Presentation on

A Hybrid Encryption Technique Based on DNA Cryptography and Steganography

By-

Shahriar Hassan

Md. Asif Muztaba

Md. Shohrab Hossain

Husnu S. Narman

13th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)
12-15 October, New York, 2022

INTRODUCTION

- Secure data transmission is a serious concern now
- **Data Encryption** is used to enhance data security
- For more security **Data Steganography** is used
- Challenging to find a specific algorithm that encrypts and hides data in such a way that it does not get intruders' attention
- We propose a secure Hybrid Data Encryption technique using **DNA Cryptography and Steganography**
- DNA will be used for both encrypting data and hiding data

IEMCON

OBJECTIVES

- To propose a robust model of data encryption so that data can be transmitted securely without getting attention of the intruder
- To analyze the security of the proposed model
- To provide a comparative analysis with some related works

IEMCON

OUTCOMES AND APPLICATIONS

- **Outcomes:** A Blind, Symmetric, DNA based encryption and steganographic technique which have a decent cracking probability
- **Applications:** Our proposed approach will help secure data transmission, especially in banking, e-commerce, authentication, and server-client secure communication sector

IEMCON

RELATED WORKS

DNA Cryptography based Works

Namdev et. al. [2]

- Proposed a DNA and Amino acid based approach with Playfair Cipher
- Message encrypted in DNA sequence further encrypted by a foursquare encryption process based on Amino acid structures

DNA Steganography based Works

Shiu et. al. [3]

- Compared 3 techniques, i.e., insertion, complementary rule and substitution
- No encryption was done

IEMCON

RELATED WORKS

Guo et. al. [4]

- Proposed a substitution based method in DNA sequence for message hiding
- They explored Motifs in a DNA sequence and substituted them with message bits

Yunus et. al. [5]

- Also proposed a Motif substitution method in DNA sequences
- Not blind and may have high modification rate

Hamed et. al. [6]

- Proposed complementary rule based DNA steganography technique
- Does not preserves biological functionality

IEMCON

RELATED WORKS

Mousa et. al. [7]

- Proposed a reverse mapping based method in DNA sequence for message hiding
- Reverse mapping is a kind of substitution technique

Hybrid Techniques

Mitras et. al. [10]

- Proposed a encryption method based on RSA algorithm and DNA encryption
- After encryption they used insertion method to hide the encrypted message into DNA sequence

Taaur et. al. [9]

- Employed 5*5 playfair cipher technique for data encryption
- Then used insertion method for data hiding

IEMCON

RELATED WORKS

Yadav et. al. [11]

- Encrypted the data with DNA encryption
- Then using KIMLA algorithm to hide the encrypted message into an image by manipulating its pixel values

