IJERMS

International Journal of Engineering Researches and Management Studies

DATA ENCRYPTION USING DNA SEQUENCES BASED ON COMPLEMENTARY RULES

MS. Amruta D. Umalkar*, Prof. Pritish A. Tijare, Prof. Swapnil N. Sawalkar

Master of Engineering Information Technology Department, Sipna College of Engg. and Technology Amravati, Maharashtra, India.

ABSTRACT

With the quick development of net technology and data process technology, the knowledge is unremarkable transmitted via the net. The vital data in transmission is definitely intercepted by unknown person or hacker. So as to reinforce the knowledge security, encryption becomes a vital analysis is direction. A message cryptography formula supported deoxyribonucleic acid (Deoxyribo Nucleic Acid) sequence for presenting during this paper. The most purpose of this formula is to write the message with the premise of complementary rules deoxyribonucleic acid sequence.

KEYWORDS: Data Encryption; DNA Sequences; Complementary Rules, Secure Transmission and reception.

INTRODUCTION

The security of a system is essential now a days. With the growth of the information technology power, and with the emergence of new technologies, the number of threats a user is supposed to deal with grew exponentially.

In the encoding and secret writing of information is completed with the help of key. The most secure and presently used technique is that the trendy ways of encoding that involves a lot of mathematical computations and two kinds of keys, the public and private keys. Nowadays, there is another recently rising encoding technique known as DNA encoding. The most objective of this technique is to write the plaintext and hide it within the DNA digital kind. DNA encoding permits the confidentiality of information a lot of the trendy ways with the utilization of complementary rules.

Today, biotechnology is applicable to several aspects of life, and the way to use biological information as a carrier to cover information has become a motivating challenge. DNA sequences have some inherent properties which will be used to cover information because it is difficult to differentiate between a true DNA sequence and a pretend one.
The DNA segments that hold this genetic data are known as genes; however other DNA sequences have structural functions or are concerned in modifying the utilization of this genetic data. Similar to a string of binary information is encoded with ones and zeros, a strand of DNA is encoded with four bases, represented by letters A (Adenine), T (Thymine), C (Cytosine) and G (Guanine). The data in DNA is keep as a code created from these four chemical bases. The combination of the bases results in purines (combination of Adenine and Guanine) and pyrimidines (combination of Cytosine and Thymine) as shown in figure1.

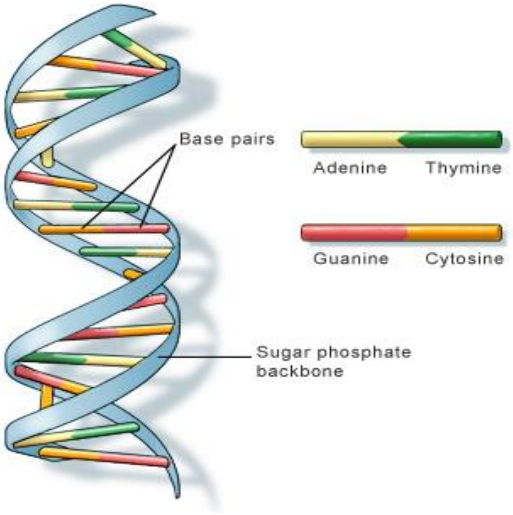


Figure 1. Structure of DNA Molecule

The two strands of a DNA molecule are antiparallel wherever every strand runs in an opposite direction. This complementarily makes DNA a singular arrangement for computation and may be exploited in many ways.

DNA encryption is emerging as a new encryption field where DNA is used to carry the information. The interesting features about the structure of DNA are the complementary rule. These rules are used for proposing message encryption methods.

In the proposed algorithm, a DNA sequence or structure is first randomly taken and complementary rules are framed so the stealth message to be sent is encoded at the sender's side. At the receiver's side, the decoding method is completed and also the original message is extracted out. A DNA sequence could be a sequence composed of 4 distinct letters, A, C, G and T. every nucleotide contains a phosphate connected to a sugar molecule (deoxyribose) and one of four bases, guanine (G), cytosine (C), adenine (A) or thymine (T). It's the arrangement of the bases during a sequence, for example like ATTGCCAT that determines the encoded gene. The natural sequence pattern with complementary secret writing and chemical classification of the nucleotides are often will be protect the data.

