Application of DNA Computing in Security

Anita Chaware, Veerti Tirpude

Post Graduate Department of Computer Science, MCA

SNDT University, Juhu Tara Road, Santacruz (West)

Mumbai-400 049, Maharashtra, India.

anitadongarwar@gmail.com , veerti.tirpude1907@gmail.com

Abstract—
In this paper we indent to present the computing technology that has a great future – DNA COMPUTING. DNA Computing is in its initial stage and its implications are only beginning to be explored.

This paper begins with the brief description of DNA and its structure. It further introduces the various applications of DNA computing with its advantages and disadvantages. Further we applied the same to security and a simple algorithm of Encryption and Decryption **was developed. This concludes that DNA computing can very well used in Security .**

Keywords— **DNA, DNA Computing. DNA Cryptography, Encryption, Decryption**

I Introduction

A. DNA (Deoxyribonucleic Acid)

Deoxyribonucleic Acid (DNA) is a genetic material found in the cells of all living organisms. DNA is the fundamental building block for life. Nearly every cell (with nucleus) in a person's body has the DNA. Most DNA is located in the cell nucleus (where it is called nuclear DNA), but DNA can also be found in the mitochondria (where it is called Mitochondrial DNA).

The information in DNA is made up of four bases which combine to form chains. These bases include two Purines and two Pyrimidines. Purines are of two types Adenine and Guanine. Pyrimidines are of two types Cytosine and Thymine. They are commonly referred to as A, G, C and T respectively. Human DNA consists of about 3 billion bases, and more than 99 percent of these bases are the same in all people. It is the order, or sequence of these bases which determine genetic characteristics.

Following figure shows the molecular structure of Purines and Pyrimidines.

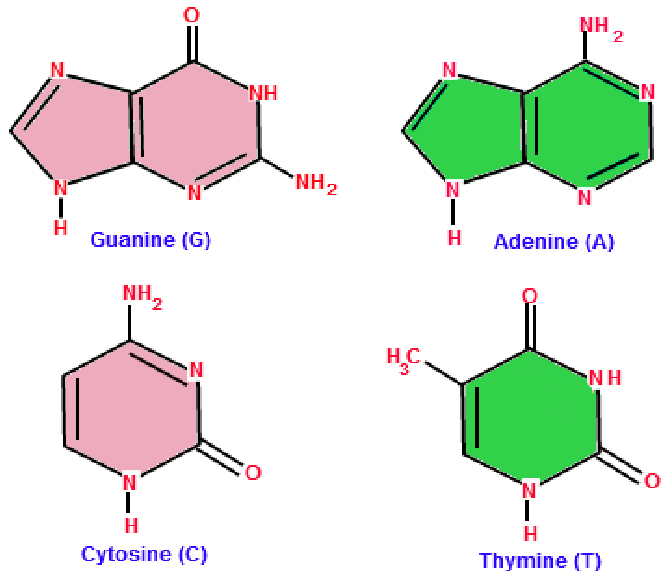

Fig. 1 Molecular Structure of Purines and Pyrimidines.

B. Structure of DNA

Each base is attached to a Sugar molecule and a Phosphate molecule. Together, a base, sugar, and phosphate are called a nucleotide. Nucleotides are arranged in two long strands that form a spiral called a double helix. The structure of the double helix is somewhat like a ladder, with the base pairs forming the ladder's rungs and the sugar and phosphate molecules forming the vertical sidepieces of the ladder.

The number of purine bases in DNA is equal to the number of pyrimidines. This is due to the law of complimentary base pairing; which is Thymine (T) can only pair with Adenine (A), and Guanine (G) can only pair with Cytosine (C). Knowing this rule, we could predict the base sequence of one DNA strand if we knew the sequence of bases in the complimentary strand.

TABLE I PURINE-PRIMIDINES BASE PAIR

Purines	**Pyrimidines**
Adenine (A)	Thymine (T)
Guanine (G)	Cytosine (C)

© 2013 IJAIR. ALL RIGHTS RESERVED

Following figure shows the structure DNA.

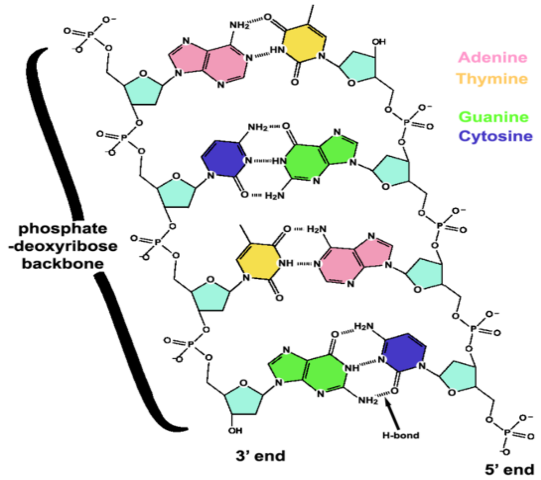

Fig. 2 Structure of DNA.

