Data Hiding Using Deoxyribonucleic Acid (DNA) Computing With Morse Code Cryptosystem

B. Adithya[1], G. Santhi[2]

[1]Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India
[2]Department of Information Technology, Pondicherry Engineering College, Puducherry, India
e-mail: [1]adithya27.07@pec.edu

Abstract: In this paper, DNA computing with Morse code encoding is proposed to bulwark the delicate information within the challenging environment. The plaintext is changed over to DNA sequences utilizing encoding table. Presently, the encoded information is transcript and translated by the RNA arrangements. Translated RNA is stego by the standard genetic code utilizing organic compounds, and the stego DNA is ciphered by Morse code pattern. The designed bio analysis and results show that the hiding capacity is high, and the execution time is less when compared to the current techniques like AES, RSA, DNA based Playfair cipher and Vigenere cipher.

Keywords: Deoxyribonucleic Acid, Morse code, cryptography, steganography, Data hiding

1. INTRODUCTION

For encouraging the delicate trade of data between any sender and beneficiary, safe correspondence is vital. These days, the web has become the discussion for all banking and electronic trade exchanges and it is extremely important that the connection is made in a profoundly secure way. A few strategies and frameworks have been created to scramble and unscramble the plain content in numerical cryptography to satisfy these security prerequisites. Such methodologies are anyway conquered utilizing procedures and strategies for DNA cryptography. DNA cryptography is a significant control of computational DNA science [1]. DNA cryptography plays a major part in the survival of the next generation. Numerous calculations offered in DNA cryptography have constraints therein some of their steps they either utilize standard math cryptography or upheld natural lab explores that do not appear to be suitable inside the advanced registering world. Proposed framework describes a new, effective, unique, and dynamic DNA calculation method to address this gap, and also give an overview of its results. This DNA cryptography uses the principle of the Central Molecular Biology Dogma (CDMB) for the computation. The continuation of the study is structured as follows. Section 2 defines the criteria that the DNA encryption algorithm must satisfy. The proposed strategy is defined in section 3. The experimental findings are discussed in Section 4. Section 5 explains how the specifications are met. Section 6 brings this research to a conclusion.

2. REQUIREMENTS FOR USING DNA COMPUTING ALGORITHM

Every algorithm for computing the DNA should follow a set of specifications. Those criteria were defined in this paper based on the constraints found in the current encryption algorithms as described in Table 1. In DNA computing algorithm, Table 2 presents the execution of a set of requirements in the existing works. The below table 2 shows that the performance of the DNA cryptography specifications of the existing works is not complete. The following discussion is evolving the same.

Table 1: Description of criteria for DNA computing algorithm to be satisfied

Criteria	Description
• Entire character set DNA encoding	• The DNA coding table must contain the complete set of characters, of which 96 elements for DNA encoding sequences.
• Dynamic generation of encoding table	• At each interval session encoding table is created according to the random principle and provides concrete DNA sequences, for each element from the set of characters.
• Unique sequence to encode DNA sequence for every plaintext character	• For each element of the character set in each encoding table generation of each session, plaintext encoding into DNA sequence is unique
• Robustness encoding	• The table of character encoding must be based on a high degree of secure random encoding table must also contain random steps generations.
• Preserving biological process in the simulation	• The DNA computing algorithm is based on the biological processes that are replicated to fit into the world of digital computing.
• Dynamicity of encryption	• Due to the unique generation of DNA encoding table, different ciphertext can be generated for each session using the same plaintext

Table 2: Execution of a set of requirements in the existing works

Authors	Entire character set DNA encoding	Dynamic generation of encoding table	Unique sequence to encode DNA sequence for every plaintext character	Robustness encoding	Preserving biological process in the simulation	Dynamicity of encryption
Agrawal et.al [3]	✓	✗	✗	✗	✗	✗
Ning [4]	✗	✗	✗	✗	✓	✗
Sabry et. al [5]	✗	✗	✗	✗	#	#
Sadeg et. al [6]	✗	✗	✗	✗	✓	✗
Zhang et. al [7]	✗	✗	✗	✗	✗	✗

Murugan et. al [8]	×	×	×	×	×	✓

× Indicates minimum acquisition # Indicates partial acquisition ✓ Indicates Appropriate acquisition

3. PROPOSED METHOD

The algorithm for DNA encoding is a complete description of a mechanism defined in [2]. For brevity's sake, this paper focuses on the algorithm of encryption and the decryption which is maintained, and not treated in depth. The computing algorithm consists of the following steps to encrypt a plaintext into ciphertext is shown in Fig. 1.

