A Robust Technique to Generate Unique Code DNA Sequence

S.V.Keerthana Priya
Computer Science and Engineering
JNTUACEP, Pulivendula, A.P, India
keerthanapriya11293@gmail.com

S.J.Saritha
Computer Science and Engineering
JNTUACEP, Pulivendula, A.P, India
sjsarithajntuacep@gmail.com

***Abstract*— DNA computing works for DNA cryptography which is an emerging field. For data secure we initiate a current technique by accepting the biological structure of advanced DNA called DNA computing. There are several problems like NP complete problem solved by Adlmen. In RSA algorithm A refer to Adleman [6]. Now a day's internet has become the media for all banking, electronic commerce and communication transactions. It is very essential to provide high security for communication. To fulfill this requirement many traditional techniques like AES, DES, RSA, HASH FUNCTIONS etc. have been used for encryption and decryption of plain text. Even though these are good for security but computation complexity is increased. To overcome this draw back a new technique based on DNA is proposing in this project work. The main Futures of DNA Cryptography are Robustness, Confidentiality, Randomness, Dynamicity, Complete Character Set Fulfillment and Uniqueness. In this DNA four volumes of nucleotides they are A-adenine, C-cytosine, G-guanine &T-thymine. These 4 nucleotides carries the genetic information of a particular living thing. In order to provide better security and reliable data transmission an effective method of DNA based cryptography is proposed. In DNA cryptography method mixture of mathematical and biological concepts are used to get the encrypted data in the form of DNA sequences. By applying those 4 nucleotides we will encrypt the data in the form of DNA sequences. First and foremost thing is we will encode those nucleotides as in two bit binary format as A-00,C-01,G-10,T-11.there are two steps to encode the plain text .Firstly find ASCII code for each letter in the plain text and convert it into binary format and from that we find DNA sequence so that we get four letter DNA sequence for each letter. So in this procedure the main limitation is the data may not be secured (i. e intruders easily covert cipher text to plain text) but advantage is transmission rate take less time due to less length. So to provide more security a "robust technique" for generating the unique code DNA sequence. This technique performs same as DNA cryptography method but here we perform some operations they are addition and multiplication of DNA characters. The cipher text generated from this will be large and so it is more secure compared to existing method. But here also one limitation i.e the transmission rate of data should take large amount of time .So further proposed "DNA compression technique" to reduce the transmission rate.**

Keywords: Data structure; character operations; DNA cryptography, DNA compression.

I. INTRODUCTION

Information security is based on cryptographic a technique which is used to build secure VPN by using this technique information will be converted into unreadable format. There are several algorithms to encrypt and decrypt the data. They are AES, DES, etc. but now these are outworn**.[1]** so DNA cryptography is a new fangled to the world on DNA cryptography. For next generation security DNA play major role. Here DNA uses "central Dogma of molecular biology" (CDMB) for encryption and decryption. CDMB is a pseudo DNA cryptography method this method is efficient in computation, transmission and storage. What CDMB mean in biological format? This means once information has passed into protein it cannot get out again. The transfer of information will do from one nucleic acid to another nucleic acid and nucleic acid to protein but not protein to nuclei.

Nucleic acid $\longrightarrow$ protein

not

Protein $\not\longrightarrow$ nucleic acid .

The above all are in biological format. DNA cryptography will be difficult to use in security world because it is a labor intensive and we want several technical lab requirements so here for encryption and decryption we are not intended to use real DNA**.[2]**In DNA both biology and computers are involved. Biology involves genetic information of a particular living thing. Computers involves that how the information should be encrypted and decrypted. [5]Mathematics, statistics, physics and biology these all are come under bioinformatics. In agricultural, medical, and many living organisms these face many problems those should be solved by bioinformatics. The study of bimolecular that is present in living cells is called bioinformatics .the bio molecules also called DNA sequences. **[3]**DNA computing is an enticing surrogate to conventional computing technologies by virtue of DNA's hefty storehouse, capacity, and gigantic parallelism. This will make DNA computing fair for cryptographic applications**.[4]**Before going to principles of DNA we have to know about what is actually DNA mean? In this world all living organisms have same blueprint of genetic information but the way is different from others. The way will show that

978-1-5386-1887-5/17/$31.00 ©2017 IEEE

whether the person is male/female. And also whether he/she was affected with any disease. It also shows whether it is a human or oak tree. For each cell one substance is present this is called DNA (Deoxyribo Nucleic Acid).

The DNA nucleotide is a combination of 4 nitrogen bases, 5 carbon sugar and phosphate group. The polymers are made by the nucleotides after nitrogen base consists of A(adenine),C(cytosine),G(guanine),T(thymine).These nucleotides always pair up C with G,T with A.