1) Linking of Nucleotides:

The DNA Nucleotides can be link in two ways:
1. Phosphodiester Bond
2. Hydrogen Bond

2) Phosphodiester Bond:

The 5' -phosphate group of one nucleotide is joined with the 3' -hydroxyl group of the other
- Strong (Covalent) bond
- Direction: 5' --- 3' or 3' ---5'

3) Hydrogen Bond:

The base of one nucleotide interacts with the base of another
- Weak Bond

4) Watson and Crick Model of DNA:

James D. Watson and Francis H. C. Crick deduced double-helix structure of DNA in 1953 and got Nobel Prize in 1962.
Adenine and Guanine always link together. Cytosine and Thymine always link together.

Watson and Crick discovered that DNA had two strands, and these strands are twisted together like a twisted ladder – the double helix. The sides of the ladder comprise the sugar-phosphate portions of adjacent nucleotides bonded together. The phosphate of one nucleotide is covalently bound (a bond in which one or more pairs of electrons are shared by two atoms) to the sugar of the next nucleotide. The hydrogen bonds between phosphates cause the DNA strand to twist. The nitrogenous bases point inward on the ladder and form pairs with bases on the other side, like rungs. Each base pair is formed from two complementary nucleotides (purine with pyrimidine) bound together by hydrogen bonds. The base pairs in DNA are adenine with thymine and cytosine with guanine.

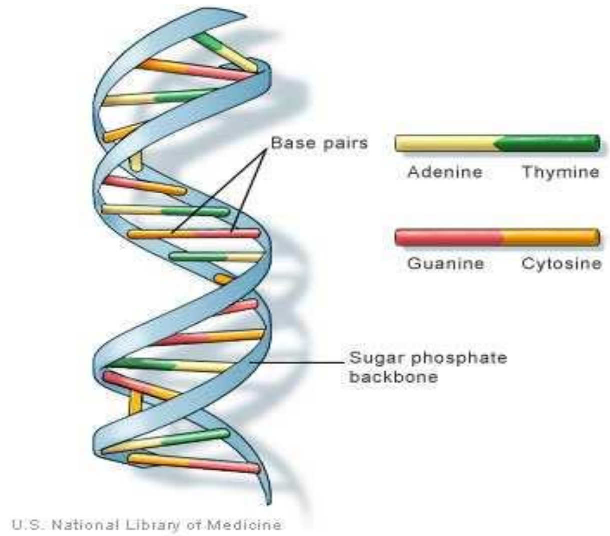

Fig. 3 DNA has a spiral staircase-like structure. The steps are formed by the nitrogen bases of the nucleotides where Adenine pairs with Thymine and Cytosine with Guanine. (Photo courtesy U.S. National Library of Medicine)

Design information is transmitted as new DNA to new cells during development and growth. The complementarity of the two DNA strands allows their information to be copied. Each old strand is used as a template in synthesizing a new complementary one.
These complimentary strands have codons as fundamental building blocks. Codons are basically triplets of nucleotide bases. Figure below shows codons forming DNA sequences in two complimentary strands.
A G C - T C T - A A G - T C C – C A A
T C G - A G A - T T C - A G G - G T T

As can be seen DNA nucleotide bases are existing in form of codons and are complimentary to each other that is A-T, G- C are complimentary to each other. We can use these codons for encoding and decoding of the data.

II. DNA COMPUTING

DNA computing is a form of computing which uses DNA, biochemistry and molecular biology, instead of the traditional silicon-based computer technologies. DNA computing, or, more generally, bio-molecular computing, is a fast developing interdisciplinary area. Research and development in this area concerns theory, experiments, and applications of DNA computing [1].

III. DNA CRYPTOGRAPHY

The DNA cryptography is a new and very promising direction in cryptography research. DNA can be used in cryptography for encoding and decoding of the information, as well as for computation. This field was initially developed by Leonard Adleman of the University of Southern California, in 1994. Adleman demonstrated a proof-of-concept use of

© 2013 IJAIR. ALL RIGHTS RESERVED

DNA as a form of computation which solved the seven-point Hamiltonian path problem also known as travelling salesman problem. Since the initial Adleman experiments, advances have been made and various Turing machines have been proven to be constructible.

IV. DNA APPLICATION

DNA computing is being used in various areas. This new emerging field is not limited to one particular field of study. Following are the few areas where DNA Computing has put its mark:

1. It is used in 2000 Sydney Olympics for Authentication of the all the products provided by the hosting country.
2. Because of the vast storage capacity of DNA molecule it is being used for storing large amount of data.
3. DNA Ink Application has been used in Steganography.
4. The use of DNA for doing Database operation is the process, scientist are still working on it for achieving the success.

