DOI: http://dx.doi.org/10.26483/ijarcs.v8i9.5006

ISSN No. 0976-5697

Volume 8, No. 9, November-December 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

ENHANCING INFORMATION ENCRYPTION WITH BIOMOLECULAR SEQUENCES USING NDES ALGORITHM

Menaka. K
Department of Computer Science
Shrimati Indira Gandhi College
Tiruchirappalli, India

Sundaravalli. V
Department of Computer Science
Shrimati Indira Gandhi College
Tiruchirappalli, India

Abstract: During Communication, Information Encryption plays a very important role. While transmitting data, unusual kinds of attacks may occur. To enhance the security during data transmission, cryptographic techniques are used. Cryptography is thus a study in which the message to be transmitted is converted into unreadable form and then sent so that unauthorized users may not be able to read the information. The information is transformed during a phase called Encryption, before being stored or transmitted, based on a Secret key. To guarantee the privacy and validity of the furtive data, a variety of cryptographic techniques have been developed and researchers are continuously working on it to afford better technique towards information security. Though many algorithms have been developed for hiding the data, biomolecular DNA (Deoxyribonucleic acid) sequences based data encryption seems to be a promising approach for satisfying the current information security needs. In this work, the enormous number of features available in DNA sequences are taken and combined with the well known cryptographic algorithm, the Data Encryption Standard (DES). The proposed technique which combines the features of DNA sequences and the Data Encryption Standard is thus named as NDES (Novel Data Encryption Standard) algorithm. This technique thus enhances the security by conversion, manipulation, substitution, confusion and hence provides complexity.

Keywords: DNA Sequence, Information Security. Encryption, Biomolecular Sequences, DNA Cryptography, Data Encryption Standards, Substitution Technique.

1. INTRODUCTION

Nowadays, information has become a very important resource and so the task of information security has become increasingly important. Cryptography is the most important factor of communication security and computer security. Information security is based on three essential factors (confidentiality, integrity, authenticity). Cryptographic techniques are helpful to hide some information in such a way which cannot be read by public groups. There are many cryptographic techniques available for securing data during transmission. Some of them used mathematical concepts and some have used the concepts of Physics [1]. As a standard with high information density, DNA was proposed for computational purposes by Adleman, 1994 [2].

Encryption using biomolecular DNA sequences is a useful tool in protecting confidentiality and integrity of information. The original meaning of the information is thus modified to prevent access from eavesdroppers. DNA supported bimolecular cryptography method is a practice that uses the substantial parallel processing way of biomolecular computation that converts short messages from hexadecimal and ASCII forms and performs **encrypt ion** and decryption process from the information. The exclusive property of DNA encoding is used for computations, which improves the security and encryption and lessens the flaws of the current security mechanism.

Forming protein from DNA sequences seems to be a challenging process which enhances the sequences and also maintains the integrity and validity of the process. This process is complex and difficult, and hence this gives the DNA based cryptography an advantage over other public key based cryptography methods. We consider this into account to go through modern security schemes and encryption algorithms to propose a novel algorithm to enhance the security and complexity of an encryption system.

2. BIOLOGICAL FRAMEWORK

In human body each cell contains a nucleus which characterizes all the physical and behavioral features of human body. They are packed into chromosomes. DNA is a molecule, within each organism. James Watson and Francis Crick formed primary 3D structure of DNA based upon an X-Ray print. They found out that DNA structure is double helix/ stranded [3] like a spiral ladder. It is made up of two strands where each strand can have either a purine or a pyramidine base. Adenine (A) and Guanine (G) are Purine bases while Thymine (T) and Cytosine (C) are Pyridimine bases. A sequence of DNA base pairs can be represented as a string made of these four characters i.e. <AAGTCGATCGATCATCGA>.