Table I. DNA Based Coding

DNA Base	Code
A	00
C	01
G	10
T	11

\

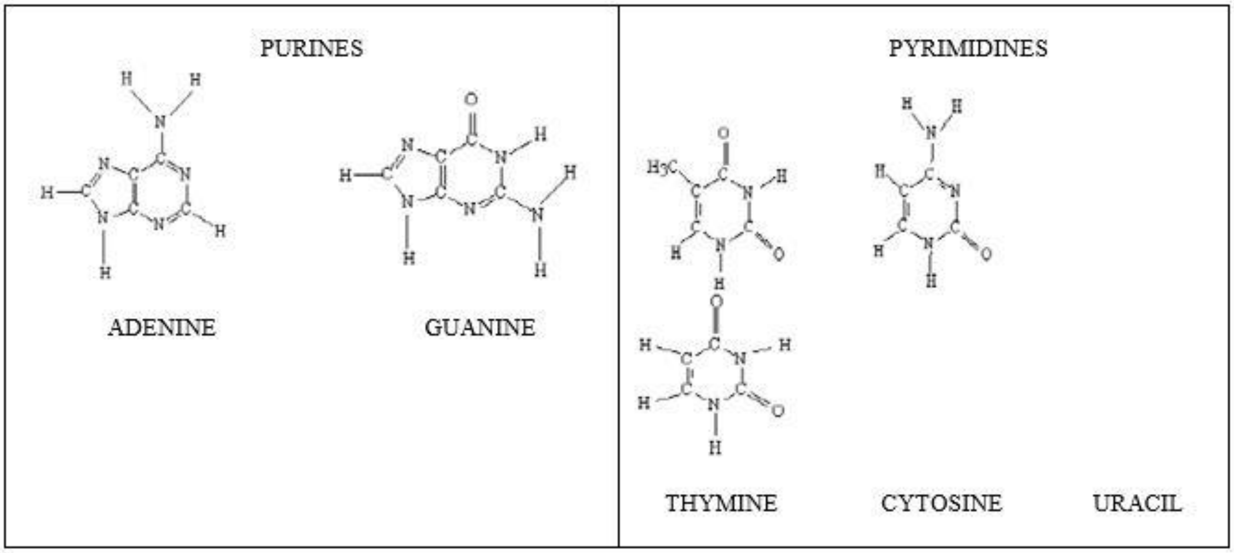


Figure 2. Structure of Purines and Pyrimidines

LITERATURE SURVEY

Message encryption using DNA sequence is a very new technique still evolving and tried out for secure transmission and reception of hided messages. The method is deemed to be so secure that it would be very difficult for any intruder to break the encrypted message and retrieve the actual message. Only the intended receiver can decrypt and receive the original message.

The following is some of the prospective DNA based messages encryption and data hiding schemes reported recently.

K. Menaka [1], proposed a data hiding method where the algorithm first randomly selects a DNA sequence. The message to be encoded is then taken and each letter in the faked DNA sequence. Each letter in the message is converted into its ASCII equivalent and they are then converted into equivalent binary form. Each two digits in the converted binary sequence are converted as per Table 1. Then, the message index position (first position of each letter) in the faked DNA sequence is applied to each letter of the converted sequence. Each digit in the resultant sequence is replaced with its equivalent three digit binary value and the equivalent alphabet value is replaced for the binary value. For example, if the obtained binary value is 010 011 101 …, then it will be replaced as C D F… where A has the value 000, B has 001 and so on. The resultant sequence of alphabets is transmitted over to the receiver. In the receiver side, the reverse process is done in which the original receiver knows the complementary rules and the randomly selected DNA sequence. The message to be sent is then encoded with the fake DNA sequence.

Debnath Bhattacharyya[2] developed an algorithm for data encryption using DNA sequencing. In their algorithm, they have used the concept of indexing the DNA Sequencing and transmitting the message to the receiver. They have not used any complementary rules.

Jin-Shiuh Taur et al. [3] proposed a way referred to as Table Lookup Substitution methodology (TLSM) that might double the capability of message activity. In TSLM, they need replaced the complementary rule with a rule table. The key plan of the TLSM is to increase the 1-bit complementary rule into a 2-bit rule table so every conversion of letters will represent 2 bits of the secret message.

In the method by Cheng Guo, Shiu,[4] the hiding procedure substitutes another letter for an existing letter on a special location set by the algorithm. The embedding algorithm encompasses a conversion operates that converts a given letter with a selected letter outlined by the complementary rule. For example, if a complementary rule is outline as (AC)(CG)(GT)(TA), then the result of θ(G) are going to be T, and therefore the result of θ(T) are going to be A. To boot, the substitution methodology can convert the letter s into s(unchanged), θ(s) and θ (θ (s)) once the secrete message is 0, 1 and no data, respectively.

Mohammad Reza Abbasy, et al. proposed [5] an information hiding methodology wherever data was efficiently encoded and decoded following the properties of DNA sequence. Complementary combine rules of DNA were employed in their methodology.

Kritika Gupta* Shailendra Singh[6] has been projected a DNA Based Cryptological Techniques for an encryption algorithm based on OTP (one-time-pad) that involve data encryption using traditional mathematical operations and/or data manipulating DNA techniques. However once an encryption algorithm has been applied and therefore the data is transmitted on the transmission media: there's a clear stage that the data, although within the cipher type gets manipulated by any interceptor.

Snehal Javheri, Rahul Kulkarni[7] proposed an algorithmic program has two phases in consequence: these are Primary Cipher text generation using exploitation substitution methodology followed by Final Cipher text generation exploitation DNA digital secret writing.

In the Primary Cipher text generation phase, the coding algorithmic program uses OTP (one-time-pad) key generation theme, since nearly one key for one piece of data is sufficient to supply voluminous strength in coding technique. The projected methodology uses indiscriminately generated symmetrical key of 8 bits size by the supposed receiver and provided to the sender. Therefore the sender can have partial information of the personal key solely and so it generates the remainder part of the keys to cipher the data.

The Byte values are extracted from the input data or message. The additional secret writing method works on unsigned byte values of the input data or text referred to as plain text. These byte values are replaced by combination of alphabets and special symbols exploitation substitution methodology. And so this substitution worths are regenerate into its binary value. So as to embed lots of security additional bits are padded at each ends of the first cipher text. These additional bits are nothing however the file size information that is provided to the receiver through key. So the secret key, the data of primer pairs are shared between sender and receiver through the secret key channel.

In the DNA digital secret writing section, the Ultimate Cipher text is generated from Primary Cipher text exploitation DNA digital encryption technique. From a process purpose of read, cannot process the DNA molecules as in sort of alphabets, therefore the DNA sequence encryption is employed during this methodology through that the binary knowledge is regenerate into DNA format and it's vice versa.

