ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 8, August 2016

Data Security Based on DNA Cryptography and Steganography

[1]Ms. Madhura W. Mangle,[2]Ms. Rashmi B. Bagdiya,[3]Ms. Samruddhi S. Rathi,[4]Mr. Ankush J. Ghodeswar, [5]Ms. Poonam B. Lohiya
[1][2][3][4][5]Computer Science and Engineering Department
[1][2][3][4][5]Prof. Ram Meghe Institute of Technology and Research, Badnera-Amravati

***Abstract*—Data hiding is the skill of hiding messages such that only the sender and the receiver of the message knows the message is hidden. In the world of security efficient techniques for data encryption and decryption are developing eventually. Though many algorithms have been developed for hiding the data, DNA sequences that are based on data encryption seems to be a promising approach for fulfilling the current information security needs. Similarly, the DNA sequences are we have used are with double encryption. This is achieved using cryptography and steganography. In this paper, an algorithm is used DNA sequences for data hiding is proposed and discussed for secure data transmission and reception. In this project, we are trying to create a cipher text using the DNA sequences consisting of nucleotides A (Adenine), G (Guanine), C (Cytosine), and T (Thymine) with the help of cryptography methods and we have embedded the cipher text in an image by using steganography.**

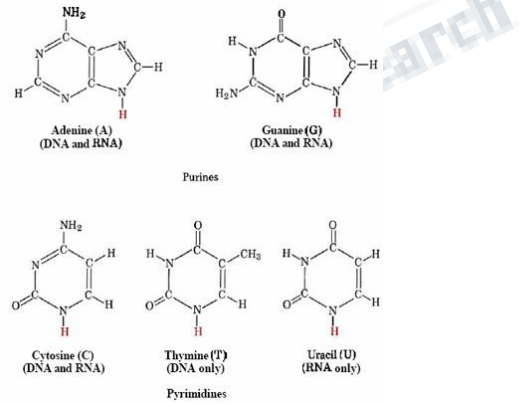
Keywords—**DNA; DNA Computing; Cryptography; Steganography.**

I. INTRODUCTION

Information Security is extremely important in today's digital era. The year 2014 is also called as "The Year of Data Breach", and also 2016 promises several major data breaches. As per statistics gathered by Breach Level Index, there were more than 2 million records per day that were breached in the year 2014, which means, 32 records were breached each second."It's apparent that a new approach to data security is needed, if organizations are to stay ahead of the attackers and more effectively protect their intellectual property, data, customer information, employees."Security approaches must be good enough to tackle the ever-changing data breaches. This is where the data security, encryption of data at rest and in motion, implementing user access control come into play.

a. Introduction to DNA:

DNA is molecular structure present in the human body, which is also called as Deoxyribonucleic acid used to carry the genetic information. DNA consists of nucleotides Adenine, Guanine, Cytosine, and Thymine. Their molecular structures are as follows:



b. DNA Computing:

DNA computing is a branch of computing which uses DNA, biochemistry, and molecular biology hardware, instead of the traditional silicon-based computer technologies. Research and development in this area concerns theory, experiments, and applications of DNA computing. The term "molectronics" has sometimes been used, but this term had already been used for an earlier technology, a then-unsuccessful rival of the first integrated circuits; this term has also been used more generally, for molecular-scale electronic technology.

DNA computing is a form of parallel computing in that it takes advantage of the many different molecules of DNA to try many different possibilities at once. For certain

ISSN (Online) 2394-2320

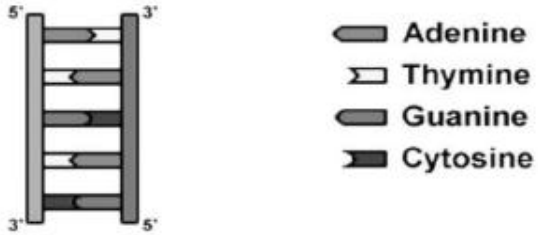
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 8, August 2016

specialized problems, DNA computers are faster and smaller than any other computer built so far. Furthermore, particular mathematical computations have been demonstrated to work on a DNA computer. As an example, DNA molecules have been utilized to tackle the assignment problem.

c. DNA Cryptography:

DNA cryptography is the rapid emerging technology which works on concepts of DNA computing. A new technique for securing data was started using the biological structure of DNA called DNA Computing (aka molecular computing or biological computing). In the year 1994, it was invented by Leonard Max Adelman for solving the difficult problems such as directed Hamilton path problem, NP - complete problem which are similar to The Travelling Salesman problem. Adelman is also known as the 'A' in the RSA algorithm,an algorithm that in some circles has become the standard for industrial-strength encryption of data sent over the Web. This technique was later on extend by various researchers for encryption and minimizing purposes of the storage size of data that made the data transmission over the network more faster and secured.

