IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 11, Issue 4 (May. - Jun. 2013), PP 44-47
www.iosrjournals.org

Crypt Sequence DNA

Mrs.S.Sujatha MCA, MPhil.,(Ph.D.),Bhadra Prabhakaran
School of IT and Science,Dr.G.R. Damodaran College of Science, India

Abstract: *The DNA sequence plays a major role in identifying each individual and making them unique. So we try incorporating DNA to encrypt the data in exclusively identifying individual encryption format. This method would highlight very basic and easy algorithms to form exclusive cipher text for each data or file. Each of the DNA component would be allocated a fixed algorithm such that the encryption would be based on the components sequence without altering the algorithms fixed.*

***Keywords** -Cipher text, Component sequence DNA, DNA Component and Encrypt*

I. INTRODUCTION

Securing data is the massive issue in today's world. Human beings are the source of data. The data obtained from or created by them is always insecure in the way of storage. The data storage was made secured by cryptography principles.Data in any format or the source has the knack to be publicized due to the advancement in the cryptanalysis. This work involves another security measure of data. Any form of cryptography for a certain amount of data would have same style of encryption throughout the entire domain. This work proposes an individual format for every data present in the data domain. The individuality needed a specific itinerary and that was found as the DNA of the individual. DNA is a unique composition of molecules for every human who is the source of data. The sequence of DNA is the key to the complete encryption method.

II. LITERATURE REVIEW

DNA is used only for the encrypted storage of data in a tiny memory according to the researchers A new method of data storage that converts information into DNA sequences allows you to store the contents of an entire computer hard-drive on a gram's worth of E. coli bacteria[1]. The DNA composition sequence is faked with the encrypted sequence of data as the researchers showed how to change the word "iGEM" into DNA-ready code. They used the ASCII table to convert each of the individual letters into a numerical value [2]. Binary data is encoded in the geometry of DNA nanostructures with two distinct conformations. Removing or leaving out a single component reduces these structures to an encrypted solution [3].Alteration of DNA sequence using one time pad and adding the encrypted message into the original DNA sequence strand.[4]The keys need to be binary string and the image file input has to be converted to binary data[5].DNA based Cryptography which puts an argument forward that the high level computational ability and incredibly compact information storage media of DNA computing has the possibility of DNA based cryptography based on one time pads[6]. An encryption scheme is designed by using the technologies of DNA synthesis, PCR amplification and DNA digital coding as well as the theory of traditional cryptography [7].

III. PROBLEM DEFINITION

The database is still not secure with the encrypted form of data.The keys or even the plain texts are identified by cryptanalysis like Brute Force. There is no specific individual encryption format for every specific data that is stored in the database.The individual encryption is not possible in huge database. Every data cannot be encrypted individually. This is the reason why hacking becomes stress-free all times and if one part of the encryption algorithm is acknowledged then the whole database can be scythed.

IV. PROPOSED WORK

First step in this search, Chargaff set out to see whether there were any differences in DNA among different species. After developing a new paper chromatography method for separating and identifying small amounts of organic material, Chargaff reached two major conclusions (Chargaff, 1950). First, he noted that the nucleotide composition of DNA varies among species. In other words, the same nucleotides do not repeat in the same order, as proposed by Levene. Second, Chargaff concluded that almost all DNA--no matter what organism or tissue type it comes from--maintains certain properties, even as its composition varies. In particular, the amount of adenine (A) is usually similar to the amount of thymine (T), and the amount of guanine (G) usually approximates the amount of cytosine (C). In other words, the total amount of purines (A + G) and the total amount of pyrimidines (C + T) are usually nearly equal. (This second major conclusion is now known as "Chargaff's rule.") Chargaff's research was vital to the later work of Watson and Crick, but Chargaff himself could not imagine the explanation of these relationships--specifically, that A bound to T and C bound to G within the molecular structure of DNA. Chargaff's realization that A = T and C = G, combined with some crucially important X-ray crystallography work by English researchers Rosalind Franklin and Maurice Wilkins, contributed to Watson and Crick's derivation of the three-dimensional, double-helical model for the structure of DNA. Watson and Crick's discovery was also made possible by recent advances in model building, or the assembly of possible three-dimensional structures based upon known molecular distances and bond angles, a technique advanced by American biochemist Linus Pauling. In fact, Watson and Crick were worried that they would be "scooped" by Pauling, who proposed a different model for the three-dimensional structure of DNA just months before they did. In the end, however, Pauling's prediction was incorrect.