Guangzhao Cui #1, Limin Qin #2, Yanfeng Wang #3, Xuncai Zhang #4 [8]proposed a secret writing theme by exploitation the technologies of DNA synthesis, PCR amplification and DNA digital secrete writing additionally because the theory of ancient cryptography. The supposed PCR two primer pairs was used because the key of this theme that not severally designed by sender or receiver, however severally designed by the entire cooperation of sender and receiver. This operation might increase the safety of this secrete writing theme. The standard secretes writing methodology and DNA digital cryptography is wont to preprocess to the plaintext. Through this preprocess operation will get fully different ciphertext from the identical plaintext, which might effectively stop attack from a potential word as PCR primers. The quality of biological troublesome issues and cryptography computing difficulties give a double security safeguards for the theme. And therefore the security analysis the secrete writing theme has high confidential strength.

Ritu Gupta, Anchal Jain [9] symmetric-key encoding algorithmic rule supported the DNA approach is projected. The initial key sequence is enlarged to desire length victimization projected key growth technique guided by the pseudo random sequence. The advantage is that there's no need to send an extended key over the channel. The variable key growth in encoding method combined with DNA addition and complement makes the technique sufficiently secure. A DNA sequence consists of four nucleic acid bases A (adenine), C (cytosine), G (guanine), T (thymine), wherever A and T are complementary, and G and C are complementary. Also use C, T, A and G to denote 00, 01, 10, 11 (the corresponding decimal digits are "0123"). By victimization this encoding technique every 8-bit component worth of the gray scale image is pictured as a nucleotide string of length four. Reciprocally to decrypt the nucleotide string will get a binary sequence simply. In total 4! = 24 forms of writing, there are only 8 of them will meet complementary rule , for instance, the decimal digits "0123" (the corresponding binary range is

"00011011") will be encoded in to one of them, like "CTAG", "CATG", "GATC", "GTAC", "TCGA", "TGCA", "ACGT" or "AGCT". There are total six legal complementary rules [3] that are as follows:
(AT)(TC)(CG)(GA),(AT)(TG)(GC)(GA),(AC)(CT)(TG)(GA),
(AC)(CG)(GT)(TA),(AG)(GT)(TC)(CA),(AG)(GC)(CT)(TA).
Any one of them for instance, (AG) (GC) (CT) (TA) is applied to projected methodology.

ANALYSIS OF PROBLEM

The message encryption algorithm has many steps to break and to get the original message. The random selection of DNA sequence can be increased to many numbers. The complementary rules which are formed based on the properties of DNA could also be increased since DNA sequence has many biological properties and using those properties also some more complementary rules can be formed.

The complementary rules which are formed based on the properties of DNA could also be increased since DNA sequence has many biological properties and using those properties also some more complementary rules can be formed for message encryption.

PRAPOSED WORK

In order to convert binary information into amino acids as a DNA sequence, the base pairing rules should be used. Synthesizing nucleotides in real surroundings (biology) is completed in constant rules:

- Purine adenine (A) forever pairs with the pyrimidine thymine (T)
- Pyrimidine cytosine (C) forever pairs with the purine guanine (G)

In binary computing area, it's possible to alter the natural rules by own decision. For instance, in biology A is synthesized to T whereas we will assume A to C or A to G, and so on, as we have a tendency to like. Increasing the complexness of the algorithm is that the main purpose of the changing the rules.

Consider A=00, C=01, G=10, and T=11 to convert binary message to DNA sequences. The way to extend the complexness is complementary pair rule. Complementary pair rule could be a unique equivalent pair that is allotted to each nucleotide base pair.

Complementary Rule: (CTC) (AGA) (TCT) (GAG)

Based on Purine and Pyrimidines

In this proposed method, the message to be encoded is taken then converted it into its 8-bit binary equivalent. Then sampling of 8-bit binary into 2-bit. Converted this 2-bit sampling according to DNA code as shown in table 1 and we get the coded DNA string. Now generate valid DNA string on the basis of nucleotide base pair like ATCGATCG.

Select and apply complementary rules on the coded DNA string. For example the coded DNA string is like CAGACGGCAGAA when we apply complementary rule on it C is replace with TC, A is replace with GA, G is replace with AG and so on. Then we get the faked DNA string like TCGAAGGATCAGAGTCGAAGGAGA

