International Journal of Advance Research, Ideas and Innovations in Technology

INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 3 - V7I3-1614)
Available online at: https://www.ijariit.com

Medical image encryption using DNA cryptography

Rohan N. Kalpavruksha
rohankalpavruksha@yahoo.com
PES University, Bangalore, Karnataka

Roshan N. Kalpavruksha
rkalpavruksha@yahoo.com
PES University, Bangalore, Karnataka

Bhavani Shankar
shankarhmbhavani@gmail.com
PES University, Bangalore, Karnataka

Under the guidance of: Prof. Rajashree, Prof. Prasad

ABSTRACT

This paper proposed a novel cryptographic method for enhancing the security while transmission of image. Image which covers highest percentage of multimedia data, its protection is at most important. In image encryption methodologies, the pixels of original images are either manipulated or information is laid inside the image using the image as a cover to protect the data from undesired receivers. Scope of the project is to ponder on different ways of securing image data. Being a part of this digital world security becomes the toughest and most important things to handle. On the other hand, hacking is also a growing domain which is used to track out a lot of information to solve a lot of digital criminal issues. DNA (Deoxy Ribonucleic acid) molecules, having the capacity to store, process and transmit information, inspires the idea of DNA cryptography. This combination of the chemical characteristics of biological DNA sequences and classical cryptography ensures the non-vulnerable transmission of data. In this paper review has been made about the present state of art of DNA cryptography.

Keywords— *Cryptography, DNA Cryptography, Cloud Security, User level security, Strong Key Generation, Data Encryption, Data Decryption*

1. INTRODUCTION

Cryptography is a method of protecting data, information and communications with the use of codes/images, so that only those who designed it and for whom the information is intended can read and process it. Encryption is one of the most formidable ways to keep our information safe between two endpoints. It makes data unread-able, so even it ends up getting in wrong hands it is mostly useless. The methodologies of encryption in which components of DNA are used to hide plain text from unauthorized users in the network come under the study of DNA cryptography .In DNA cryptography, the 4 chemical bases of DNA namely adenine [A], guanine [G], cytosine [C], and thymine [T] has so far majorly been used in addition to algorithms of cryptography. Considering 'ATGC' to be a code we can have 24 combinations of it like 'TACG', 'TGCA' etc. (4! = 24).Although among the 24 combinations only 8 are possible in a real-life occasion as 'T' can combine only with 'A' and 'G' combines only with 'C' following the properties of DNA. But here we shall use all the 24 combinations to encrypt our data. This will increase the complexity of detecting patterns in the encrypted text.

In the study of cryptography, an image which contains information is converted to an unreadable or unrecognizable form by using algorithms. The image also has the capacity of carrying information. This field is gaining much popularity as image can carry vital information with them. Image encryption enables passing of data and information over unsecured networks like the internet. Without the correct key, the correct decryption to retrieve the original is a major challenge to high-end super computers. Image encryption plays a vital role in securing the transmission of images of military, healthcare, important government document images and other private images.

Security services comprises of confidentiality, authentication and data integrity, and digital signature. In case, a person "A" wants to send a message to another person "B", secretly they need to follow the above said security service mechanisms. In confidentiality the data security is provided using symmetric or asymmetric method in two different ways such as block cipher and stream cipher. Symmetric method uses a single key {K} for both encryption and decryption, whereas in asymmetric method uses a pair of keys {KU, KR} for encryption and decryption process separately.