Using cardboard cutouts representing the individual chemical components of the four bases and other nucleotide subunits, Watson and Crick shifted molecules around on their desktops, as though putting together a puzzle. They were misled for a while by an erroneous understanding of how the different elements in thymine and guanine (specifically, the carbon, nitrogen, hydrogen, and oxygen rings) were configured. Only upon the suggestion of American scientist Jerry Donohue did Watson decide to make new cardboard cutouts of the two bases, to see if perhaps a different atomic configuration would make a difference. It did. Not only did the complementary bases now fit together perfectly (i.e., A with T and C with G), with each pair held together by hydrogen bonds, but the structure also reflected Chargaff's rule.

Deoxyribonucleic acid (**DNA**) is a molecule that encodes the genetic instructions used in the development and functioning of all known living organisms and many viruses. Along with RNA and proteins, DNA is one of the three major macromolecules essential for all known forms of life. Genetic information is encoded as a sequence of nucleotides (guanine, adenine, thymine, and cytosine) recorded using the letters G, A, T, and C. Most DNA molecules are double-stranded helices, consisting of two long polymers of simple units called nucleotides, molecules with backbones made of alternating sugars (deoxyribose) and phosphate groups (related to phosphoric acid), with the nucleobases (G, A, T, C) attached to the sugars. DNA is well-suited for biological information storage, since the DNA backbone is resistant to cleavage and the double-stranded structure provides the molecule with a built-in duplicate of the encoded information.

These two strands run in opposite directions to each other and are therefore anti-parallel, one backbone being 3' (three prime) and the other 5' (five prime). This refers to the direction the 3rd and 5th carbon on the sugar molecule is facing. Attached to each sugar is one of four types of molecules called nucleobases (informally, *bases*). It is the sequence of these four nucleobases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of the amino acids within proteins. The code is read by copying stretches of DNA into the related nucleic acid RNA in a process called transcription.

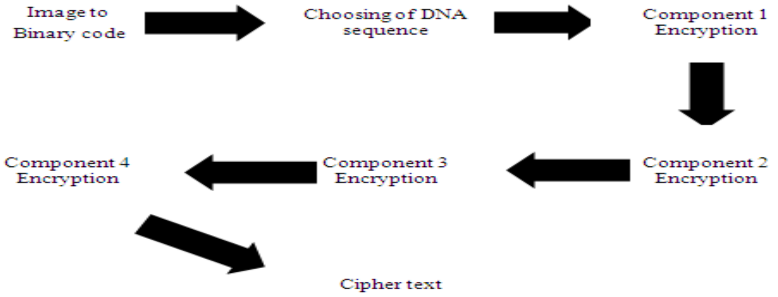
These four compositions can be provided with specific fixed algorithms and the encryption sequence would be based on the DNA composition sequence.The algorithm would be fixed for each individual component. The encryption sequence would be based on the sequence of the components present in the individual’s DNA. This would lead to reduction of complicate algorithms and also a lot of keys to be remembered. This makes data so secure that even if the keys are identified the plain text could not be retrieved.

The DNA composition consists of four major components

- Adenine
- Thymine
- Guanine
- Cytosine

These four compositions can be provided with specific fixed algorithms and the encryption sequence would be based on the DNA composition sequence.The algorithm would be fixed for each individual component. The encryption sequence would be based on the sequence of the components present in the individual’s DNA. This would lead to reduction of complicate algorithms and also a lot of keys to be remembered. This makes data so secure that even if the keys are identified the plain text could not be retrieved.

