %16 = (5+7) %16=12%16=12 and r = (n2+7) %16= (12+7) %16=19%16=3.

Step 6: The numbers q and r will be used as the index of the static arrays Q & R. E.g. here the value of q =12 and r=3. We get:



Thus the byte code '92' is converted into *'M@'*

Step 7: After getting the cipher of each byte code, concatenate all byte codes in order to get the primary cipher text (PCT), which is the result of first part of our algorithm.*(Note: Primary cipher will always be different for same plaintext because of our secure random key generation scheme. This feature makes data safe because it does not give hint due to its OTP nature.)*

Step 8: Now convert the byte codes obtained into binary bits. Use Primer pairs to change primary cipher sequence and then DNA digital coding is performed on reshaped data to get final cipher text in DNA sequences.

End

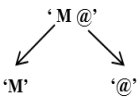
6.5 Procedure for Decryption

Procedure: *Receiver's side Computation*

Begin

Step 1: Covert Final Cipher into Primary Cipher Text First.

Step 2: Read the input cipher file, two bytes at a time. Split the code as shown below:



Step 3: Search the array Q to get the index of '*M*' and array R to get the index of '@'. Therefore we will get the index of 'M'=12 and the index of '@'=3.

Step 4: Subtract the key from each of the index value. If the result becomes negative, add 16 with it.
For example, n1=12–7=5 and n2= 3–7=-4. Since n2<0, so add 16 with it. So now, n2= -4+16= 12.

Step 5: Now multiply n1 by 16 and add n2 with it to get the byte code. So Byte_code = (n1*16) + n2.
e.g. Byte_code = (5*16) +12 = 92.

Step 6: Convert the byte code '92' into its corresponding ASCII code.

End

7. PERFORMANCE ANALYSIS

7.1 Key Strength Analysis

The key generation scheme uses two different keys at different levels of computations, where- *Key1* ranges from 0 to 255 i.e. requires 8 bits to represent also *Key2* is used for sending Primer pair information, ranges from 0 to 255 i.e. requires 8 bits. So in total the key size is 16 bit, that is there are 2^16 =65536 possible keys. Now assume that a hacker have a very fast computer using which our decryption algorithm can be executed in 1 micro second for all possible key trials. Even if he tries half the set of keys then also he is quite successful in decrypting. But then also the hackers require more than half year for decrypting the cipher text as shown: *Assume,*

In one second = 10 possible key trials
In one hour = 36 x 10^3 =36000 possible key trials
In one day = 864 x 36 x 10^3=3.1104 x 10^9 possible key trials.

As after one day all the possible keys are tried, but our model works so effectively that it would have already transmitted the information and the receiver had read it earlier to the intrusion attack.

7.2 Result Analysis

The proposed procedures are implemented in Java platform for its platform independence property and available in-built cryptography functionalities. The procedures are implemented successfully to specified sized input plain texts. Initially, there were constraints for large files such as image or videos where the required primary memory of the system could create a problem in the execution and conversion of plain text into cipher text. But later on that problem are also resolved by extracting the byte code values of the compressed file of input plaintext and then dividing the compressed file into fixed sized chunks and then performing all the procedures of encoding and decoding by joining the chunked sub files. This algorithm is applicable for all most all documents.

The results of generation of cipher text after encrypting the plaintext on few data sets are given below:

Table 2. Datasets for performance analysis

File	File size(Byte)	Cipher size(Byte)	Encrypt time(ms)	Decrypt time(ms)
Test1	49	90	44	45
Test2	1,443	2,878	455	370
Test3	2,454	4,892	785	721
Test4	3,960	11,860	4,096	1522

8. CONCLUSION

The proposed method of encoding is far better and faster than conventional cryptography like DES and other DNA based encryption algorithms. As DNA computing is a very promising field that keeps the ability to overcome many limitations of silicon computers, the proposal can surely be enhanced with much more advanced concept such as realization in several security technologies of encryption, steganography, signature and authentication by using DNA molecular as information medium. The strength of the proposed method is the complex cipher generation from two strong keys. Although, this method is efficient and powerful against certain attacks; the partial information contained in the cipher text makes the method much stronger. The conversion of byte values into the combination of alphabets and special symbols makes the cipher text more complex, but a percentage of increase in cipher size will be there. However several extension and variation can be made to enhance this technique, so that it can be used for secured data communication in any type of network. It is truly mentioned that the DNA based cryptosystem cannot totally be replaced by traditional cryptosystem which is currently being used. This field requires a lot of research and work, to have a position in which it can be implemented and used for practical purposes. There is a need that people from traditional cryptography and DNA technology should exchange knowledge among each other and cryptosystems should be devised in such a way that they can enjoy benefits from both the fields.