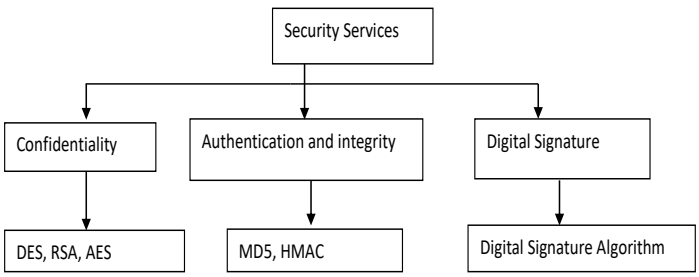


Fig. 1. Introduction to Cryptography (Ref. Geeks to Geeks)

2. CHALLENGES IN TRADITIONAL CRYPTOGRAPHY

Modern computers store data using a binary format. The size of the keys used in recent cryptographic applications is too big. It is very much difficult to crack a key when a billion calculations perform at a second as the combination to crack the key is larger and takes more time. Quantum computation is a new phenomenon which stores data using quantum bits. This performs calculations faster and hence the codes which take more time to break can be cracked speedily. Some of the challenges of traditional cryptographic methods are in which infrastructure it is executed, key size, and the quality of the algorithm. Recent days cloud computing and all other networking applications need information security for protecting the data and user validation. User validation validates the user and authenticates them after validity. As traditional encryption algorithm has severe security problems. The field of information security give importance to the new way of protecting the data. The main objective of DNA cryptography is to provide confidentiality when the persons send data over a network

3. DNA CRYPTOGRAPHY

DNA Cryptography is a combination of modern biotechnology and cryptology. DNA Cryptography can be defined as hiding information with the use of DNA Sequence. DNA Cryptography is one of the rapidly evolving technologies in present world. Adelman was the one showed the world how it can be used to solve complex problems like directed Hamilton path problem (for ex. Travelling Salesman problem).It makes it possible to break the unbreakable algorithms.

This is because DNA computing in terms of codes offers more speed, minimal storage and power requirements. DNA stores memory at a density of about 1 bit/nm3 whereas conventional storage media requires 1012 nm^3/bit. DNA computing does not require any power while the computation is taking place. The Surprising part is that one gram of DNA contains 1021 DNA bases which is equivalent to 108 TB of data and hence we can store all the data in the world in a few milligrams. To encode data in a DNA strand this is made up of 4 nitrogenous bases namely:

Adenine[A]Thymine[T] Cytosine[C] Guanine[G]

The easiest way to encode is to represent these 4 units as four figures:

A (0) –00 T (1) –01 C (2)–10 G(3)–11

4. XOR ONE TIME PAD

With the use of an example of XOR One Time Pad and see how it's implemented using DNA Cryptography. For example – Let M denote the message and K denote the key. The expression M xor K = C gives the CipherText. The user can revert back the encoded message by doing: C xor K = M xor K xor K= M, therefore getting the original message back.

The steps involved in implementing it are:

- The message and the OTP key are converted to ASCII bits
- Zero Padding is added to the message and the key in order to make the size of their binary codes even.
- The message and the key are XORed together
- The XOR output is represented in DNA bases format. This is our enciphered text.

The decryption process involves the following processes and hence it is also prone to eavesdropping:

- All the DNA bases are transformed into bits.
- These bits are then XORed with the OTP key bits to reproduce the original plain text.
- This text so obtained in binary format is then converted into a sequence of ASCII characters.

Instead of storing data as a sequence of 0s and 1s, storing them as a sequence of nitrogenous bases. Storing information in the form a DNA enables us to store a lot of data in a small area.

Step 1: The key sequence S (one-time pad), which contains random R number of bits, is distributed to both sender and receiver in advance. These bits are used for encryption of the plaintext. L is a variable which denotes number of unused bits in S. Initially, L = R.

Step 2: P which represents the plain text contains N bits such that N < L. Each bit Pi (where, i = 1... N) isXOR'ed with Ki = - S_{R-L+i} to generate the encrypted cipher bit Ci , where Ci = Pi ⊕ Ki .

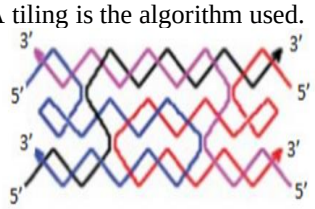
Table.1. Ci = Pi ⊕ Ki

Pi	*Ki*	*Ci*
0	0	0
0	1	1
1	0	1
1	1	0

Step 3: From the source pad sequence S the user bit is destroyed.