Encoding processes for plaintext to DNA sequence conversion to be performed before the encryption begins. For example: Let's expect the plaintext is, "Tree" and the plaintext are equitably isolated into two parts. If the plaintext is not divided equally, an element of randomness is applied to make them equal. Normally, plaintext is first changed over to DNA sequence by utilizing the following binary code rule A-00, T-01, C-10, and G-11 [9]. The cycle of transformation is simply a plaintext encoded with DNA, with the respective intron sequences. The turned DNA sequence is translated into mRNA sequence by supplanting Thymine (T) with Uracil (U) on both the left and right sides of the DNA groupings. It is a method of simulating a biological transcriptions mechanism.

The mRNA sequence is translated into a tRNA sequence by substituting each DNA from the letter set with its DNA letter set complement. For illustration, changes will be performed to A-U, U-A, G-C, C-G. The tRNA sequence is changed over to DNA sequence by the substitution of Uracil (U) with Thymine (T). This is an organic method of reverse-transcription of the simulation. The reverse-transcription DNA sequence is moved once on the two sides of the right-hand side. The universal basic amino acids table consists of 20 amino acids; in this work, it is extended to 256 elements is shown in Table 3.

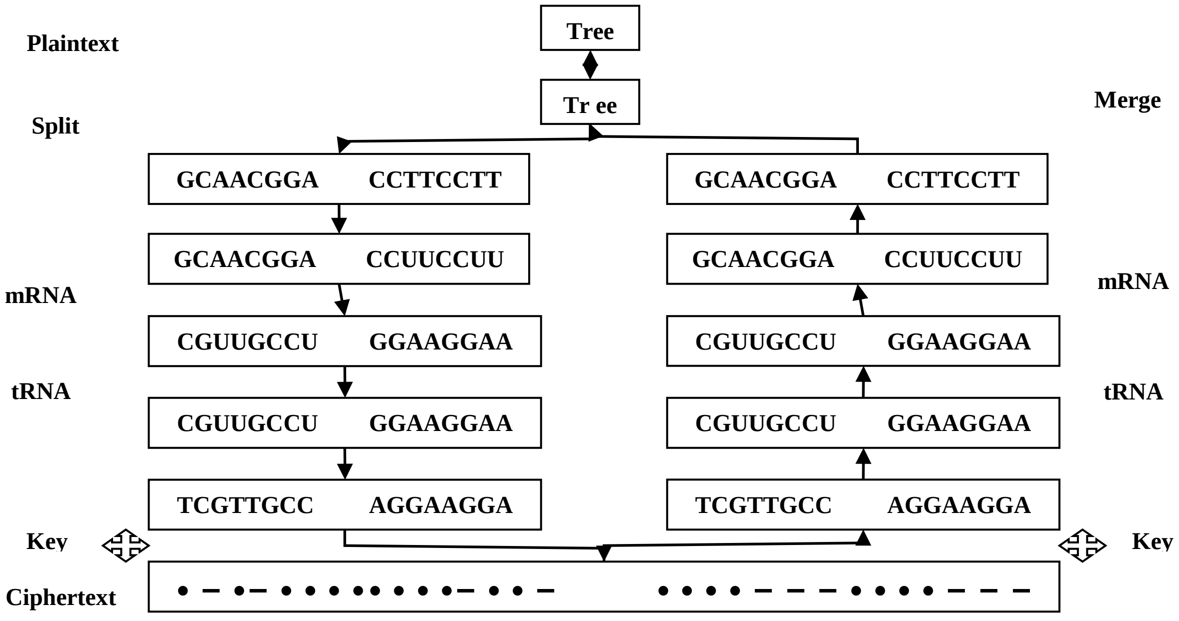

Fig. 1. Block Diagram of Data Hiding using DNA Computing with Morse Code Cryptosystem