9. FUTURE SCOPE

In the recent year, the most challenging field of information security is wireless sensor networks. The light weight computational sensor nodes used in wireless network leads to many security applications. The proposal can be further enhanced to include in security mechanism of wireless networks and analyzing its performance to basic cryptanalytic attacks and comparing it with existing cryptosystems to know exactly how much improvement is achieved.

10. ACKNOWLEDGEMENT

All compliments to the almighty God who gives us the ability and knowledge to do this hard work. Greatest gratitude and profound respect to Mr. Rahul R. Kulkarni, Assistant Professor, Dept. of Computer Engineering and Dr. Mrs. V. M. Rohakale, Dept. of E&TC ,SIT Lonavala Pune, University of Pune, for their proper and perfect guidance and valuable suggestions.

11. REFERENCES

[1] Ashish Gehani, T. LaBean and J. Reif, "*DNA-based cryptography"*, DIMACS DNA Based Computers V, American Mathematical Society, 2000.

[2] Ashish Gehani et al , "*DNA-based cryptography"*, Lecture Notes in Computer Science, vol.2950, pp.167-188, 2004.

[3] Atul Kahate, *"Computer and Network Security",* Third Edition,Tata McGraw Hill Publication Company Limited,2013.

[4] Beenish Anam,Kazi Sakib,Md. Alamgir Hossain,Keshav Dahal, *"Review on the Advancements of DNA Cryptography"* arXiv:1010.0186v[cs.CR], 1st Oct 2010.

[5] Bibhash Roy, Gautam Rakshit, Pratim Singha, Atanu Majumder, Debabrata Datta,"*An improved Symmetric key cryptography with DNA Based strong cipher*"-ICDeCom-2011, Feb' 24-25'2011, pp.1-5.

[6] Bibhash Roy et al, "*A DNA based Symmetric key Cryptography*"-ICSSA- 2011, 24-25 Jan'11.

[7] Bibhash Roy, Gautam Rakshit, Pratim Singha, Atanu Majumder, Debabrata Datta,"*An Enhanced key Generation Scheme based cryptography with DNA Logic*"-IJICT-2010-11, Volume 1 No. 8, Dec' 2011.

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

[9] Er. Ranu Soni Er. Vishakha Soni,Sandeep Kumar Mathariya, "*Innovative field of cryptography: DNA cryptography*",- DOI:10.5121/csit.2012.2115,CSCP-2012.

[10] G. Cui, L. Qin, Y. Wang, and X. Zhang, "*An encryption scheme using DNA technology,"* in IEEE 3rd International conference on Bio-Inspired Computing: Theories and Applications (BICTA08), Adelaid, SA, Australia, 2008, pp. 37–42.

[11] Gehani Ashish, La Bean, Thomas H. Reif, John H,"*DNA-Based Cryptography*",Department of Computer Science, Duke University, June 1999.

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

[13] Guangzhao Cui Limin Qin Yanfeng Wang Xuncai Zhang. *"An encryption scheme using DNA technology."* Bio-Inspired Computing: Theories and Applications, 2008. BICTA 2008. 3rd International Conference on Publication Date: Sept. 28 2008-Oct. 1 2008 ISBN: 978-1-4244-2724-6, page(s):37-42; Adelaide, SA.

[14] Guangzhao Cui Limin Qin, Yanfeng Wang, Xuncai Zhang, "*An Encryption Scheme Using DNA Technology*", IEEE, 978-1-4244-2724-6/08, 2008.

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

[16] L. Eschenauer and V. D. Gligor. *"A key-management scheme for distributed sensor networks."* Proceedings of the 9th ACM conference on Computer and communications security, Washington, DC, USA, pp. 41– 47, November 18-22 2002.

[17] Leroy Hood and David Galas, "The digital code of DNA", vol 421, no. 6921, 2003.

[18] Miki Hirabayashi, Akio Nishikawa, *"Analysis on Secure and Effective Applications of a DNA-Based Cryptosystem"*, IEEE computer Society, 978-0-7695-4514-1/11, 2011.

[19] Nucleotide base pairing of strands, http://dedunn.edblogs.org, 2012.

[20] P Pankaj Rakheja,*"Integrating DNA Computing in International Data Encryption Algorithm"*, IJCA, Volume 26-No.3, 2011.

[21] Tushar Mandge, Vijay Choudhary, *"A DNA Encryption Technique Based on Matrix Manipulation and Secure key Generation Scheme"*, ICICES Journal, 2013.