These two standards are in anti parallel and run in an opposite direction fig1 shows two standard of DNA and fig2 shows double helix shape of DNA.

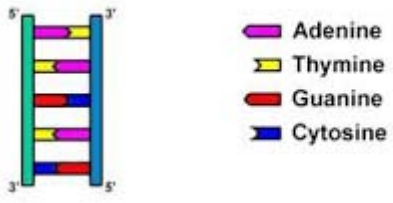


Fig. 1 Two Standard of DNA

Fig. 2 Double Helix of DNA

Features of DNA:

A. Robustness

The Encoding Technique is not breakable.

B. Confidentiality

The DNA encoding technique should help to encode the plain text in such a way that it is not possible to be deciphered.

C. Randomness

Randomness suggests that disarray or muddled in arrangement of steps is mandatory to reach more fickleness in each and every step of DNA encoding table so that the encoding table cannot be interpret.

D. Dynamicity

The DNA encoding table spawn should be dynamic across peculiar Sessions whereby every session between a sender and receiver generates and uses a new encoding table.

E. Complete Character Set Fulfillment

The DNA encoding table should lend for DNA encoding sequences for the complete character set, which contains 96 elements.

F. Uniqueness

The encoding of plaintext into DNA sequence is unique in every generation of encoding table.

One gram of DNA can cache 108 terra bytes. DNA computing is a advanced estimating model by control potential massive parallelism, high density information of bio-molecules & low power consumption which bring challenge to cryptography. Several DNA algorithms are proposed for cryptography, crypt analysis & steganography problems. Steganography mean masking of unpublished message with in an familiar message.

Nevertheless in view of DNA computing is situated on bimolecular reactions unusually it is a causes for errors.

II. LITERATURE REVIEW

Several DNA-based cryptographic techniques have been proposed [6],[7],[8],[9],[10]. Some of these use Polymerase Chain Reaction (PCR) [6],[7], while others use DNA chip technology [8],[9]. A DNA tile assembly model has also been used for one-time pad cryptosystems [8],[10]. One-time pad systems are especially attractive since, if used correctly, they are almost unbreakable. But they require a large library of DNA sequences. In [5] DNA – BASED RSA ALGORITHM is implemented. This algorithm initially succeeded in overcoming some main problems in "RSA". Here they were not intended to use real DNA strands to implement cryptographic algorithm, but only to simulate some mechanisms of the process of the central dogma of molecular biology. In [11] DNA-based Cryptography using DNA sequence for secure communication and key distribution, and using the chemical information of biological alphabets for steganography – Information Hiding. In [12] Pseudo DNA cryptography is used to overcome the limits of using the DNA cryptography. In [13] DNA computing methods are implemented. Here they are implemented DNA-based signature scheme, a contract for playing mental poker on the wetware, and an RNA-based zero-knowledge proof system situated on determining the Sudoku problem. Hand over basic of classical cryptography as it provides a variety of advantages over conventional silicon-based computing paradigms. In [16] PCR-based amplification technology of DNA In order to solve the key space-constrained problem that the PCR amplification technology of DNA has, the authors used a method for building a chaotic system. This system includes a logistic chaotic map and a Henon chaotic map. We can generate a chaotic pseudo-random sequence which could handle the plaintext for eliminating the statistical rules in it with the two maps. Improving security and the key space, and it provides an operational test of it.

Author	Method	Title	Limitations
O.Tornea &M.E.Borda	Plaintext to ASCII to binary to DNA	A novel DNA computing based encryption and decryption algorithm	Except complete character set fulfillment remaining all features of DNA is not supported
Zhang, Qiang et.al	Plaintext to ASCII to binary to DNA	Review on DNA based encryption and steganography	Except security remaining features of DNA is not support
Xianglian Xue & gehani et.al	Original image to binary to DNA	A novel image encryption algorithm based on DNA subsequences	Except security remaining features of DNA is not support
Guangzhao cui	Plaintext to	DNA	

Author	Method	Title	Limitations
et.al	hexadecimal to binary to DNA	cryptography based on symmetric key exchange	Except confidentiality remaining all features of DNA is not supported
Sherif T.amin et.al	Plaintext to ASCII to binary to DNA	DNA based secure data hiding technique for cloud	Except complete character set fulfillment remaining all features of DNA is not supported
Akanksha Agarwal	Preferred manual table to DNA	Novel DNA based encryption and decryption	Except complete character set fulfillment remaining all features of DNA is not supported

In [21] defined a one-time pad cryptosystem using DNA self-assembly and showed that self-assembly is more efficient than PCR for generating the DNA sequences needed by our system. In [22] they are providing reliability in DNA based computations and it improves previously reported methods in [23].