As we have prior said this technology is still in the process of development, we are trying to use it in the area of Cryptography. We have designed a small easy to use and understand algorithm where a plain text is getting converted into the Cipher text which is the sequence of DNA codons.

Using DNA Cryptography can also be used in combination with other algorithm which will add an additional level of security in the process of decryption.

The major advantage of this could be for every encrypted plain text we would be having two sets of DNA cipher text, this is because of the Complementary nature of DNA Sturucture.

V. Mechanism Designed

Here we have designed an algorithm which makes use of DNA Computing to encrypt the given text. In this algorithm the alphabets are represented by the sequence of DNA codons. The following table shows the one of such DNA codon mapping table:

TABLE II CODON MAPPING TABLE

ASCII	Letter	Codon Pair	ASCII	Letter	Codon Pair
65	A	CGT	78	N	TTT
66	B	ATC	79	O	CGG
67	C	TGC	80	P	CGC
68	D	TCC	81	Q	CTG
69	E	GCC	82	R	TGA
70	F	CAG	83	S	GCA
71	G	ATC	84	T	GCC
72	H	TAA	85	U	GTC
73	I	AGT	86	V	CCA
74	J	CGA	87	W	ACC
75	K	GAA	88	X	TCG
76	L	GGT	89	Y	AAA
77	M	CCT	90	Z	ATG

Fig. 4 Codon Mapping Table

The Table is generated by the randomly choosing 26 codon pair from the available 64 codon pair. They are chosen by using a seed value.

A. Algorithm Flow Chart

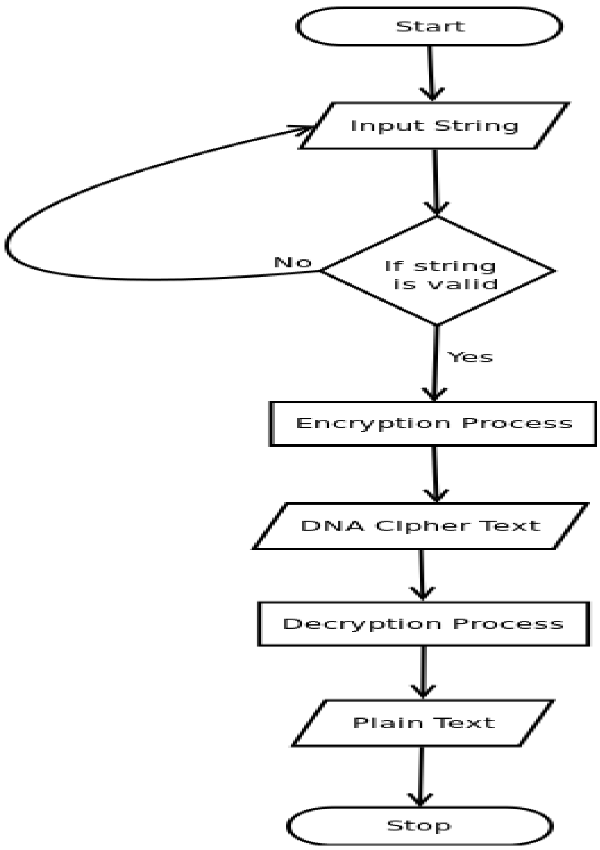

Fig. 5 Flow Chart

B. Encryption Process Steps

1. Enter letter to be encrypted.
2. Map the corresponding DNA pair for the letter from the mapping table.
3. Generate the ASCII value for each letter of the DNA pair.
4. Generate corresponding binary value for it.

© 2013 IJAIR. ALL RIGHTS RESERVED

5. Reverse the 8 digit binary number.
6. Take a 4 digit divisor (1000) as a key.
7. Divide the reversed number with the divisor.
8. Store the remainder in first 3 digits and quotient in the next 5 digit.
9. How do the binary to decimal conversion i.e. you will get the ASCII value and then get the corresponding letter.
10. Map the letter in the mapping table to get the corresponding DNA.

C. Decryption Process Steps

1. Received cipher text
2. Map the cipher with the mapping table.
3. Covert the result into the binary.
4. Multiply the last 5 bits of the binary with the key.
5. Add the first 3 bits to the result of step 3.
6. Reverse the bits which are obtained from the step
7. Perform the binary to decimal conversion.
8. Map it to the ASCII value and get the corresponding letter.
9. After above all steps are done map the DNA pair to the mapping table to get the corresponding letter (plain text).