DNA is used to store and transmit data. The concept of using DNA computing in the area of cryptography and steganography has been identified as a potential technology that may bring forward a new hope for unbreakable algorithms. DNA strands are polymers of millions of linked nucleotides. These nucleotides consist of one of four nitrogen bases, a five carbon sugar and a phosphate group. The polymers are made up of different nucleotides named after the nitrogen bases that it consists of : Adenine, Cytosine, Guanine and Thymine.



d. Steganography:

Steganography is the practice of hiding the private or sensitive information or messages within something that appears to be nothing unusual. Steganography is generally befuddled with cryptology as both of them are similar in such a way that they are used to protect important information. The difference is that, steganography involves hiding information so it appears that no information is hidden at all. There are no traces of hidden message. For example, If a person views the object that the information is hidden inside of, will have no clue that there is any hidden information, hence the person will not try to decrypt the information. What steganography fundamentally do is, it exploits human perception, the human senses which are not trained to look for files that have information inside of them, although the software is available, that can do Steganography. The most common use of steganography is to hide a file inside another file. The main goal of this project is it to communicate securely in an undetectable manner and to avoid drawing attention to even the transmission of a hidden data.

vessel image

After entering the confidential message

steg image

II. LITERATURE REVIEW

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 8, August 2016

In 2008, Guangzhao Cui et al., [7] proposed the public key encryption technique that uses DNA synthesis, DNA digital coding and PCR amplification to provide the security safeguard during the communication. This encryption scheme has high confidential strength.

In 2008, Lai Xin-she et al., [8] proposed A novel generation key scheme based on DNA using key expansion matrix. They used random key generation scheme to increase computational speed. In this algorithm author used block cipher, data signature, identity authentication, DNA sequences, which randomized database.

In 2010, Lai Xuejia et al., [9] proposed Asymmetric encryption and signature method with DNA technology. This paper proposed DNA public key cryptosystem, an asymmetric encryption and signature cryptosystem. DNA (PKC) uses encryption and signature. Key and cipher-text is biological molecule in DNA (PKC). In DNA (PKC) key and cipher-text are transmitted physically and it's difficult to replicate. DNA public key cryptosystem is based on DNA microarray chip. It is fabricated with probes for encryption and decryption. Existing probes are used as a key. If the probes have intensity greater than some fixed value then it is denoted as probe 1 and if the probes have intensity lesser than some fixed value, it is denoted as probe 0. For encryption process two key are used PKs and PKr. First plaintext converted into its ASCII code and then it converted into binary code, binary code is arranged in the form of matrix.

In 2011, Deepak Kumar et al., [10] proposed a new secret data writing techniques based on DNA sequences. They have explained about one-time-pad (OTP) technique for secure data transformation and DNA coding technique. They used cryptography and steganography technique for encryption and hiding data.

In 2011, Bibhash Roy et al., [11] proposed an improved symmetric key cryptography with DNA based strong cipher. Author focused on DNA computational logic, used for encrypting, storing and transmitting the data. This paper proposed about the unique cipher-text procedure and key generation procedure. Author discussed only about DNA cryptography and DNA computing.

In 2012, Yunpeng Zhang et al., [12] proposed a DNA cryptography based on DNA fragment assembly. Author mentioned features of DNA molecular, key bio technologies, DNA digital coding and related software. Using the DNA digital coding and DNA fragmentation author designed symmetric system algorithm, it converted plaintext into binary ASCII code and then into DNA sequences.

In 2013, Wang Zhong et al., [13] proposed an Index based DNA encryption algorithm. They used Block cipher and Index of string for encrypting message into DNA sequences, which is send to the receiver by a secure communication medium. First message converted into ASCII code then converted into binary code, which is further converted into DNA sequence. DNA sequence search in the key sequence and writes in index number.