Step 4: The cipher sequence C = C1 , C2, … … , Cn is generated after encrypting all the bits of P.

Step 5: To decrypt the cipher text the receiver again performs XOR using each bit of cipher sequenceCi and the pad sequence Ki to generate the plaintext bit Pi i.e. Ci⊕Ki =Pi. The self-assembly of DNA tiling is the algorithm used.



TAO triple-helix tile

Fig. 2.TAO triple helix tile (Ref Research Paper, DNA based cryptography by Iran)

The Verman Cipher is implemented in DNA Cryptography using the following steps.

Step 1: The n bits of the plaintext are encoded in the form of DNA.Sequences and using the scaffold stranda1, a2 … … an is generated.

Step 2: The further portion of the scaffold strand a1' + a2'+ … an' +can be formed using arbitrary inputs which is actually the One-time pad of the scheme.

Step 3: The input assembly, the tile structure is formed by using the two types of input scaffold strands.

Step 4: An opening for hybridization of a single output tile is created by the input assembly because of the un-complemented sticky ends, after the introduction of the output tiles (i.e. encoded ciphertext) because of the self-assembly of DNA tiles, it can attach to the desired sticky ends by itself.

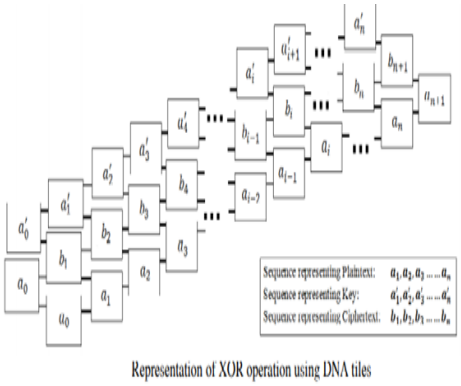
Step 5: Reporter strand R=a1a2 … an. a1'a2' … an'.b1b2..bn

Where bi= ai⊕ai'

Step 6: The reporter strand contains three domains. The first domain (a1a2…an) encodes the input plaintext, the second domain (a1'a2'…an') encodes one-time pad (key) and the third domain(b1b2..bn) represents the cipher text. If a restriction site between the second and third domain is restricted, the cipher text can be recovered from the reporter sequence

Step 7: According to the principle of the Vernam cipher, the same key n sequence (a1'a2'…an') is used for decryption. The plaintext can be recovered by using scaffold strands representing ciphertext(b1b2..bn) and key sequence as input assembly. The intended output is, ai= bi ⊕ai'

Step 8: After self-assembly of DNA tiling, the reporter strand, + is recovered. R'=a1a2 … an. a1'a2' … an'.b1b2..bn



Representation of XOR operation using DNA tiles

Fig. 3. Representation of XOR operation using DNA tiles (Ref Report on DNA cryptography by Kumar)

5. TRIPLE STAGE DNA CRYPTOGRAPHY

Soni et al proposed a new algorithm of DNA cryptography based on the concept of Moore machine in automata theory. The authors claim that the designed cryptosystem is more dependable and the security of this technique relies on the three encryption stages which use Secret key, Auto generated Moore machine and Password. Automata theory can be defined as the study of abstract self-propelled devices which follow predetermined sequence of operations to solve computational problem automatically. Final state machine (FSM) is a type of automaton that can be defined as a machine with finite number of states and the automaton can be in exactly one state at any given time. Moore machine produces output depending only on the present state. It can be defined by a 6 tuple were

$$M = (Q , \sum , \lambda, \Delta, \delta, q_0)$$

Q is a nonempty finite set of states.
$\sum$ is a nonempty finite set of input symbols.
λ is a output transition function, where λ is $Q \rightarrow \Delta$
Δ is a finite set of output symbols.
δ is a input transition function, where δ is $Q \times \sum \rightarrow Q$
q0 is the initial state, where $q0 \in Q$.