Morse code is a way to transfer information by using the regular sequence of short and long marks or pulses of letters, numbers and special characters of a message, usually referred to as “dots and dashes”. The Morse code pattern is applied over the encoded DNA sequences to form the ciphertext. The receiver receives the encrypted message and the idea of the alternate network from the safe link, and the recipient employments the DNA encoding calculation to build two DNA encoding tables from their claim clue and the sender clue. Finally, both the left and the right side plaintext are blended.

Table 3: Extended amino acid encoding table with 256 elements

	T	**G**	**A**	**C**	**T**	**G**	**A**	**C**	**T**T	**G**	**A**T	**C**T	**T**	**G**	**A**	**C**
	A	**A**	**A**	**A**	**G**	**G**	**G**	**G**		T			**C**	**C**	**C**	**C**
T	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC	TC
C	TA	G	A	CA	TG	G	A	CG	TT	GT	AT	CT	TC	GC	AC	CC
		A	A			G	G									
A	A	AC	AC	AC	A	AC	AC	AC	A	A	A	A	A	AC	AC	A
C	CT	G	A	CA	CT	G	A	CG	CT	C	C	CC	CT	GC	AC	CC
	A	A	A		G	G	G		T	GT	AT	T	C			C
G	G	GC	GC	GC	G	GC	GC	GC	G	G	G	G	G	GC	GC	G
C	CT	G	A	CA	CT	G	A	CG	CT	C	C	CC	CT	GC	AC	CC
	A	A	A		G	G	G		T	GT	AT	T	C			C
C	CC	CC	CC	CC	CC	CC	CC	CC	C	CC	CC	CC	CC	CC	CC	CC
C	TA	G	A	CA	TG	G	A	CG	CT	GT	AT	CT	TC	GC	AC	CC
		A	A			G	G		T							
C	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT
T	TA	G	A	CA	TG	G	A	CG	TT	GT	AT	CT	TC	GC	AC	CC
		A	A			G	G									
G	GT	GT	GT	GT	GT	GT	GT	GT	GT	GT	GT	GT	GT	GT	GT	GT
T	TA	G	A	CA	TG	G	A	CG	TT	GT	AT	CT	TC	GC	AC	CC
		A	A			G	G									
A	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT
T	TA	G	A	CA	TG	CG	A	CG	TT	GT	AT	CT	TC	GC	AC	CC
		A	A				G									
T	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
T	TA	G	A	CA	TG	G	A	CG	TT	GT	AT	CT	TC	GC	AC	CC
		A	A			G	G									
T	TG	TG	TG	TG	TG	TG	TG	TG	TG	TG	TG	TG	TG	TG	TG	TG
G	TA	G	A	CA	TG	G	A	CG	TT	GT	AT	CT	TC	GC	AC	CC
		A	A			G	G									
A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
G	GT	G	G	GC	GT	G	G	GC	GT	G	G	G	GT	G	G	G
	A	G	A	A	G	G	A	G	T	GT	AT	CT	C	GC	AC	CC
		A	A			G	G									
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	GT	G	G	GC	GT	G	G	GC	GT	G	G	G	GT	G	G	G
	A	G	A	A	G	G	A	G	T	GT	AT	CT	C	GC	AC	CC
		A	A			G	G									
C	C	CG	CG	CG	C	CG	CG	CG	C	C	C	C	C	CG	CG	C
G	GT	G	A	CA	GT	G	A	CG	GT	G	G	G	GT	GC	AC	G
	A	A	A		G	G	G		T	GT	AT	CT	C			CC