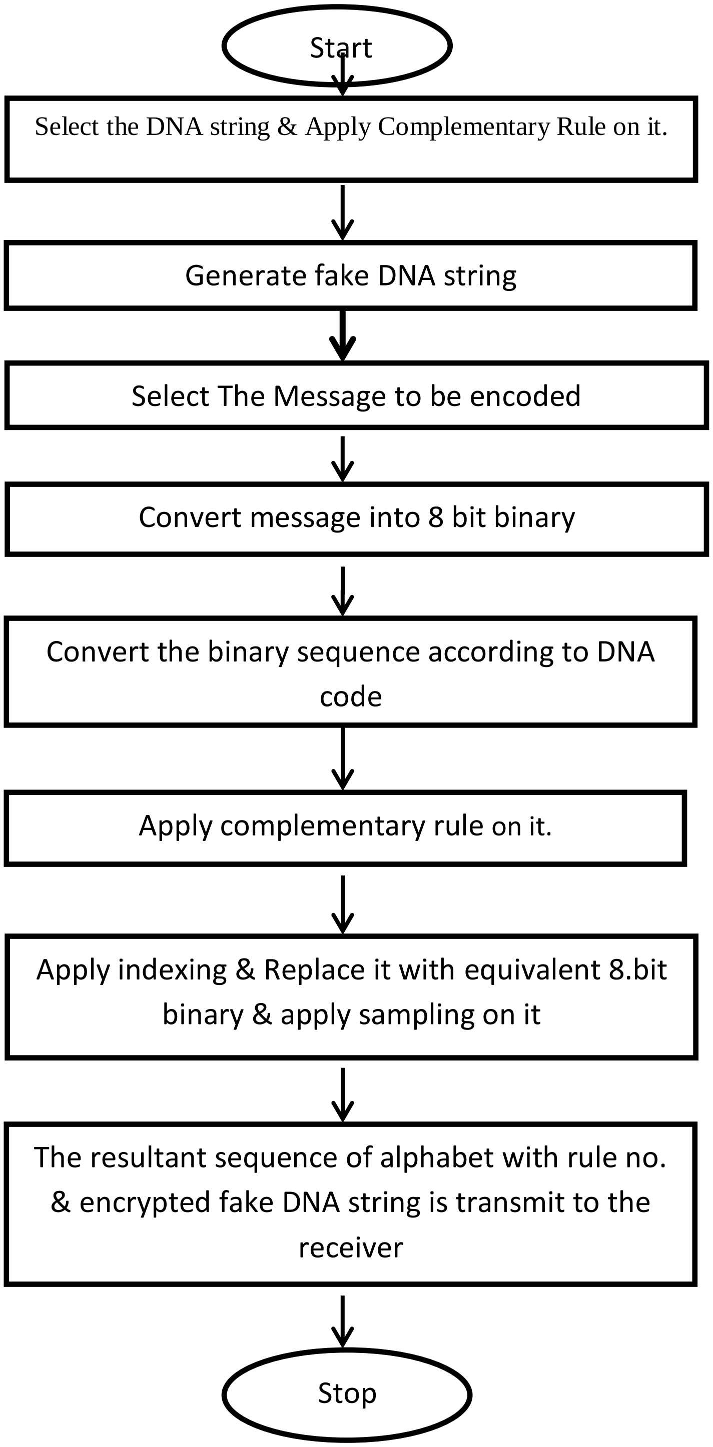
.Then apply indexing on the valid DNA string like $A_0T_1C_2G_3A_4T_5C_6G_{7.}$ After applying indexing converted it into 8-bit binary equivalent and then sampling it into 4- bit binary.

Each digit in the resultant sequence is replaced with its equivalent four digit binary value and the equivalent alphabet value is replaced for the binary value. For example, if the obtained binary value is 0010 0011 0101 … , then it will be replaced as C D F… where A has the value 0000, B has 0001 and so on. The resultant sequence of alphabets is transmitted over to the receiver.

In the receiver side, the reverse process is done in which the original receiver knows the complementary rules and the randomly selected DNA sequence. The message to be sent is then encoded with the fake DNA sequence and transmitted.

In the following flow chart showing the process of encryption and decryption.

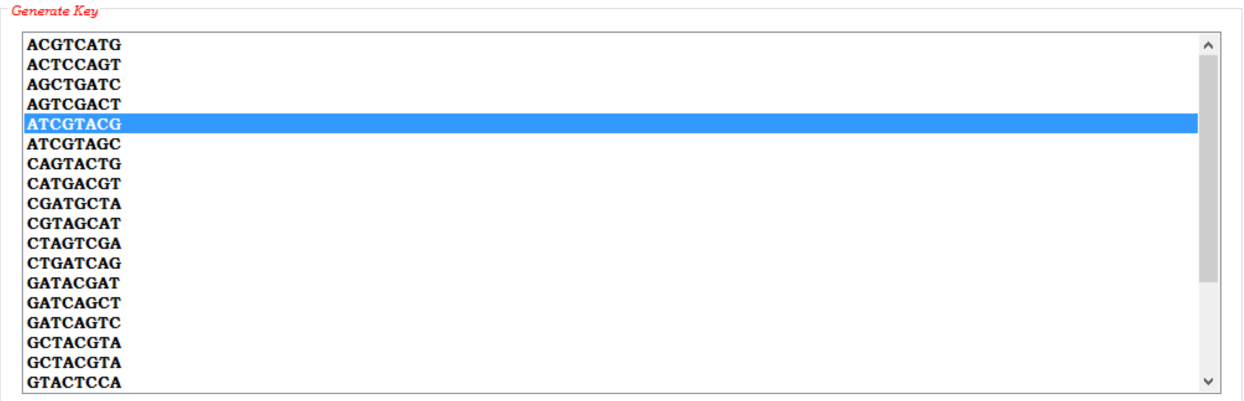
THE FLOWCHART OF ENCRYPTION / DECRYPTION:


Start
Select the DNA string & Apply Complementary Rule on it.
Generate fake DNA string
Select The Message to be encoded
Convert message into 8 bit binary
Convert the binary sequence according to DNA code
Apply complementary rule on it.
Apply indexing & Replace it with equivalent 8.bit binary & apply sampling on it
The resultant sequence of alphabet with rule no. & encrypted fake DNA string is transmit to the receiver
Stop

RESULT AND ANALYSIS

In Encryption:-

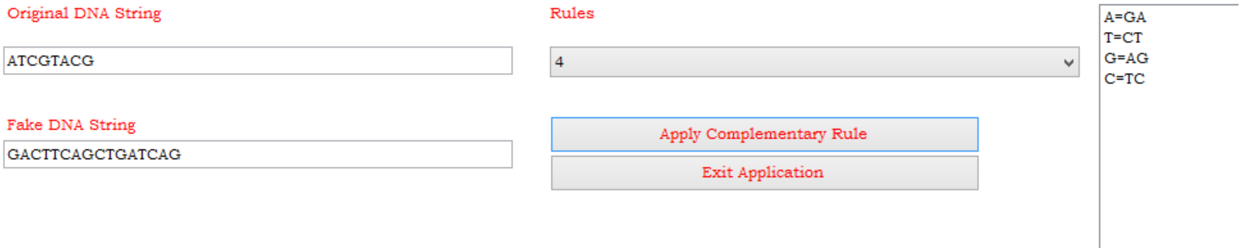
- **First select the DNA string.**



Generate Key

```
ACGTCATG
ACTCCAGT
AGCTGATC
AGTCGACT
ATCGTACG
ATCGTAGC
CAGTACTG
CATGACGT
CGATGCTA
CGTAGCAT
CTAGTCGA
CTGATCAG
GATACGAT
GATCAGCT
GATCAGTC
GCTACGTA
GCTACGTA
GTACTCCA
```