IEMCON

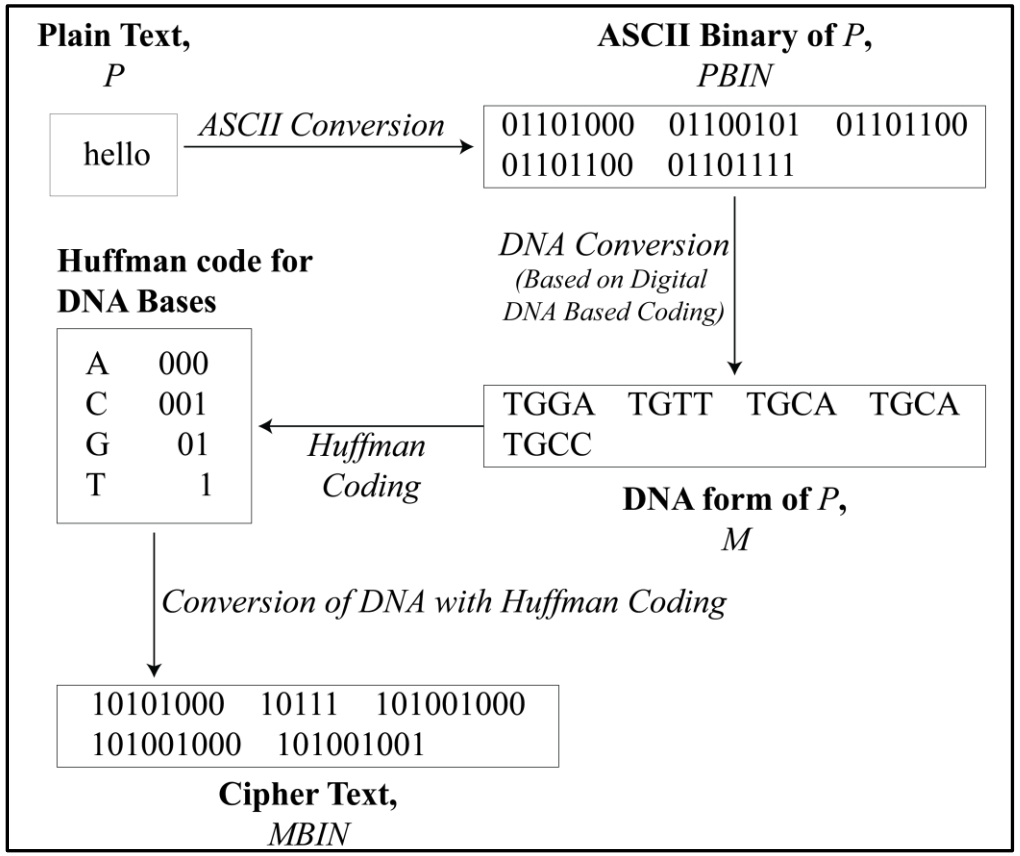
PROPOSED METHOD

- Our proposed method has **two phases**:
    1. Encoding the data using DNA encryption
    2. Hiding the data in real DNA sequence
- For DNA encryption we used **2-bit Binary Encoding** technique
- Also we employed Huffman Coding to hide the encoding procedure
- For data hiding we used a modification of **3:1 LS-Base** method

IEMCON

PROPOSED METHOD

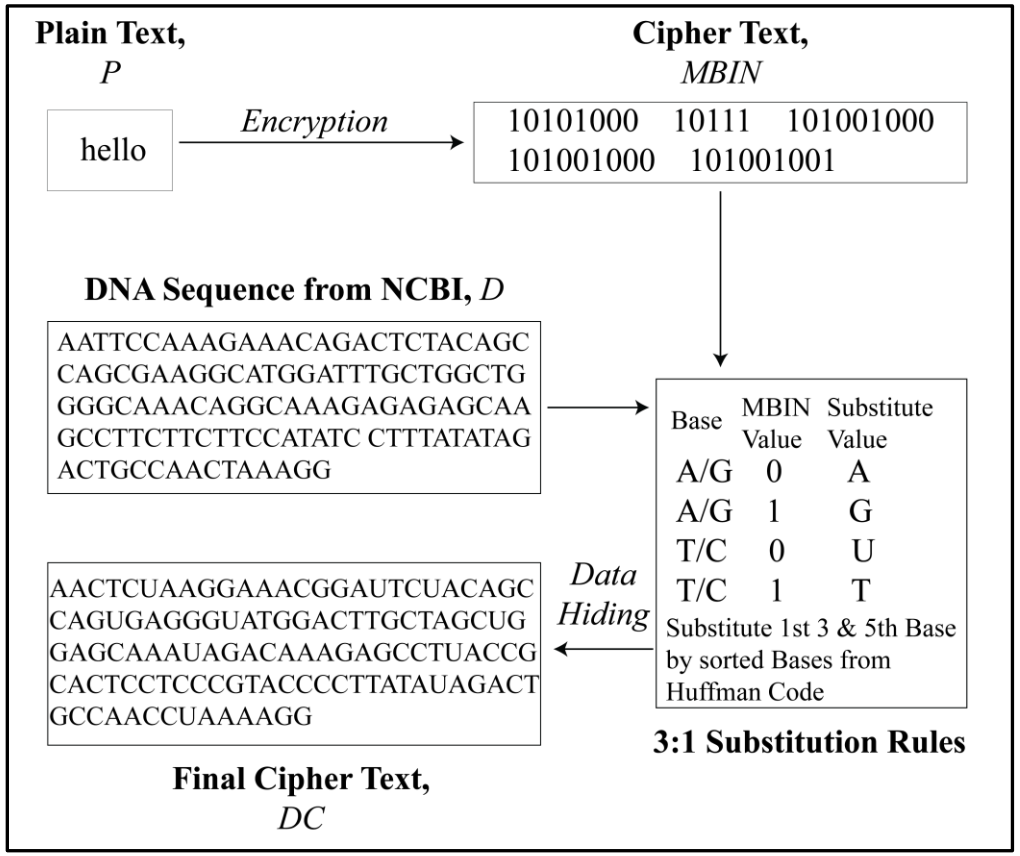
- Flowchart explaining the data encryption method with an example