Step 1: By using user statically/dynamically generated secret key the input hidden message is encrypted first.
Step 2: The encrypted data is converted into binary sequence. The binary data is partitioned in n number of blocks of size 256 bits each.
Step 3: Now on the blocks the XOR operation is performed.
Step 4: Following the codebook-I, output of XOR operation is converted into DNA sequence.
Step 5: In this step Moore machine takes the resultant DNA sequence of step 4 as input and produces user password data as output. The transition table for generating user password data is given below.
Step 6: The newly created user password data is again converted in DNA sequence following

DNA codebook-II. The resultant sequence is the final encrypted data which is sent to the intended recipient. Because of the multiple layers in the encryption methodology, the authors assert that the proposed algorithm is very difficult to break. The added advantage of the algorithm is handling a broad range of plaintext in terms of data size.

Table.2. State Transition Table (Ref Research Paper, DNA based cryptography by Iran)

Current State	Next State				Output
	Input = A	Input = G	Input = T	Input = C	
q_0	q_1	q_7	q_{12}	q_{17}	Null
q_1	q_2	q_3	q_4	q_5	Null
q_2	q_1	q_7	q_{12}	q_{17}	W_1
q_3	q_1	q_7	q_{12}	q_{17}	W_2
q_4	q_1	q_7	q_{12}	q_{17}	W_3
q_5	q_1	q_7	q_{12}	q_{17}	W_4
q_7	q_8	q_9	q_{10}	q_{11}	Null
q_8	q_1	q_7	q_{12}	q_{17}	W_5
q_9	q_1	q_7	q_{12}	q_{17}	w_6
q_{10}	q_1	q_7	q_{12}	q_{17}	W_7
q_{11}	q_1	q_7	q_{12}	q_{17}	W_8
q_{12}	q_{13}	q_{14}	q_{15}	q_{16}	Null
q_{13}	q_1	q_7	q_{12}	q_{17}	W_9
q_{14}	q_1	q_7	q_{12}	q_{17}	W_{10}
q_{15}	q_1	q_7	q_{12}	q_{17}	W_{11}
q_{16}	q_1	q_7	q_{12}	q_{17}	W_{12}
q_{17}	q_{18}	q_{19}	q_{20}	q_{21}	Null
q_{18}	q_1	q_7	q_{12}	q_{17}	W_{13}
q_{19}	q_1	q_7	q_{12}	q_{17}	W_{14}
q_{20}	q_1	q_7	q_{12}	q_{17}	W_{15}
q_{21}	q_1	q_7	q_{12}	q_{17}	W_{16}

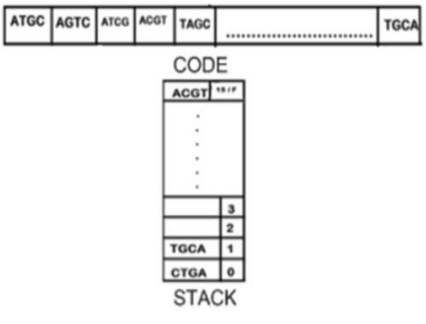
State transition table for user password data generation

6. DNA STEGANOGRAPHY

Step 1: The plaintext is deleted by input DNA strands.
Step 2: The input strands are tagged with secret key which are also in the form of DNA sequence.
Step 3: These strands are mixed with random DNA strands which are specified as distracter.
Step 4: If the secret key is known to the receiver then the strands representing the plaintext can be extracted from the mixed up solution following affinity purification protocol. The single stranded sequence used in the experiment is the complementary sequence of the secret key.
Another technique of entropy reduction is, if making the plaintext to imitate the distracter. This can be achieved by compressing the plaintext with lossless algorithm.