In a DNA sequence, every three adjacent nucleotide bases forms a codon which maps to a unique amino acid that is used in protein synthesis. When Adleman [2] started to inspect in molecular biology and found that these four characters (A, T, C and G) keeps the entire information needed by an organism and can be employed for computation, computing in DNA established to commence. DNA computing, in the factual sense, is the use of DNA molecules which encode genetic information for all living things, in computers. This is accomplished in a suspended solution of DNA, where certain combinations of DNA molecules are interpreted as a particular result to a computational problem encoded in the original molecules present. DNA computing is presently one of the fastest emerging fields in both Computer Science and Biology and

its future looks tremendously promising. A highly interdisciplinary study incorporating the research results of computer scientists and biologists

2. DNA CRYPTOGRAPHY

DNA cryptography is a division of biological science, which has large data storage capacity. Cryptography when combined with molecular biology, gives more secure data transmission and data hiding. DNA cryptography skill is needed in information security to guard and hide data. In conventional cryptography methods, encrypted messages can be detectable by attacker. DNA has ability to store enormous information rather than existing algorithm.

DNA cryptography is an emerging field in which a variety of researches are going on in order to make strong cryptographic techniques. Concept of DNA cryptography can be realized with either DNA computing or conventional cryptographic approach. The enormous parallelism and eccentric information capacity available in DNA molecules are used for cryptographic principles such as encryption, authentication, and signature. DNA cryptography based on conventional cryptographic approach consists of key generation, encryption and decryption process [4]. But the differences exist in it from conventional one is we use key sequences in a DNA format like ATCGCCAG. Ciphertext produced during encryption process by converting plaintext is also in the DNA form and decryption process converts the DNA cipher into its original plaintext. DNA cryptography is based on both symmetric and asymmetric key cryptography.

The conventional cryptographic algorithms do not provide much security when compared with DNA cryptography. To make data more secure against certain attacks while transmitting through networks, DNA based encryption strategy and some new methods are adopted. The message or plaintext will process through the algorithm proposed in this work and will produce a new form of encoded message.

3. LITERATURE SURVEY

The following is some of the prospective DNA based data hiding schemes reported recently. Hossain [5] adopted the concept of DNA cryptography in which the concept of using ASCII characters is supported Several rounds with substitution and transposition technique were used, in which the generation of dynamic table of nucleotides sequence was adopted. Chen [6] used OTP (one time padding) in which an image has been encrypted using DNA cryptography. Mohammad Reza Abbasy, et al. proposed [7] a data hiding method where data was efficiently encoded and decoded following the properties of DNA sequence.

One of the authors of this paper, Menaka [8] proposed a message encryption scheme using DNA sequences in which Complementary rules of DNA sequences are taken and indexing rules are applied to hide the message. The author has also proposed a methodology [9] on DNA cryptography and devised a novel algorithm in which 8x8 playfair cipher is used along with Ramanujam Square Matrix for the purpose of encrypting the message. Amal Khalifa and Ahmed Atito et al. discussed a way of hiding text where the text is encrypted using amino acid and DNA based on playfair cipher and also used complementary rules to hide the resultant ciphertext in a DNA sequence [10]. Mohammad Reza Abbasy, Pourya Nikfard et al. has given a DNA based cryptographic approach in which a sort of indexing method over the complementary DNA sequence was used [11]. Abhishek Majumdar and Meenakshi Sharma derived an encryption method in which high level of encryption is done using DNA sequences in the field of cryptography [12].

4. PROPOSED METHODOLOGY

The proposed methodology starts with accepting a plaintext and then converting each and every character of the plaintext into its ASCII equivalent. They are then converted into equivalent binary form. Each two digits in the converted binary sequence are taken and converted as per Table-I. Two secret keys are used in the encryption process both of them are also DNA sequences.

Table I. Dna Based Coding

Bits	***DNA Code***
00	A
01	C
10	G
11	T

The converted plaintext in DNA form, the secret keys (key1 & key2) are all stored in arrays which have high power for doing computations. The DNA codos from the corresponding positions as in Fig. 1 (for example, A[0], B[0] and C[0]—here, CGA) are taken from the arrays and converted into an equivalent amino acid (protein) representation (‘R’ here for CGA – Fig. 2). This is done for all the elements in the arrays and the corresponding protein sequence is obtained. The obtained protein sequence is now converted into ASCII equivalent which is then converted to binary form.