[25]Compression Techniques Applied to DNA Data of Various Species. 4 contrasting algorithms for DNA data compression: LZW (Lampel-ziv-Welch) algorithm, run length encoding algorithm, Arithmetic coding and Substitution method. The compression results on these algorithms are conferred and correlated on DNA sequence data of 10 divergent species. Compression of these types of target phenoms could be carried out applying a reference genome for better compression results.

[26]A luminous density secure image encryption scheme based on disordered & DNA computing luminous density secure cryptographic scheme for secure image communication. In this scheme the transparent image is allowed first accepting a sequence of pseudo random number (PRN) and encrypted by Deoxyribo Nucleic Acid (DNA) computation. Two PRN sequences are provoke by a Pseudo Random Number Generator (PRNG) situated on cross coupled chaotic logistic map accepting two sets of keys. The first PRN sequence is accepted for allowing the plain image considering the second PRN sequence is approved for provoking irregular DNA sequence.

[12]DNA Solve the Gnomic problems in the paper DNA Sequence Compression

A Compression Model for DNA various Sequence Alignment Blocks squeezing modeling path for the multiple sequence alignment(MSA) blocks, which form maximum of these datasets. Our method is positioned on a batter of finite-context models. At variance with other new approaches, it addresses both the DNA bases and gap symbols at once, better exploring the existing correlations.

[14]DNA Lossless Compression Algorithms: Review

Here implemented the DNA by the comparison such as row column and Standard comparisons with the elements of the DNA.

[13]DNA lossless differential compression algorithm positioned on affinity of genomic sequence database. Here achieved DNA for the Genomic data. More work is desired to prefer the prefect hint for immense data to enhance difference compression generation, and to further scrutinize the use of the late method in realistic genomic applications.

[16]Vertical DNA Sequences Compression Algorithm Based on Hexadecimal Representation

The main strength of our algorithm is that it allows, using a simple arithmetic operation, to measure a pat of DNA sequences and spot regions of affinity between them. The algorithm is also very obvious to resolve and his acceptance is fortunate because a hexadecimal character code two nucleotides. In the forthcoming, we will shot to accomplice our algorithm to alternative vertical compression algorithms positioned on analytical way in order to choose the sequences that can finest produce a set of DNA sequences.

III. EXISTING METHOD

For existing method we take fast three level DNA security there are three levels. In the first level shift the text by using the shift key and perform its complement. In the second level operate the LBP. Later the LBP take the binary format of the value. In the third level convert this LBP result into DNA sequence.

Encryption

/* First level of security*/

- The plain text has to be shifted with some key length then we get shifted text.
- For the shifted text ASCII value is to be finding out.
- ASCII value will be converting as binary value.
- The binary value should be converted into 1's complement.

/*Second level*/

- Attain the LBP operation on the first level output.
- Get the new binary format.

/*Third level*/

- Convert this value to DNA sequence format by using the DNA conversion table. Shown below

Alphabets	Binary Representation
A	00
C	01
G	10
T	11

- Then obtained values send as encrypt text.

The following are the sequence of steps in decryption process.

Decryption

- By applying the same rule sequence from the table we get the DNA into binary format
- On binary data LBP operation should be performed.
- Convert the binary form which will get after LBP operation into 1's complement.
- Covert the binary values into segment of 8 and replace with correspondent ASCII value.
- With the length of key we will shift the obtained text.
- The obtained value is plain text.

Limitations

In fast three level security time complexity is low but here we get four DNA sequences for one letter so it take less time to encrypt data and there is a chance for intruders to hack data easily so to overcome that I go for robust technique to generate unique code DNA sequences.

IV. PROPOSED METHOD

In this paper it will overcome the limitations of existing method(fast three level security) i.e. this provide more security compare to existing one this entiles that for existing we get four DNA sequence for one letter and for this present method it will get 8 DNA sequences for one letter. But here also one limitation i.e. time complexity is high, storage also have to large. For future enhancement we will use best compression technique. In this paper also have three steps. For getting a large many DNA sequences we have to perform operations like addition and multiplication. Here we take string matrices format for addition operation no need to arrange in matrix format but for multiplication we arrange it in matrix format i.e. rows and columns arrangement.

Encryption

/* First level of security*/

- Take the plain text and that text should be shifted with some key length.
- Convert the shifted text into relevant ASCII value.
- Convert that ASCII into binary format.
- Convert the binary into DNA sequence as given in table below considered it as "x".