- **Apply complementary Rule on it and generate the fake DNA string.**



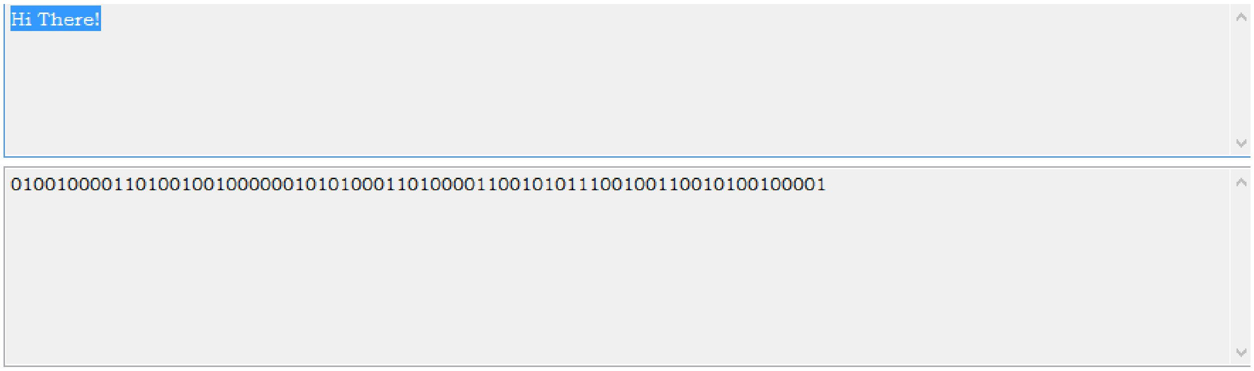
Original DNA String: ATCGTACG

Rules: 4

A=GA
T=CT
G=AG
C=TC

Fake DNA String: GACTTCAGCTGATCAG

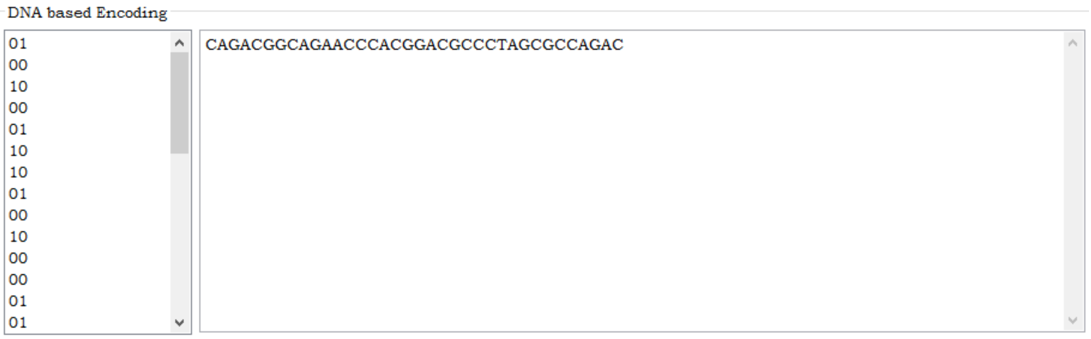
- **First select message to be encode and converted it into 8-bit binary.**



```
Hi There!
```

```
010010000110100100100000010101000110100001100101011100100110010100100001
```

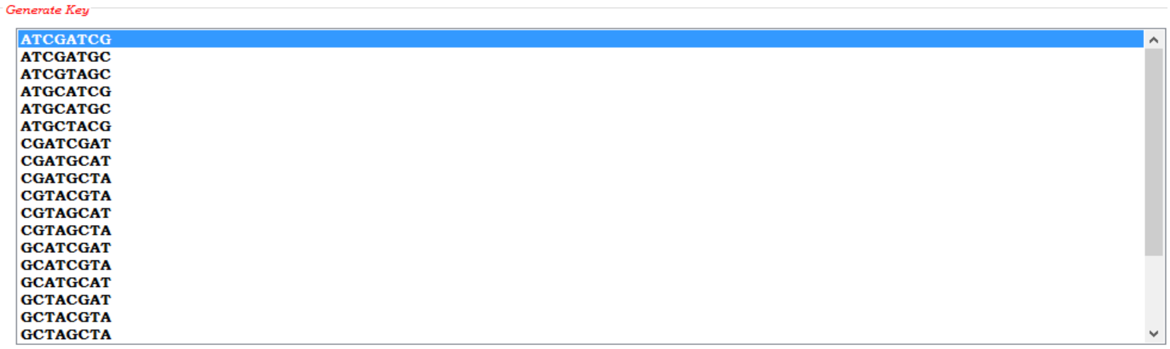
- **Apply sampling and converted as per DNA coding.**



```
01
00
10
00
01
10
10
01
00
10
00
00
01
01
```