The next stage of this work proceeds with the effective use of the primary concept of the DES (Data Encryption Standard) algorithm which is one of the popular cryptography algorithms to protect data. DES is a symmetric encryption system in which 64-bit blocks size is used. The algorithm also uses combination, substitution and permutations along with the text to be encoded and a key is generated. DES algorithm is strongly secured and this security is mainly based on the complexity of the encryption key [14].

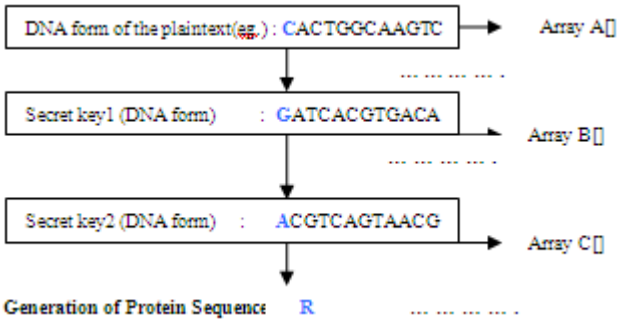

Figure. 1. Initial Stage of Encryption Process

With this usage, the algorithm advances by dividing the binary form of the message into 12 blocks of 8 bits each.

$$B_1B_2B_3B_4B_5B_6B_7B_8B_9B_{10}B_{11}B_{12}$$

Inverse table for the standard genetic code

Amino acid	Codons	Amino acid	Codons
Ala / A	GCT, GCC, GCA, GCG	Leu / L	TTA, TTG, CTT, CTC, CTA, CTG
Arg / R	CGT, CGC, CGA, CGG, AGA, AGG	Lys / K	AAA, AAG
Asn / N	AAT, AAC	Met / M	ATG
Asp / D	GAT, GAC	Phe / F	TTT, TTC
Cys / C	TGT, TGC	Pro / P	CCT, CCC, CCA, CCG
Gln / Q	CAA, CAG	Ser / S	TCT, TCC, TCA, TCG, AGT, AGC
Glu / E	GAA, GAG	Thr / T	ACT, ACC, ACA, ACG
Gly / G	GGT, GGC, GGA, GGG	Trp / W	TGG
His / H	CAT, CAC	Tyr / Y	TAT, TAC
Ile / I	ATT, ATC, ATA	Val / V	GTT, GTC, GTA, GTG
START	ATG	STOP	TAA, TGA, TAG

Figure. 2 Codon Translation Table

Substitution boxes familiarly called as S-boxes in DES are constructed with 16 rows and 16 columns and the values to be stored in the S-boxes are computed using a formula and stored in a matrix array. The 8 bits of each of the blocks $B_1B_2\ldots B_{12}$ are taken in which the first 4 bits determines the row of the S-box and the last four bits determines the column in it. By identifying a particular cell in the matrix, the value in that cell is taken and substituted for that block. Thus for each block B_i, the following is calculated.

$$S(B_1)S(B_2)S(B_3)S(B_4)S(B_5)S(B_6)S(B_7)S(B_8)$$
$$S(B_9)S(B_{10})S(B_{11})S(B_{12})$$

The values that appear in the S-box for every block are the final cipher text message of the given plaintext.

5. RESULTS AND DISCUSSION

In the proposed algorithm, a message is encrypted with numerous phases of conversion to diverse formats with insertion of secret key. Using the biological properties of the DNA sequences, the secret keys are generated.. This adds extra complexity and difficulty to encryption system. The algorithm has been implemented with ASP.NET as front end tool with Intel Pentium Dual Core processor. As in Fig. 2, representing the message in amino acid form can have massive capacity for data manipulation, substitution, and insertion to extended key generation which can be used to encode confidential information sent over public channel. From the screen shots given below, it is seen that the obtained cipher text 0 92 12 116 0 92 37 0 0 92 12 21 is entirely different from the plaintext taken ('HAI' here). The focus of the methodology is to use the combined power of the DNA sequences and the DES algorithm. The DNA and the protein codes formed intermediately also strengthen the algorithm. We first entered the keys, which are passwords known only to the sender and the intended receiver and then we gave the plaintext. The snapshots below show the generation of ciphertext after performing the procedures of the proposed algorithm. For Decrypting the message, the inverse process of the encryption is performed.