V. METHODOLOGY



VI. Algorithms Chosen And Their Compatibility

	AES	Triple DES	Blow Fish	Serpent
Derived from	Square	DES	-	Square
Key sizes	128, 192 or 256 bits	168, 112 or 56 bits	32–448 bits	128, 192 or 256 bits
Block sizes	128 bits	64 bits	64 bits	128 bits
Structure	Substitution-permutation network	Feistel network	Feistel network	Substitution-permutation network
Rounds	10, 12 or 14 (depending on key size)	48 DES-equivalent rounds	16	32
cryptanalysis	Brute Force, Linear, Boomerang and Differential	Brute Force, Linear, Boomerang and Differential	Brute Force, Linear, Boomerang and Differential	Brute Force, Linear, Boomerang and Differential
Security	All known attacks are computationally infeasible. For AES-128, the key can be recovered with a computational complexity of $2^{126.1}$ using bicliques. For biclique attacks on AES-192 and AES-256, the computational complexities of $2^{189.7}$ and $2^{254.4}$ respectively apply.	The original DES cipher's key size of 56 bits was sufficient but the computational power made brute-force attacks feasible. Triple DES provides a method of increasing the key size of DES making it stronger without changing cipher blocks	Four rounds of Blowfish are susceptible to a second-order differential attack but when the keys are still stronger then it becomes strong enough	All known attacks are computationally infeasible. An attack in 2011 could break only till 12 Round of the Serpent decryption

VII WORKING PRINCIPLE

1. The image file is converted to binary string and keys are generated
2. DNA sequence is chosen from the complete sequence
3. Encryption of data according to the DNA component sequence chosen
4. Adenine- AES Algorithm
5. Thymine- Triple DES
6. Cytosine- Serpent Algorithm
7. Guanine- Blowfish Algorithm

VIII PERFORMANCE

- Produces a very complex cipher code such that it is hard to break
- The algorithms are unique in their keys and also their encryption format which makes cryptanalysis more stressful
- Cipher code can be identified or cracked only when the DNA random generated sequence components are found with the keys and the algorithms allocated to them, which would be really hectic.

IX CONCLUSION

This method would help in encrypting the data individually though all the data belong to the same database. The common encryption format for a database can be totally eradicated. The security would be at a very high stage. Government confidential material can be at a higher rate of confidentiality and in very simple steps.

REFERENCES

Journal Papers:

1)Bio encryption can store almost a million gigabytes of data inside bacteria-*IO9 Journal Vol 2 by Alasdair Wilkins* on Nov 26 2010 2)Binary DNA Nanostructures for Data Encryption- *PLOS ONE* Published: September 11, 2012 by *Ken Halvorsen, Wesley P. Wong* 3)DNA-Based Data Encryption and Hiding Using Play fair and Insertion Techniques- *Journal of Communications and Computer Engineering Vol2, 3 (2012) by A Atito* 4)DNA-based Cryptography -DNA Based Computers- *June 1999 by Ashish Gehani, Thomas H. LaBean and John H. Reif* 7) An encryption scheme using DNA technology Guangzhao Cui Coll. of Electr. Inf. Eng., *Zhengzhou Univ. of Light Ind., Oct. 1 2008,*

Books:

Advanced Encryption standard Algorithm by VincentRijmen, Joan Daemen, Triple Data Encryption Standard, Serpent Algorithm by Ross Anderson, Eli Biham, Lars Knudsen, Blow Fish Algorithm, Text book on Identity-Based Encryption By Chatterjee Sanjit, SarkarPalash, A text book A Novel DNA based Encrypted Text Compression By D. Prabhu, M. Adimoolam, P.Saravannan, A text book DNA Structure and Function By Richard R Siden,

Articles:

ARTICLE ON INNOVATIVE FIELD OF CRYPTOGRAPHY: DNA CRYPTOGRAPHY BY ERSONI – 2012 ER.RANUSONI, ER.VISHAKHASONI AND ER.SANDEEPKUMARMATHARIYA, ARTICLE ON STABILIZING SYNTHETIC DATA IN THE DNA OF LIVING ORGANISMS SYST SYNTH BIOL. 2008 JUNE ,NOZOMUYACHIE, YOSHIAKIOHASHI, AND MASARUTOMITA