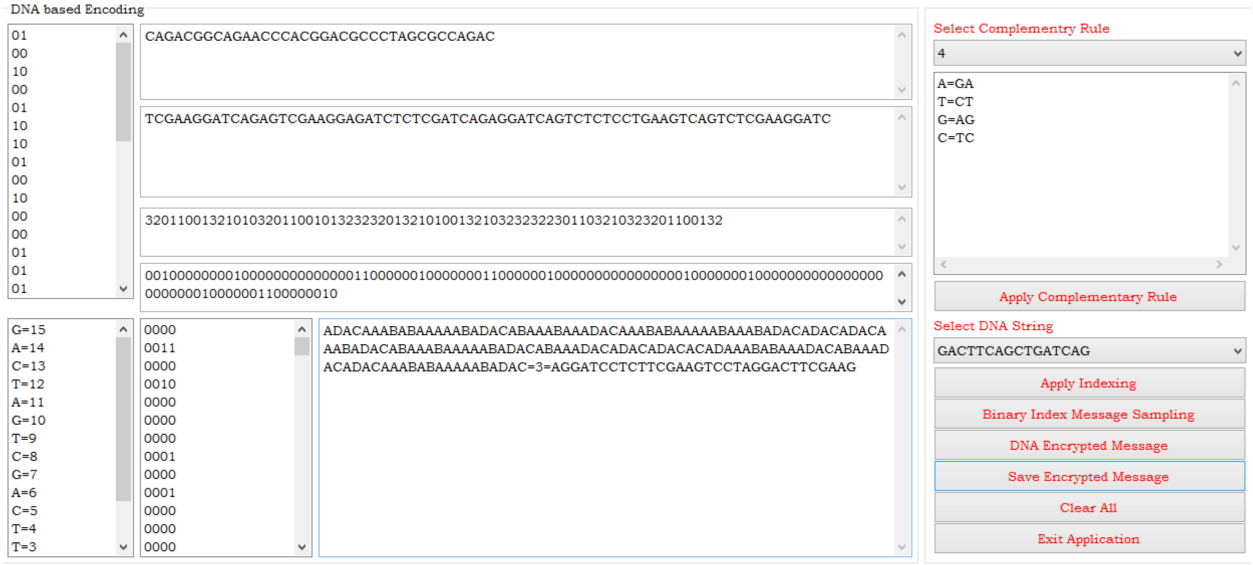
CAGACGGCAGAACCCACGGACGCCCTAGCGCCAGAC

Select DNA String



```
ATCGATCG
ATCGATGC
ATCGTAGC
ATGCATCG
ATGCATGC
ATGCTACG
CGATCGAT
CGATGCAT
CGATGCTA
CGTACGTA
CGTAGCAT
CGTAGCTA
GCATCGAT
GCATCGTA
GCATGCAT
GCTACGAT
GCTACGTA
GCTAGCTA
```

DNA based encoding



```
01
00
10
00
01
10
10
01
00
10
00
00
01
01
01
```

CAGACGGCAGAACCCACGGACGCCCTAGCGCCAGAC

TCGAAGGATCAGAGTCGAAGGAGATCTCTCGATCAGAGGATCAGTCTCTCCTGAAGTCAGTCTCGAAGGATC

32011001321010320110010132323201321010013210323232230110321032320110013 2

0010000000010000000000000001100000010000000110000001000000000000000010000000100000000000000000
0000000100000011000000010

```
G=15
A=14
C=13
T=12
A=11
G=10
T=9
C=8
G=7
A=6
C=5
T=4
T=3
```

```
0000
0011
0000
0010
0000
0000
0000
0001
0000
0001
0000
0000
0000
```

ADACAAABABAAAAABADACABAAABAAADACAAABABAAAAABAAABADACADACADACA
AABADACABAAABAAAAABADACABAAADACADACADACACADAAABABAAADACABAAAD
ACADACAAABABAAAAABADAC=3=AGGATCCTCTTCGAAGTCCTAGGACTTCGAAG

Select Complementry Rule: 4

```
A=GA
T=CT
G=AG
C=TC
```

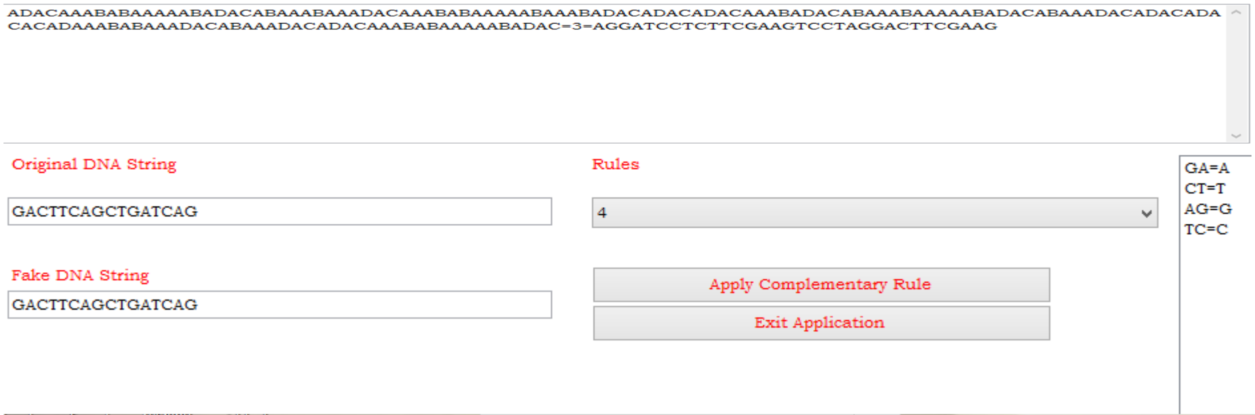
Select DNA String: GACTTCAGCTGATCAG

Apply complementary rule on DNA coded string. And apply indexing on the valid DNA string then converted it into 8-bit binary and sampling it into 4-bit binary. Each digit in the resultant sequence is replaced with its equivalent four

digit binary value and the equivalent alphabet value.The resultant sequence of alphabets is transmitted over to the receiver.
In the receiver side, the reverse process is done.