In 2014 K. Menaka et al., [14] proposed Message Encryption Using DNA Sequences. They described about data hiding based on DNA sequence. DNA sequences based data encryption algorithm fulfilling the current information security needs. In this paper, an algorithm using DNA sequences for data hiding is proposed and discussed for secure data transmission and reception.

III. METHODOLOGY

In this paper, we are using two encryption techniques i.e. Cryptography and Steganography. So there are two different methods for the implementation. We have divided the implementation into two parts:

A IMPLEMENTATION OF CRYPTOGRAPHY

In this paper we have , based on the chemical structures divided the four nucleotides A,C,G,T into two classes:

Purine R={A,G} and Pyrimidine Y= {C,T}

Or

Amino Group N = {A,C} and Keto Group K={G,T}.

In addition to these, the division can also be made according to their strength of the hydrogen bonds
i.e. **Weak H-bonds W = {A, T} and Strong H-bonds S = {G, C}**

as suggested by Ing-an He et al. [5].
Using these properties of DNA sequences, three complementary rules are formed and later can be used to generate fake DNA sequences. And they are as follows:

Complementary Rule 1: (AG) (GA) (CT) (TC) _ based on Purines and Pyrimidines.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 8, August 2016

Complementary Rule 2: (AC) (CA) (GT) (TG) _ based on Amino and Keto groups.

Complementary Rule 3: (AT) (TA) (GC) (CG) _based on Strong and Weak bonds.

There is also one special property for DNA sequences i.e. the real DNA sequence and the faked DNA sequence that will almost look like the same. And, there are a large number of DNA databases which are publicly available. By using these facts, in this paper, a new methodology is formed for encrypting messages using DNA sequences. DNA sequences propose a exclusive method for encrypting messages or information. The main advantage of DNA sequences is,that they are composed of letters which are meaningless for most people. The DNA sequence is a combination of A, C, G and T base pairs. In the proposed algorithm, the complementary rules are formed based on the properties like, Purine, Pyrimidine, Amino, Keto; Strong, Weak bonds. Purines and Pyrimidine are two of the building blocks of nucleic acids. The methodology can be implemented as follows:

In this proposed method, the algorithm first randomly selects a DNA sequence for example, TAGCATGACT, this sequence is called as message index. Each letter is then given a subscript index starting from 0. Message index is the first positional index value of the DNA sequence. As we proceed to the next step, any complementary rule from the above rules is selected. As per the

T0A1G2C3A4T5G6A7C8T9
(AG) (GA) (CT) (TC)
(Randomly Selected DNA string) (Randomly selected complementary rule: Rule1)

A CGATGCAGTC
(Faked DNA sequence obtained after applying complementary rule along with an 'A' as the MSB for selecting the 1st complementary rule)

CGATGCAGTC
MS75
(Message to be encoded)

01001101 01010011 00000111 00000101
(8 bit binary coded message)

CATC CCAT AACA AACG TGCT TTGC GGTG GGTA
(DNA coded message as per Table 1)
(Converting as per complementary rule1)

0230 0023 2202 2201
(After applying Message Index)

000 010 011 000 000 000 010 011 010 010 000 010 010 010 000 001
(Three digit Binary Equivalent)

ACDAAACDCCACCCAB
(Message received at the receiver side – corresponding alphabet value applied)

Fig: Sender's side

ACDAAACDCCACCCAB and
A CGATGCAGTC
(Message received from the sender)
(Faked DNA sequence where MSB tells the complementary rule applied)

000 010 011 000 000 000 010 011 010 010 000010 010 010 000 001
(Converted to equivalent three digit Binary value for the
alphabet values)

0230 0023 2202 2201
(Converted to decimal form)

TGCT TTGC GGTG GGTA
(Converted to DNA coded message by using the index value – only the correct recipient knows the indexed DNA
sequence and the complementary rule applied)

CATC CCAT AACA AACG
(Applying complementary rule 1)

01001101 01010011 00000111 00000101
(Corresponding Binary coded value – 2 bits for each letter)

MS75
(Original Message received)
183

Fig: Receiver's side

Algorithm, a single letter is replaced with a specific letter as as Defined by the complementary rule.
For example, if the complementary rule 1 is selected and the above DNA sequence is taken, then the result will be CGATGCAGTC (faked DNA sequence). Also, if the complementary rule 1 is selected, then, as a first bit (most significant bit) apart from the obtained sequence, a letter 'A' is inserted which implicitly tells the receiver that rule 1 is selected. Likewise, if letter 'C' is inserted, then it tells that rule 2 is used and 'G' is used for rule 3. The message to be encoded is then taken and each letter in the faked DNA sequence is given subscript.