D. Results

For carrying out implementation of the cryptographic mechanism designed I have used Java1.7.0_21. DNA Cipher which is designed here relies on the Mapping table. The plain text after undergoing through the encryption and the decryption process has been successfully recovered into its original plain text. Following figure shows the output of the implementation:

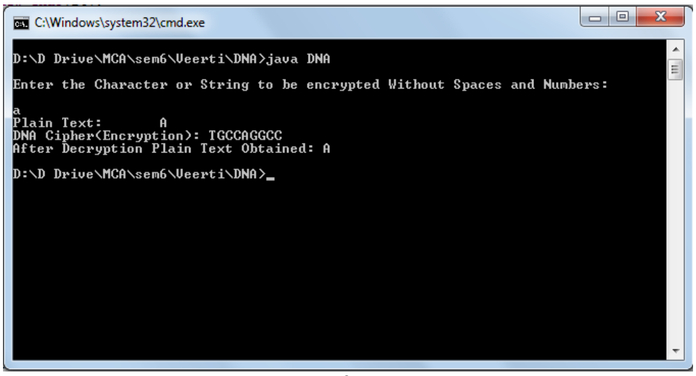

Fig. 6 Output of Computation

E. Strength of this algorithm

1. Every letter is represented by series of DNA codon pair.
2. Every letter is represented by 3 codons, which make it little difficult to crack.

F. Advantages of DNA Computing

1. Perform millions of operations simultaneously (Parallel Computing).
2. Over 100 times faster than fastest computer.
3. Minimal Storage and Power requirement.

G. Disadvantages of DNA Computing

1. DNA Computers could not replace (at this point) traditional computer.
2. They are not programmable and the average dunce cannot sit down at a familiar keyboard and get to work.
3. It requires human assistance.

VI. CONCLUSION

The field of DNA computing is still in its preliminary stage and the applications for this technology are still not fully understood.
Due to the development and advance of science, the possibilities are rapidly growing. DNA in connection to cryptography is a fast developing interdisciplinary area.
DNA cryptography is the future of the information security. Its complexity and randomness provides a great uncertainty which makes encoding of data in DNA format better than other mechanism of cryptography.

VII. Future work

1. Presently using this algorithm only text i.e. alphabets can only be encoded which can be extended to numbers and special characters too.
2. Currently the 4 bit divisor being used is fixed but it can be change in future.
3. Presently the alphabet ‘G’ cannot be decoded by following the steps of the algorithm but it can be done in future.

Example:
Encryption

1. Input :- A
2. Map DNA codon pair corresponding to A:-CGT
3. Consider the first letter of the codon i.e. C find the ASCII value of C i.e. 67
4. Change the decimal value 67 to binary i.e. 01000011
5. Reverse the binary 11000010
6. Divide the reverse number by the 4bit divisor 1000
7. Remainder: 010, Quotient: 11000
8. Convert this to decimal i.e. 01011000 which 88 i.e. ‘X’
9. Map DNA Codon pair corresponding to ‘X’.
10. Follow the same steps for ‘G’ and ‘T’.
11. Therefore DNA cipher for plain text A will be ‘**TGCCAGGCC**’.

Decryption

1. Received cipher TGCCAGGCC
2. Map the DNA Codon table to get the corresponding codon: TGC->X, CAG->F, GCC->E.
3. Get ASCII value of X i.e. 88 convert it into binary 01011000
4. Get the last 5 bits of the binary and multiply it with the 4 bit key.
5. Add 1st 3 bit to the result of the above step.
6. Reverse the result obtained and convert it into decimal.
7. Follow the same steps for the remaining sequence and you will get CGT
8. Map the result into the DNA codon look up table and you will get the plain text i.e. 'A'.

References

http://en.wikipedia.org/wiki/DNA_computing [1]

REFERENCES

[1] Atul Kahate ,"Cryptography and Network Security", Tata Macgraw Hill,2009

[2] Xing Wang, Qiang Zhang ,"DNA computing-based cryptography", Fourth International Conference on BioInspired Computing, 2009.

[3] L.M. Adleman, Molecular computation of solutions to combinational problems. Science, (1994), 266(4), pp.1021-1025.

[4] Pankaj Raheja "Integrating DNA Computing in International Data Encryption Algorithm (IDEA)" in *International Journal of Computer Applications (0975 – 8887)Volume 26– No.3, July 2011.*

[5] Piper," Basic principles of cryptography" , IEEE Colloquium on Public Uses of Cryptography, 1996

[6] Ayushi "Symmetric Key Cryptography Algorithm" in 2010 International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 15.

[7] Computing with DNA *The manipulation of DNA to solve mathematical problems is redefining what is meant by "computation"* by Leonard M. Adleman

[8] Guangzhao Cui, Cuiling Li, Haobin Li, Xiaoguang Li "DNA Computing and Its Application to Information Security Field" in 2009 Fifth International Conference on Natural Computation.

[9] Herbert Schildt " The Complete Reference java" Seventh Edition. 2006

© 2013 IJAIR. ALL RIGHTS RESERVED