IEMCON

PROPOSED METHOD

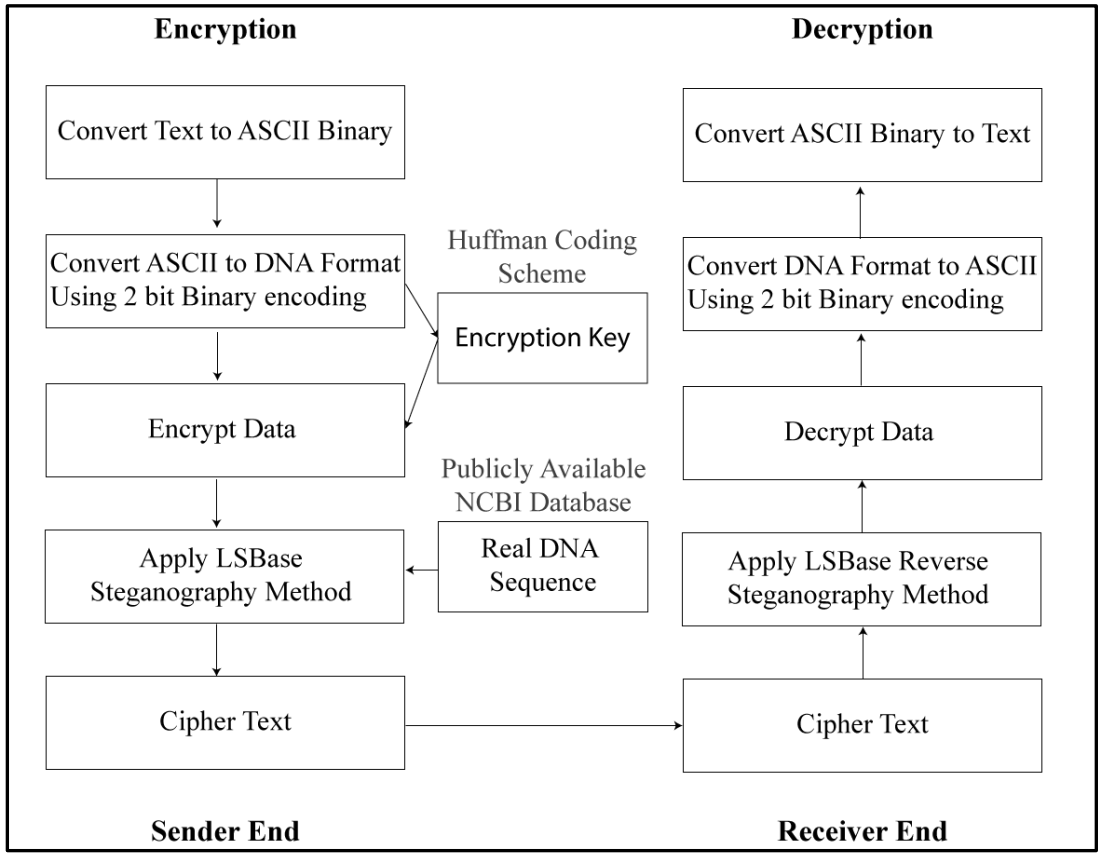
- Flowchart explaining the data hiding process with an example