Figure 3: Screen Shots of the Proposed Technique

6. CONCLUSION

In this paper, we have pointed out that the DNA sequences have the unique properties which we can make use of for encryption purposes. The secret information has been hidden with more strength so that it will be impossible to an intruder to know about the message. Also, it is almost impossible to an intruder to predict DNA and protein sequences. Furthermore, if the ciphertext is accessed and content is revealed, the true meaning of the message will not be revealed without the key and the knowledge of the biomolecular sequences and DES algorithm. The security of the encryption process is thus enhanced with this novel approach.

REFERENCES

[1] Leuenberger, M. N., "Quantum computing in molecular magnets", Vol . 410, 12 April 2001, Nature - International Journal of Science, pp.789-793, doi:10.1038/35071024..

[2] Adleman, L. M., "Molecular computation of solutions to combinatorial problems", Vol. 266, Issue 5187, 11 Nov. 1984, Science, pp. 1021-1024, http://www.jstor.org.

[3] Watson, J. D. and Crick, F . H. C., "Molecular Structure of Nucleic Acid", Vol. 171, 1953, Nature, pp. 737-738.

[4] M.X. Lu, "Symmetric-key cryptosystem with DNA technology", Vol.50, Issue 3, June 2007, Science in China Series F: Information Sciences, pp. 324-333.

[5] Hossain, E. M., "A DNA cryptographic technique based on dynamic DNA sequence table" , 18-20 Dec. 2016, 19th International Conference on Computer and Information Technology (ICCIT), IEEE Xplore, Digital Library, pp. 270 – 275, DOI: 10.1109/ICCITECHN.2016.7860208.

[6] Chen, J., "A DNA-based, bio molecular cryptography design. Circuits and Systems", 25-28 May 2003, Proceedings of the International Symposium on Circuits and Systems, 2003. ISCAS '03, IEEE Xplore, Digital Library, pp. 822-825, DOI**:** 10.1109/ISCAS.2003.1205146

[7] Mohammad Reza Abbasy, Azizah Abdul Manaf, and M.A. Shahidan, "Data Hiding Method Based on DNA Basic Characteristics", Vol. 194, 20-22 July 2011, International Conference on Digital Enterprise and Information Systems, pp. 53–62.

[8] Menaka, K, "Message Encryption Using DNA Sequences", 27 Feb. – 1 March 2014, World Congress on Computing and Communication Technologies (WCCCT), IEEE Xplore Digital Library, pp. 182-184, DOI**:** 10.1109/WCCCT.2014.35.

[9] Meena ,K and Menaka,K, "A New Approach to DNA Cryptography Using 8x8 Playfair Cipher and Ramanujam Square Matrix", Vol. 5 Issue 8, August 2016, International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064, pp 444-446.

[10] Amal Khalifa and Ahmed Atito, "High-Capacity DNA-based Steganography", 14-16 May 2012 , 8th International Conference on informatics and Systems (INFOS2012) IEEE Xplore Digital Library, INSPEC Accession Number**:** 12864186.

[11]. Mohammad Reza Abbasy, Pourya Nikfard, Ali Ordi, Mohammad Reza Najaf Torkaman. "DNA Base Data Hiding Algorithm", Vol. 2, No. 1, Jan. 2012, International Journal on New Computer Architectures and their Applications, pp. 183-192.

[12] Abhishek Majumdar, Meenakshi Sharma, "Enhanced Information Security using DNA Cryptographic Approach", Vol. 4 Issue-2, July 2014 , International Journal of Innovative Technology and Exploring Engineering (IJITEE), pp. 72-76.