/*Second level*/

- Convert the generated binary format into 1's compliment
- Attain the LBP operation on the generated 1's complement.
- Get the new binary format.
- Convert the new binary value into DNA sequence as given in table consider it as y.

/*Third level*/

- Convert this value to DNA sequence format by using the DNA conversion table. Shown below

Alphabets	Binary Representation
A	00
C	01
G	10
T	11

- Here we perform operation on both DNA sequences generated above like addition, multiplication etc. i.e. x+y, x*y
- Then obtained values send as encrypt text

The following are the sequence of steps in decryption process.

Decryption

- Divide the generated DNA into two parts.
- The both parts will be converted into binary format by applying same rule sequence..
- For first BINARY format we find ASCII value
- On second binary data LBP operation should be performed.
- Convert the binary form which will get after LBP operation into 1's complement.
- Convert this value also in ASCII format
- With the length of key we will shift the obtained text.

The obtained value is plain text.

V. EXPERIMENTAL RESULTS

A. Existing method

- Consider the plain text as HEN.
- Let the length of the shift key as 5.
- The shifted text is: KZQ
- The ASCII values of

  M (77) → 01001101
  J (74) → 01001010
  S (83) → 01010011

- Complement of the binary values

  01001101 → 10110010
  01001010 → 10110101
  01010011 → 10101100

- LBP for the complement data of the shifted text

  The order in which the binary values are arranged is:

For M			For J			For S		
1	0	1	1	0	1	1	0	1
1	M	0	1	J	0	0	S	1
0	1	0	1	0	1	1	0	0

Here the data will be arranged by the row wise and get the output by the column wise.

- LBP values for
  M → 11001100
  J → 11100101
  S → 10100110
- DNA sequence for the LBP values by using DNA conversion table shown below.

Alphabets	Binary Representation
A	00
C	01
G	10
T	11

Conversion from LBP output to DNA sequence is 11001100
Here take the MSB and LSB values 10 it is equal to G

1) Reaming data is 100110 here MSB &LSB are 10 equal to G
2) Perform the above step until we complete the all the bits

Finally we get the sequence

11001100 →10100011 →GGAT
11100101 →11101001 → TGGC
10100110 →10011001 → GCGC

Cipher text is **GGATTGGCGCGC.**

B. Propose method

- Consider the plain text as HEN.
- Let the length of the shift key as 5.
- The shifted text is: KZQ
- The ASCII values of
  M (77) → 01001101
  J (74) → 01001010
  S (83) → 01010011
- Convert this into DNA sequence by using DNA sequence table as shown below

Alphabets	Binary representation
A	00
C	01
G	10
T	11

For M: CATC
For J: CAGG
For S: CCAT

- Complement of the binary values
  01001101 → 10110010
  01001010 → 10110101
  01010011 → 10101100

- LBP for the complement data of the shifted text:
  The order in which the binary values are arranged is:

For M			For J			For S		
1	0	1	1	0	1	1	0	1
1	M	0	1	J	0	0	S	1
0	1	0	1	0	1	1	0	0

Here the data will be arranged by the row wise and get the output by the column wise.

- LBP values for
  M → 11001100
  J → 11100101
  S → 10100110
- DNA sequence for the LBP values by using DNA conversion table shown below.

Alphabets	Binary Representation
A	00
C	01
G	10
T	11

Conversion from LBP output to DNA sequence is 11001100
Here take the MSB and LSB values 10 it is equal to G

1) Reaming data is 100110 here MSB &LSB are 10 equal to G
2) Perform the above step until we complete the all the bits

Finally we get the sequence

11001100 →10100011 →GGAT
11100101 →11101001 →TGGC
10100110 →10011001 →GCGC

- Here we perform operations on 5th & 9th step we get final output:

CATC + GGAT =CGAGTACT

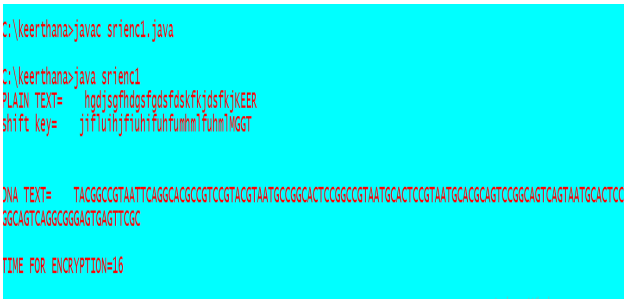
CAGG + TGGC =CTAGGGGC

CCAT + GCGC =CGCCAGTC

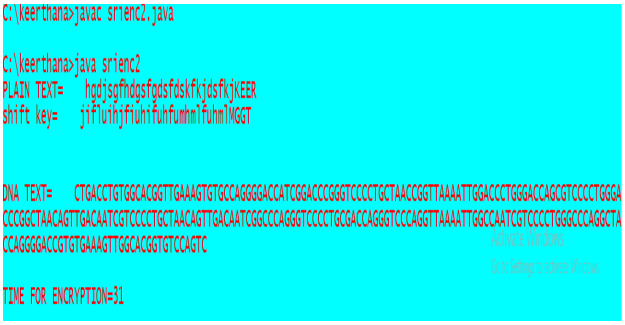
Final cipher text is **CGAGTACTCTAGGGGCCGCCAGTC**

The Sample Output for Existing Method and Proposed Method is shown below:

Existing Method Output:



Proposed Method Output:



Comparing both outputs, time complexity and lines of code are high in proposed method. Due to lines of code is high intruders will not hack the data easily. Hence the Proposed Method is more secure than Existing Method

VI. CONCLUSION AND FUTURE SCOPE

This paper describes a firm technique to bring about unique code DNA sequences for encryption and decryption. During the time of encryption and decryption work of security plays a major role. In this paper an unfamiliar method is introduced to bring about cipher text. This technique wills impairment operation on DNA sequences like add and multiply. Extremely we get 8 letters of DNA sequences from single letter so it afford better security correlate to existing. Existing equip 4 from single letter. But in this paper limitation is storage capacity had to be high. For future enhancement best compression technique has to be finding to reduce storage and time complexity.

References

[1] Representation" published in proceedings of the world congress on engineering and computer science 2015 vol.II. Noorul Hussain UbaidurRahman,ChithralekhaBalamurugan,Rajapandian Mariappan."A novel DNA computing based encryption and decryption algorithm" published in ICICT 2014.

[2] Vilas Machhi and Maulika S patel."Compression technique applied to DNA data of various species" published in international journal of bio-science and bio-technology.vol.8. no.3(2016).

[3] R. pradeep kummar reddy, C.Nagaraju and N.Subramanyam "text encryption through level based privacy using DNA stegnography "published in international journal of emerging trends and technology in computer science-2014.vol.3

[4] Bacem Saada, jing Zhang "Vertical DNA sequences compression algorithm based on hexadecimal

[5] A. AppaRao, "DNABIT compress-compression of DNA sequences" in proc, the biomedical informatics,2011.

[6] G. Cui, L. Qin, Y. Wang and X. Zhang, "An encryption scheme using DNA technology", Bio Inspired Computing: Theories and Applications, 2008, pp. 37-42.

[7] D. Prabhu and M. Adimoolam, "Bi-serial DNA encryption algorithm (BDEA)", [online]. Available <arXiv:1101.2577v1> [Jul 12, 2012].

[8] A.Gehani, T. LaBean and J. Reif, "DNA based cryptography" in Aspects of Molecular Computing 2950, N. Jonoska, G. Paun and G. Rozenburg, Ed. Germany: Springer, 2004, pp.34-50.

[9] M. Shyam, N. Kiran and V. Maheswaran, "A novel encryption scheme based on DNA computing", 14th IEEE International Conference on High Performance Computing, 2007.

[10] Z. Chen and J. Xu, "One-time-pads encryption in the tile assembly model," Bio-Inspired Computing: Theories and Applications, 2008, pp.23-30.

[11] Biological Alphabets and DNA-based Cryptography Qinghai Gao Department of Security Systems, Farmingdale State College, SUNY

[12] Auday H. Saeed Al-Wattar, Ramlan Mahmod, Zuriati Ahmad Zukarnain & Nur Izura Udzir University Putra Malaysia, Faculty of Computer Science & Information Technolgy, Selangor-Serdang, Malaysia

[13] Implementing Modern Cryptographic Protocols Using DNA and RNA Information Processing ArashKarimi †, †Iran University of Science and Technology (IUST), Narmak, Tehran, Iran

[14] Research on DNA Cryptography Yunpeng Zhang* and Liu He Bochen Fu College of Software and Microelectronics, Northwestern Polytechnical University, Xi'an, China

[15] AnushaMantha, "Improving reliability of DNA based computations with applications to cryptography", MS thesis, University of Cincinnati, 2012.

[16] Improving Reliability in DNA-based Computations AnushaMantha, George Purdy, and Carla Purdy School of Electronic and Computing Systems University of Cincinnati Cincinnati, USA manthaaa@mail.uc.edu, george.purdy@uc.edu,

[17] "DNA data cryptography"