IEMCON

PROPOSED METHOD

- Flowchart of the proposed method



IEMCON

IMPLEMENTATION

- We have implemented the method using Python programming language
- The algorithm was tested on Intel(R) Core (TM) i5-8300H CPU @ 2.30 GHz personal computer with 8 GB RAM
- We took a text message containing letter, digits and symbols in a 5KB file
- We encrypted the text message and hide it in 8 real DNA sequences of different length collected from NCBI database
- Then we investigated capacity, payload, bit per nucleotides, encryption time and decryption time for each DNA sequences
- Also we derived the cracking probability of our proposed model and done comparative study with some recent works.

IEMCON

DNA SEQUENCES

- Following are the 8 DNA sequences collected from the NCBI database

Locus	Number of Nucleotides(bp)	Species Definition
AC166252	149,884	Mus musculus 6 BAC RP23-100G10
AC168901	191,456	Bos taurus clone CH240-1851
AC168907	194,226	Bos taurus clone CH240-19517
AC153526	200,117	Mus musculus 10 BAC RP23-383C2
AC168897	200,203	Bos taurus clone CH240-190B15
AC167221	204,481	Mus musculus 10 BAC RP23-3P24
AC168874	206,488	Bos taurus clone CH240-209N9
AC168908	218,028	Bos taurus clone CH240-195K23