CA	CATA	CAGA	CAAA	CACA	CATG	CAGG	CAAG	CACG	CATT	CAGT	CAAT	CACT	CATC	CAGC	CAAC	CACC
GA	GATA	GAGA	GAAA	GACA	GATG	GAGG	GAAG	GACG	GATT	GAGT	GAAT	GACT	GATC	GAGC	GAAC	GACC
AA	AATA	AAGA	AAAA	AACA	AATG	AAGG	AAAG	AACG	AATT	AAGT	AAAT	AACT	AATC	AAGC	AAAC	AACC
TA	TATA	TAGA	TAAA	TACA	TATG	TAGG	TAAG	TACG	TATT	TAGT	TAAT	TACT	TATC	TAGC	TAAC	TACC

4. PERFORMANCE ANALYSIS

The proposed DNA computing algorithm has been differentiated with the results of the current techniques of Advanced Encryption Standard (AES), Rivest-Shamir-Adleman (RSA), DNA based playfair cipher, and DNA based vigenere cipher. The exploratory comparison shows that the proposed cryptosystem is the foremost compelling encryption strategy to be combined with DNA steganography concerning the greatest sizes of bits that can be implanted for hiding capacity and runtime of the cover objects. The proposed cryptosystem was assessed on the same seven measurements that received in the NCBI database [10] [11]. As demonstrated in fig. 2, 3 the x-axis represents DNA sequences of a different length utilized for information covering up through base pair (bp). In fig. 2, y-axis represents the reference DNA sequence of seven measurements. In fig. 3, y-axis represents the information hiding execution time for the estimate of 10kb.

5. NUMERICAL RESULTS

In proposed strategy, the hiding capacity is expanded to exceed 2.4 bits per nucleotide (bpn). The current strategy hides 1.5bpn, which is 40% less than the proposed strategy. For example, by utilizing the current procedure, a reference DNA of 181,432bp can cover up a message of length up to 35.05 kilo bytes (kb) whereas the proposed one can hide up to 57.03kb. The proposed methods also illustrates its dominance over the AES cipher and the hiding capacity of the RSA cipher, as these conventional strategies require extra padding bits in their encryption calculation, so the AES cipher can cover up to 47.02kb and the RSA cipher can cover up to 42.08kb by utilizing the same length of reference DNA sequence 181,432bp.

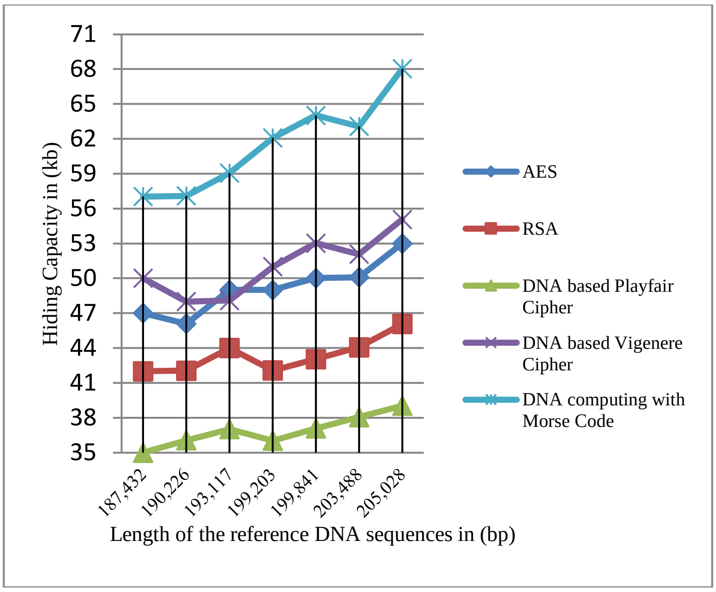

Fig. 2. Comparison of the proposed method with current techniques using hiding capacity