In Decryption :-

- In decryption, encoded message is converted into its equivalent 4-bit binary then sampling it into 8-bit. 8 bit binary is converted into its equivalent decimal value.



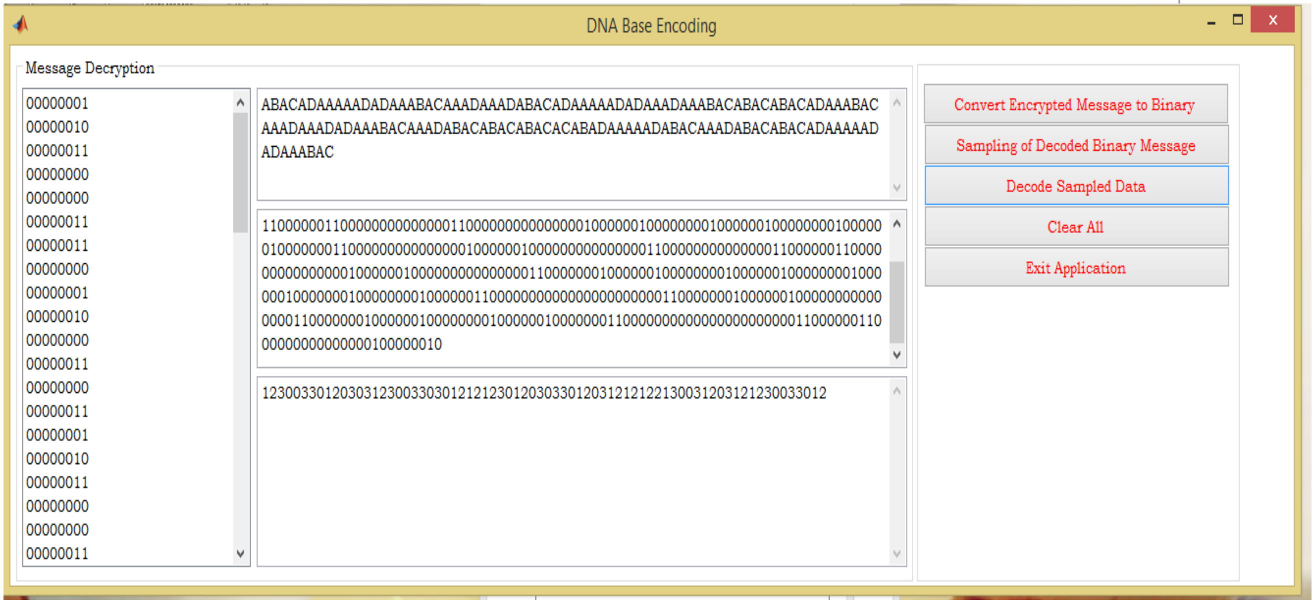
ADACAAABABAAAAABADACABAAABAAADACAAABABAAAAABAAABADACADACADACAAABADACABAAABAAAAABADACABAAADACADACADA
CACADAAABABAAADACABAAADACADACAAABABAAAAABADAC=3=AGGATCCTCTTCGAAGTCCTAGGACTTCGAAG

Original DNA String: GACTTCAGCTGATCAG

Rules: 4

Fake DNA String: GACTTCAGCTGATCAG

GA=A
CT=T
AG=G
TC=C



```
00000001
00000010
00000011
00000000
00000000
00000011
00000011
00000000
00000001
00000010
00000000
00000011
00000000
00000011
00000001
00000010
00000011
00000000
00000000
00000011
```

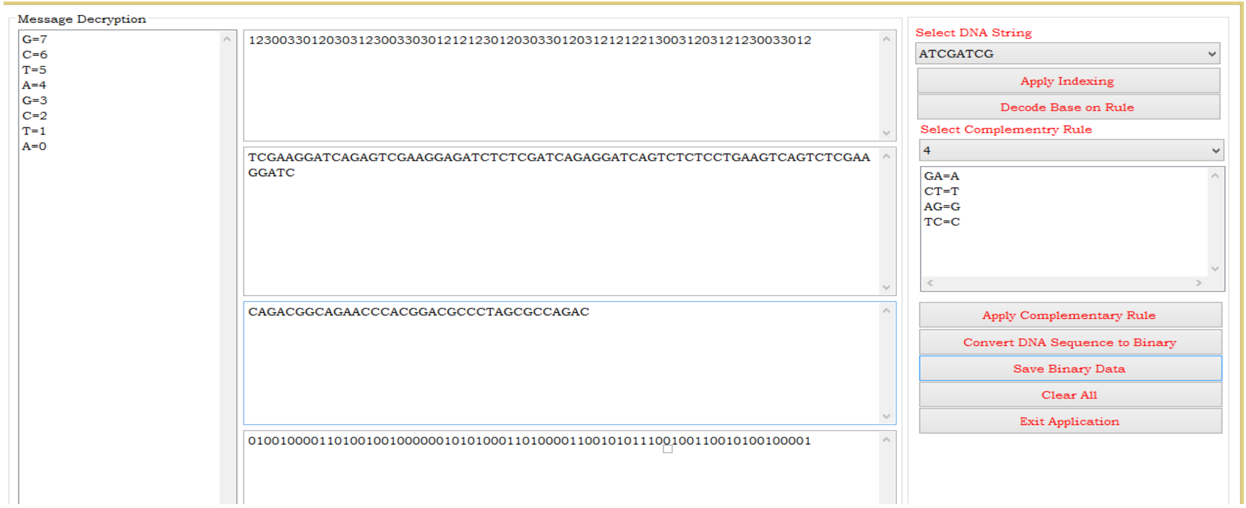
```
ABACADAAAAADADAAABACAAADAAADABACADAAAAADADAAADAAABACABACABACADAAABAC
AAADAAADADAAABACAAADABACABACABACACABADAAAAADABACAAADABACABACADAAAAAD
ADAAABAC
```

```
110000001100000000000000011000000000000000100000010000000010000001000000000100000
010000000110000000000000001000000100000000000000001100000000000000110000001100000
000000000001000000100000000000000001100000001000000100000000100000010000000001000
000100000000100000000100000011000000000000000000000001100000001000000100000000000
000011000000010000001000000000100000010000000011000000000000000000000011000000110
000000000000000100000010
```

```
123003301203031230033030121212301203033012031212122130031203121230033012
```