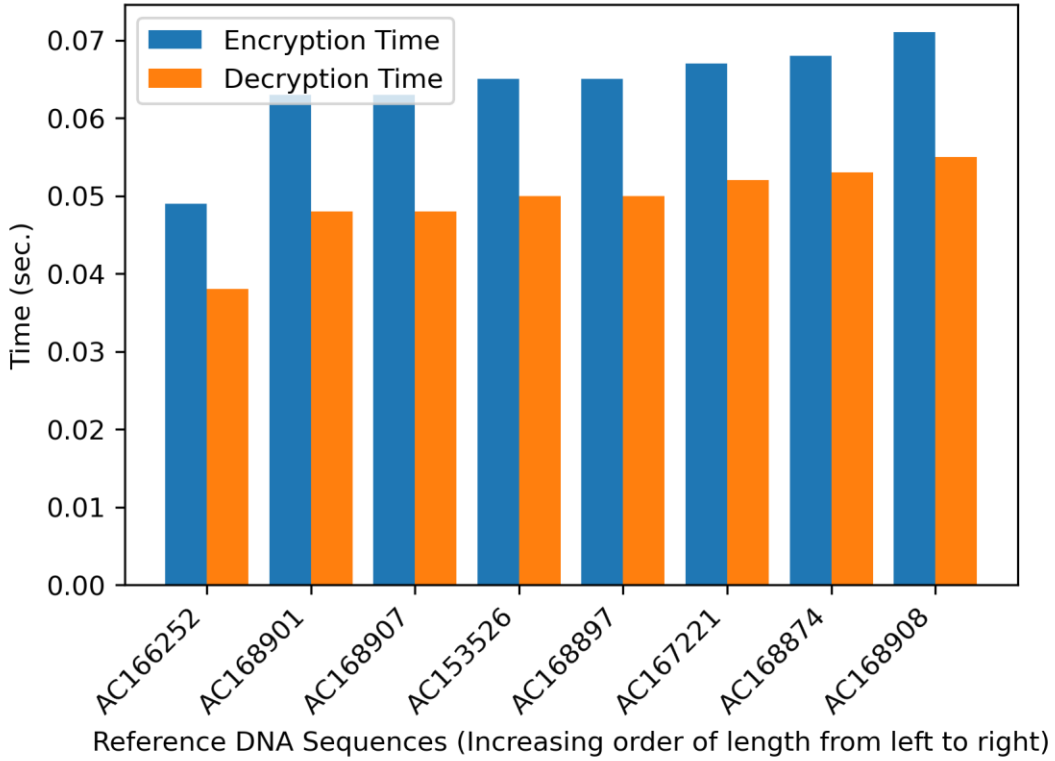
IEMCON

EXPERIMENTAL RESULTS

Locus	Capacity(bits)	Payload	bpn = (M+K)/C	Encryption Time(Sec)	Decryption Time (Sec)
AC166252	49965	0	3.6	0.049	0.038
AC168901	63822	0	2.8	0.063	0.048
AC168907	64746	0	2.8	0.063	0.048
AC153526	66709	0	2.7	0.065	0.050
AC168897	66738	0	2.7	0.065	0.050
AC167221	68284	0	2.6	0.067	0.052
AC168874	68833	0	2.6	0.068	0.053
AC168908	72680	0	2.5	0.071	0.055

IEMCON

EXPERIMENTAL RESULTS



IEMCON

SECURITY ANALYSIS

- **DNA Reference Sequence**

  There are around 163 million DNA sequences into the public database and the first 6 bases of the sequence might be fully changed in our model. So, intruder needs to analysis the rest n-6 bases of a DNA sequence. Hence, probability of guessing correct DNA sequence is:

$$\mathrm{P}(DNA_{Ref}) = \frac{1}{1.63*10^{8}*(n-6)}$$

- **Binary Encoding Rule**

  DNA sequence has only 4 symbols A, T, C, G. Huffman code for those can be 000, 001, 01, 1. Again binary encoding creates a code of 00, 01, 10 and 11 for them. So, the probability of guessing the correct code each time is:

$$\mathrm{P(BER)} = \frac{1}{4!*4!}$$

IEMCON

SECURITY ANALYSIS

- **LS Base Substitution Rule**

  There are possibilities for pyrimidine base substitution are 2*1. The possibilities are sme for purine bases. Hence, probability of guessing correct substituted nucleotides is:

$$P(N) = \frac{1}{4}$$

- **System Cracking Probability**

$$P(S) = \frac{1}{1.63*10^{8}*(n-6)*4!*4!*4}$$

IEMCON

COMPARATIVE STUDY

Comparison Criteria	P1: Enhanced Double Layer Security using RSA over DNA based Data Encryption [10]	P2: DNA Base Data Encryption and Hiding using Playfair and Insertion Techniques [9]	P3: Proposed Steganography Approach using DNA Properties [6]	P4: A New Data Hiding Scheme Based on DNA Sequence [5]	P5: The Proposed Method
Secret Text Type	Any Type of Data	Any Type of Data	Any Type of Data	Binary Data	Any Type of Data
Binary Coding Rule	2-Bit Binary Coding Rule	2-Bit Binary Coding Rule	2-Bit Binary Coding Rule	Binary Coding Rule Independent	2-Bit Binary Coding Rule
Encryption Type	Symmetric	Asymmetric	Not Applicable	Not Applicable	Symmetric
Encryption Algorithm	Encrypting secret data by mapping it to DNA and amino acids	5*5 Playfair cipher based on DNA and amino acids	No Encryption	No Encryption	DNA Based Huffman Coding Encryption
Data Hiding Algorithm	Insertion	Insertion	Complementary rules based hiding method, which is the rule that specifies the strand of DNA directly opposite a specified sequence	Substitution method using repeated nucleotides to hide the secret message bits	Substitution method using the least significant base of each codon in the DNA reference sequence
Blind/Not Blind	Not Blind	Blind	Not Blind	Not Blind	Blind
System Cracking Probability	$P(S) = 1/(1.63 * 10^8 * (n-1) * 24 * 2^(m-1) * 2^(s-1))$	$P(S) = 1/(1.63 * 10^8 * (n-1) * 24 * 2^(m-1) * 2^(s-1))$	$P(S) = 1/(1.63 * 10^8 * (n-1) * 24 * 24)$	$P(S) = 1/(1.63 * 10^8 * (n-1) * 24 * 6)$	$P(S) = 1/(1.63 * 10^8 * (n-6) * 4! * 4! * 4)$
Security Level	Double Layer	Double Layer	Single Layer	Single Layer	Double Layer
Modification Rate	High	High	Moderate	High	Low
Biological Functionality	Does not Preserve	Does not preserve	Does not preserve	Does not preserve	Preserves
Capacity	High	High	Moderate	Moderate	Moderate