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

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 8, August 2016

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 they 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 and transmitted.

B.IMPLEMENTATION OF STEGANOGRAPHY

Steganography can be applied to any type of computer file. The most usually used formats are image , audio, video files. Image files are used for this project because we have used (C#) that has allowed us for easy implementation and manipulation of the image files.

It's also important to maintain the integrity of the image, to keep it intact after the hidden file is embedded, meaning the quality is not noticeably altered. The main intention of steganography is not to make a file look suspicious or draw any attention to it. The ordinary person looks at and can only see a normal image file.

The first thing that had to be considered is how the header of a bitmap file is structured, so we would know when to begin embedding the hidden data. The bitmap header is 54 bytes and structured as follows:

Byte Offset	Size (in bytes)	Description
0	2	Signature. Must be the letters "BM" in ASCII
2	4	Size of BMP file
6	2	Reserved. Must be 0
8	2	Reserved. Must be 0
10	4	Offset to start of image data (in bytes)
14	4	Size of BITMAPINFOHEADER structure, must be 40
18	4	Image width in pixels
22	4	Image height in pixels
26	2	Number of planes in the image, must be 1
28	2	Number of bits per pixel
30	4	Compression type (0=none, 1=RLE-8, 2 = RLE-4)
34	4	Size of image data in bytes (including padding)
38	4	Horizontal resolution in pixels per meter
42	4	Vertical resolution in pixels per meter
46	4	Number of colors in image, or 0
50	4	Number of important colors, or 0

We have chosen least significant bit insertion method for this project. It works by using the last bit of each red, green, and blue byte in each pixel for our storage purposes. Since a 24-bit image has so many colors, this small change by altering the last bit is not detectable by the human eye. The file being embedded can even be compressed beforehand using ZIP, RAR, or a similar format for even greater storage capacity within the steg file.

Let's assume we have 3 pixels and want to embed the letter "R" in a 24-bit bitmap using LSB insertion. They could be represented as bits:

	Pixel 1	Pixel 2	Pixel 3
Red	00100101	11010011	01010110
Green	10110000	11111101	00001011
Blue	00011101	11100100	00000001

To insert "R" (represented in binary as 01010010), we go through each pixel and set the last bit to what we need for "R." This would result in:

	Pixel 1	Pixel 2	Pixel 3
Red	00100100	11010011	01010111
Green	10110001	11111100	00001010
Blue	00011100	11100100	00000001

In order to not draw any suspicion to the stag file, we have copied over the original header precisely as it was in the original image.

The next step was to find a way to store the length of the embedded data so that we can find out when to stop extracting. We also considered storing the file extension and the file name, but decided contrary to this because the person that is deliberate to receive the stag image should know what the file contains already.

So the best way to store this information without noticeably altering the structure of the bitmap image was, to create our own 32 byte "stag header" and place it at the top of the image, such that it comes directly after the bitmap header. The steg header is only stored in the last bit of the first 32 bytes of the image data, similar to how the embedded data is stored. A person would never know the first 32 bytes of the image data are used for this purpose without any knowledge of our embedded algorithm. It simply looks like normal pixel data. We begin storing the embedded data directly after the steg header is written.

To work out the project we have used an object oriented approach ,such that, when we separate the user interface code from the "back end code", it has allowed us to work on them separately. Since the user interface code have nothing to do with steganography, the hub of this section will be on the "back end."

VI CONCLUSION

The anticipated 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 regain the original message as intended by the sender. Message encryption using DNA sequence is a very latest technique still evolving and tried out for secure transmission and reception of hided messages. It is so secure that it would be very complicated for any intruder to break the encrypted message and retrieve the actual message and image steganography can be a challenging venture. While maintaining complete secrecy may be impossible, creating the illusion that an image has not been tampered with is quite possible. With some effort, users can make use of LSB insertion to hide information within the lowest significant bits of any bitmap image format. LSB image steganography provides users ample return on investment with respect to both time and efficiency. By replacing the lowest significant bits of each byte in a bitmap image a stenographer can efficiently hide relatively large amounts of data within an image without significantly degrading the quality. Although all steganographic methods are partially susceptible to binary diff attacks, LSB insertion is by far the most cost effective method of image Steganography.