7. ENCRYPTION

Step 1: Finding all the 24 combinations of ATGC and storing them in an array named CODE of size 24.
Step 2: STACK of size 16 is declared. Using random function, generate numbers between 0 and 23. The number generated the corresponding ATGC combination of that index CODE would be placed at the top of the STACK.
Step 3: The pixel values of the image are fetched in matrix form.
Step 4: Finding the first seven most occurring colours from matrix that are present in matrix.
Step 5: From the remaining eight combinations in CODE, seven are again taken randomly and each of the most occurring colour is specially assigned a combination. The combinations being stored in an array MOST_OCCUR.



CODE and STACK

Fig. 4. Encryption Stack (Ref DNA Image encryption using modified symmetric key,IEEE,2017)

Step 6: The symbol '#' is given a code which gets left after assigning all the 23 combinations corresponding to 0 to F and the most occurring

Step 7: The width and height of image are encoded with the codes corresponding to 0 to F in CODE and written in a file named CODE_FILE with code of '#' after width and height.

Step 8: The pixel values of the image (considering already in hexadecimal) in horizontal (X-axis) direction is picked and stored in a variable 'Hex'. Three cases can occur:

Case A*:* The color is one of the most occurring colors, and then it is encoded with its corresponding code stored in MOST_OCCUR and written in the CODE_FILE.

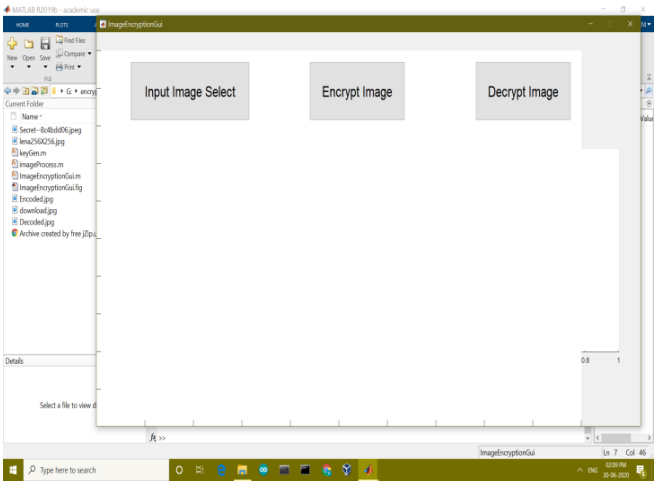
Case B*:* The color is not among the most occurring seven colors. Then each character from 'Hex' will be extracted and encoded with the codes corresponding to its value (equal to index number, for 'A' to 'F' codes of 10–15 will be assigned) in CODE.

Case C: The color is same as that of previous color. Then a counter will count the number of times it is repeated. The final number will encoded with the codes corresponding to 0–9 in CODE and written in CODE_FILE with the code of '#' concatenated before and after it. For example, suppose '1010F1' occurs 10 times serially. The above steps will get repeated till the end of the matrix containing the hexadecimal codes of the pixels of the image.

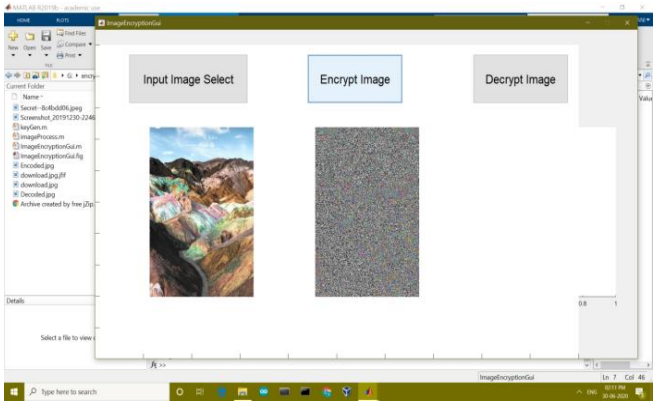
Step 9: The 7 most occurring colours will be treated as the normal colours and encoded similar to Case B of Step 8 and written in a separate file MOST_OCCUR_FILE.