CONCLUSION AND FUTURE WORK

- In this work, we proposed a novel cryptographic technique combining DNA cryptography and steganography
- The technique encrypts the data in its first stage and then hides the encrypted message into an actual DNA sequence
- The encryption method uses DNA bases to encrypt the message, followed by a variable length code generation and assignment for each DNA base using Huffman coding
- The proposed method is blind and it does not expand the actual DNA sequence while keeping its biological functionality
- Experimental results and analysis shows that our proposed method gives a decent level of security which is quite impossible to break without having full knowledge of the steps involved in particular encryption
- The proposed method can be modified in our future work to increase its data hiding capabilities and security

IEMCON

REFERENCES

[1] Niu, Ying, et al. "Review on DNA cryptography." *International Conference on Bio-Inspired Computing: Theories and Applications*. Springer, Singapore, 2019.

[2] Namdev, Sonal, and Vimal Gupta. "A Dna and Amino-Acids Based Implementation of Four-Square Cipher." *Int. Journal of Engineering Research and Applications* 6 (2016): 90-96.

[3] Shiu, Hung-Jr, et al. "Data hiding methods based upon DNA sequences." *Information Sciences* 180.11 (2010): 2196-2208.

[4] Guo, Cheng, Chin-Chen Chang, and Zhi-Hui Wang. "A new data hiding scheme based on DNA sequence." *Int. J. Innov. Comput. Inf. Control* 8.1 (2012): 139-149.

[5] Yunus, Yunura Azura, Salwa Ab Rahman, and Jamaludin Ibrahim. "Steganography: a review of information security research and development in muslim world." *American Journal of Engineering Research* 11 (2013): 122-128.

IEMCON

REFERENCES

[6] Hamed, Ghada, et al. "DNA based steganography: survey and analysis for parameters optimization." *Applications of intelligent optimization in biology and medicine*. Springer, Cham, 2016. 47-89.

[7] Mousa, Hayam, et al. "Data hiding based on contrast mapping using DNA medium." *Int. Arab J. Inf. Technol.* 8.2 (2011): 147-154.

[8] Vijayakumar, P., V. Vijayalakshmi, and R. Rajashree. "Increased level of security using DNA steganography." *International Journal of Advanced Intelligence Paradigms* 10.1-2 (2018): 74-82.

[9] Taur, Jin-Shiuh, et al. "Data hiding in DNA sequences based on table lookup substitution." *International Journal of Innovative Computing, Information and Control* 8.10 (2012): 6585-6598.

[10] Mitras, Ban Ahmed, and A. Abo. "Proposed steganography approach using DNA properties." *International Journal of Information Technology and Business Management* 14.1 (2013): 96-102.

IEMCON

REFERENCES

[11] Yadav, Vikash, and Indresh Kumar Gupta. "A hybrid approach to metamorphic cryptography using KIMLA and DNA concept." *International Journal of Computational Systems Engineering* 5.4 (2019): 218-229.

[12] Yadav, Vikash, and Indresh Kumar Gupta. "A hybrid approach to metamorphic cryptography using KIMLA and DNA concept." *International Journal of Computational Systems Engineering* 5.4 (2019): 218-229

[13] Hamed, Ghada, et al. "Hybrid technique for steganography-based on DNA with n-bits binary coding rule." *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*. IEEE, 2015.

[14] Sajisha, K. S., and Sheena Mathew. "An encryption based on DNA cryptography and steganography." *2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA)*. Vol. 2. IEEE, 2017.

IEMCON

THANK YOU

IEMCON