REFERENCES

[1] L. M. Ad leman, "Molecular computation of solution to combinatorial problems Science, (1994) 11, (266): 1021-1024.

[2] Chen Jie, “A DNA-based bio molecular cryptography design,” Proceedings of IEEE International Symposium, Vol. 3, pp. III-822, (2003).

[3] Pramanik Sabari, and Sanjit Kumar Setua, "DNA cryptography," In Electrical & Computer Engineering (ICECE), 7th IEEE International Conference on, pp. 551-554, (2012).

[4] Tushar Mandge, Vijay Choudhary. “A DNA encryption technique based on matrix manipulation and secure key generation scheme”. Information Communication and Embedded Systems (ICICES), International Conference on 21-22 Feb. (2013).

[5] Kazuo Tanaka, Akimitsu Okamoto, and Isao Saito, “Public-key system using DNA as a one-way function for distribution”. Bios stems 81, 1, pp. 25-29, (2005).

[6] Sherif T. Amin, Magdy Saeb and El-Gindi Salah, “A DNA-based implementation of YAEA encryption algorithm,” In Computational Intelligence, pp. 120-125, (2006).

[7] Cui, Guangzhao, Liming Qin, Yanfeng Wang, and Xuncai Zhang, "An encryption scheme using DNA technology," In Bio-Inspired Computing: Theories and Applications, IEEE International Conference on, pp. 37-42, (2008).

[8] Lai Xin-she, Zhang Lei, “A novel generation key scheme based on DNA”. Computational Intelligence and security, IEEE, International conference on 13-17 Dec. (2008).

[9] Lai, XueJia, "Asymmetric encryption and signature method with DNA technology," Science China Information Sciences 53.3, page 506-514, (2010).

[10] Deepak Kumar, and Shailendra Singh, "Secret data writing using DNA sequences," In Emerging Trends in Networks and Computer Communications (ETNCC), IEEE International Conference on, pp. 402-40, (2011).

[11] Bibhash Roy, Pratim singha, “An improved symmetric key cryptography with DNA based strong cipher”. Devices and Communications (ICDeCom), IEEE, 2011 International Conference on 24-25 Feb. (2011).

[12] A. K. Verma, Mayank Dave, R.C. Joshi, “Securing Ad hoc Networks Using DNA Cryptography”, IEEE International Conference on Computers and Devices for Communication (CODEC06), pp. 781-786, Dec. 18-20, (2006).

[13] Yunpeng Zhang, Bochen Fu, and Xianwei Zhang, "DNA cryptography based on DNA Fragment assembly," In Information Science and Digital Content Technology (ICIDT), IEEE International Conference on, vol. 1, pp. 179-182, (2012).

[14] Wang Zhong, Zhy Yu, “Index-based symmetric DNA encryption algorithm”. Image and Signal Processing (CISP), 2011 4th International congress on image and signal processing, 15-17 oct. (2011).

[15] Olga Tornea, and Monica E. Borda, "Security and complexity of a DNA-based cipher," IEEE Roedunet International Conference (RoEduNet), 11th, pp. 1-5, (2013).

[16] Borda, Monica, and Olga Tornea, "DNA secret writing Techniques," In Communications (COMM), 8th IEEE International Conference on, pp. 451-456, (2010).

[17] Ashish kumar kaundal, "Feistel Inspired structure for DNA cryptography" in June (2014).

[18]T. Korkel, J.H.P. Eloff, M.S. Olivier. "An Overview of Image Steganography". (2005).

[19]"BMP (Windows) Header Format." Fastgraph Home Page. (2001).

[20]Brundick, Frederick S., and Lisa M. Marvel. "Implementation of Spread Spectrum Image Steganography." Army Research Laboratory, Mar.(2001).

[21]Cummins, Johnathan, Patrick Diskin, Samuel Lau, and Robert Parlett. "Steganography and Digital Watermarking." (2004).

[22] Johnson, Neil F., and Sushil Jajodia. "Exploring Steganography: Seeing the Unseen." (1998).

[23]Lenti, Jozsef. "Steganographic Methods." 05 June (2000).