Convert Encrypted Message to Binary
Sampling of Decoded Binary Message
Decode Sampled Data
Clear All
Exit Application

- **Now apply indexing on DNA string then apply complementary rule on it. Converted it as per DNA code.**



Message Decryption

```
G=7
C=6
T=5
A=4
G=3
C=2
T=1
A=0
```

```
123003301203031230033030121212301203033012031212122130031203121230033012
```

```
TCGAAGGATCAGAGTCGAAGGAGATCTCTCGATCAGAGGATCAGTCTCTCCTGAAGTCAGTCTCGAA
GGATC
```

```
CAGACGGCAGAACCCACGGACGCCCTAGCGCCAGAC
```

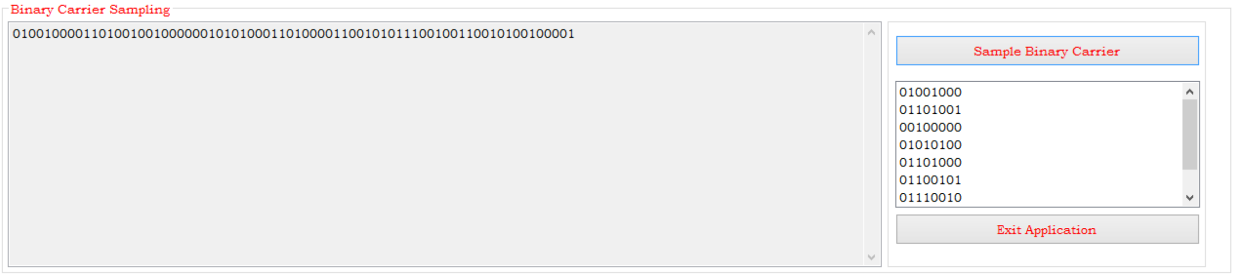
```
010010000110100100100000010101000110100001100101011100100110010100100001
```

Select DNA String: ATCGATCG

Select Complementry Rule: 4

```
GA=A
CT=T
AG=G
TC=C
```

- **DNA coded message sample into 8-bit.**



Binary Carrier Sampling

```
010010000110100100100000010101000110100001100101011100100110010100100001
```

```
01001000
01101001
00100000
01010100
01101000
01100101
01110010
```

CONCLUSION

The entire proposed algorithm has many steps to break and to get the original message. So, any intruder who receives the intermediate message will never be able to retrieve the original message as intended by the sender. The random selection of DNA sequence can be increased to many numbers. The complementary rules which are formed based on the properties of DNA could also be increased since DNA sequence has many biological properties and using those properties also some more complementary rules can be formed.

Message encryption using DNA sequence is a very new technique still evolving and tried out for secure transmission and reception of hided messages. The method is deemed to be so secure that it would be very difficult for any intrude to break the encrypted message and retrieve the actual message. Only the intended receiver can decrypt and receive the original message.

REFERENCES

1. K.Menaka—Message Encryption Using DNA Seqence‖ 978-1-4799-2977-4. 201 I EEE.
2. Debnath Bhattacharyya, Samir Kumar Bandyopadhyay, —Hiding Secret Data in DNA Sequence‖, International Journal of Scientific &Engineering Research Volume 4, Issue 2, February-2013 ISSN 2229-5518.
3. H. J. Shiu, K. L. Ng, J. F. Fang, R. C. T. Lee and C. H. Huang, —Data hiding methods based upon DNA sequences‖,Information of Science, vol.180, no.11, pp.2196-2208,2010.

4. Cheng Guo, Chin-Chen Chang and Zhi-Hui Wang ―A New Data Hiding Scheme Based On DNA Sequence‖ International Journal of Innovative Computing, Information and Control ICIC International Volume 8, Number 1(A), January 2012.
5. Mohammad Reza Abbasy, Azizah Abdul Manaf, and M.A.Shahidan, ―Data Hiding Method Based on DNA Basic Characteristics‖, International Conference on Digital Enterprise and Information Systems, July 20-22, (2011), London, UK,pp. 53–62.
6. Kritika Gupta* Shailendra Singh "DNA Based Cryptographic Techniques" International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 3, March 2013 ISSN: 2277 128X
7. Snehal Javheri, Rahul Kulkarni ―Secure Data communication and Cryptography based on DNA based Message Encoding‖ International Journal of Computer Applications (0975 – 8887) Volume 98– No.16, July 2014
8. Guangzhao Cui #1, Limin Qin #2, Yanfeng Wang #3, Xuncai Zhang *4 ―An Encryption Scheme Using DNA Technology‖ 978-14244-2724-6/08/2008 IEEE.
9. Ritu Gupta, Anchal Jain ―A New Image Encryption Algorithm based on DNA Approach‖ International Journal of Computer Applications (0975 – 8887) Volume 85 – No 18, January 2014