8. PROGRAM RESULT

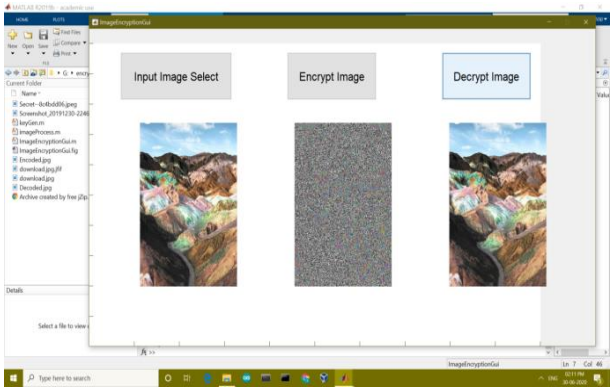
Stage 1: Here the image is selected for encryption.



Stage 2: Here the image is selected for encryption and the result is encrypted image.



Stage 3: So here we decrypt the image so as to get earlier image.



9. CONCLUSION

DNA cryptography is in its infancy. In the last few years DNA computing is experiencing real progress. DNA cryptography is not studied that well yet, but ramped up work in cryptography over the past several years has laid good groundwork for applying DNA methodologies to cryptography and steganography. Researches and studies are being carried out to identify a better and unbreakable cryptographic standard. A lot of schemes are being proposed regarding DNA cryptography and is being explored. Presently the work in DNA cryptography is focused on using DNA sequences to encode binary data in some or the other way. The field is still in the developmental stages extremely as it is complex and in current work there is a lot of hope that DNA computing will act as a good technique for Information Security.

10. ACKNOWLEDGEMENT

Our warm gratitude to PES University Computer Science Dept. IFSCR team for allowing us to work with them on image encryption and decryption using DNA cryptography and special thanks to our respected guide Prof.Rajshree helped us enrich our knowledge about the topic by sharing her valuable understandings and also Mrs.Vineetha who enlightened us with her skills.

10. REFERENCES

[1] http://en.wikipedia.org/wiki/Cryptography
[2] https://www.hindawi.com/journals/scn/2019/8694678/
[3] https://www.tandfonline.com/doi/abs/10.1080/19393555.2020.1718248
[4] https://resources.infosecinstitute.com/dna-cryptography-and-information-security/#gref
[5] https://link.springer.com/article/10.1007/s11042-017-4741-7
[6] http://en.wikipedia.org/wiki/Plaintext
[7] http://en.wikipedia.org/wiki/Ciphertext
[8] http://en.wikipedia.org/wiki/Encryption
[9] Jadhav D, Ochani A, Gulwani R (2017) DNA Image encryption using modified symmetric key (MSK). In: International conference on inventive computation technologies, IEEE
[10] Biradar S, Akkasaligar PT (2017) Secure medical image encryption based on intensity level using chaos theory and DNA cryptography. In: International conference on computational intelligence and computing research, IEEE, Chennai 8.
[11] Tamaki, K., Jeffreys, A.J., MacLeod, A., Neil, D.L. and Monckton, D.G. (1991). Minisatelliterepeat coding as a digital approach to DNA typing. Nature 354, pp. 204- 209.
[12] Hei RM, Niyat AY, Jahan MV (2016) Chaos-based image encryption using a hybrid cellular automata and a DNA sequence. In: 2015 International congress on technology,

communication .IEEE, 11–12 November 2015, Mashhad, Iran 9.

[13] Lindell, Yehudaand Katz, Jonathan (2014). Introduction to Modern Cryptography. Second Edition, CRC Press.

[14] R. Prajapati, Soni, G., Khan, A. and Kulhare, D. (2013). Triple Stage DNA Cryptography Using Sequential Machine. International Journal of Advanced Research in Computer Science Engineering.