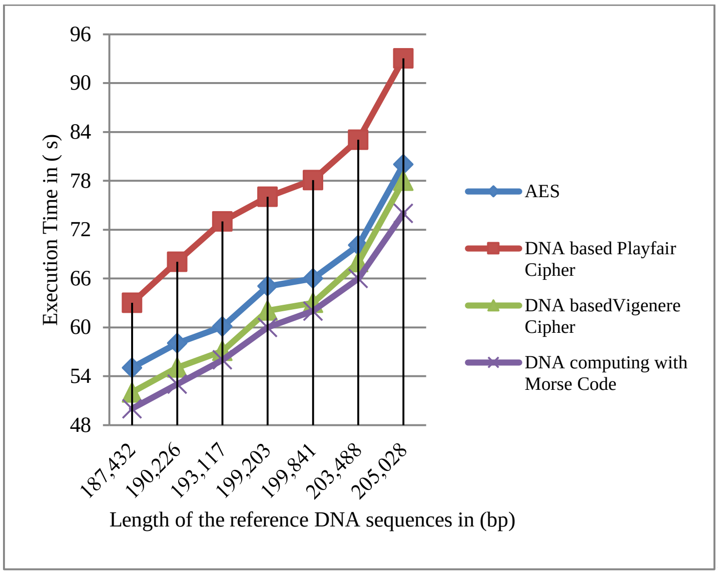

Fig. 3. Comparison of the proposed method with current techniques using execution time

In vigenere cipher there is no need of additional bits, where it can hide up to 50.06kb. Therefore, it is obvious from this calculation that as the length of the reference DNA sequence increases, so does the hiding capacity of any cipher technique will also be increased. In fig. 3, four outlined curves reflect the output of all the comparative ciphers except the RSA cipher as it absorbs very high time complexity that could not be integrated with the other ciphers in the same graph. It is clearly noted that the proposed algorithm is better than current techniques when compared to the execution time.

6. CONCLUSION

A novel and unique organic simulation based procedure for DNA computing has been created in this paper, which satisfies all the functional and non-functional traits that ought to be characteristic of an encryption algorithm based on DNA computing. Performance examination appears the algorithm quality through hiding capacity and execution time.

REFERENCES

[1] N. Kar, K. Mandal, B. Battacharya, Improved Chaos-Based Video Steganography using DNA Alphabets, ICT Express, (2018), https://doi.org/10.1016/j.icte.2018.01.003

[2] U.N. Hussain, T. Chithralekha, A Novel DNA Encoding Technique and System for DNA Cryptography, India Patent 5107, CHE, 2012.

[3] A. Agrawal, A. Bhopale, J. Sharma, M.S. Ali, D. Gautam, Implementation of DNA algorithm for secure voice communication, *International Journal of Scientific & Engineering Research.* 3 (2012).

[4] K. Ning, A Pseudo DNA Cryptography Method, (2009). http://arxiv.org/abs/0903.269

[5] M. Sabry, M. Hashem, T. Nazmy, Three Reversible Data Encoding Algorithms based on DNA and Amino Acids Structure, International Journal of Computer Applications, 54 (2012) 0975 – 8887.

[6] S. Sadeg, M. Gougache, N. Mansouri, H. Drias, An encryption algorithm inspired from DNA, In: Proceedings of the International Conference on Machine and Web Intelligence. IEEE, 2010, 344-349.

[7] Zhang, Qiang, Wang, Qian, Wei, Xiaopeng, A Novel Image Encryption Scheme based on DNA Coding and Multi-Chaotic Map, Advanced Science Letters, 3 (2010) 447-451.

[8] A. Murugan, R. Thilagavathy , Securing cloud data using DNA and Morse code: A triple encryption scheme, International Journal of control theory and applications. Science press. 10 (23) 2017 31-38.

[9] B. Adithya, G. Santhi, Bio-inspired Deoxyribonucleic Acid based data obnubilating using Enhanced Computational Algorithms, In: Proceedings of the International Conference on Computer Networks, Big Data and IoT. Springer, 2020, 597-609.

[10] A. Khalifa, A. Atito, High-capacity DNA-based steganography, In: The 8thInternational Conference on INFOrmatics and Systems, 2012.

[11] NCBI database. [link]. URLhttp://www.ncbi.nlm.nih.gov.