J Med Syst (2018) 42:17
https://doi.org/10.1007/s10916-017-0851-z

IMAGE & SIGNAL PROCESSING

DNA Cryptography and Deep Learning using Genetic Algorithm with NW algorithm for Key Generation

Shruti Kalsi[1,2] · **Harleen Kaur**[1] · **Victor Chang**[3]

Received: 15 August 2017 / Accepted: 25 October 2017

Abstract Cryptography is not only a science of applying complex mathematics and logic to design strong methods to hide data called as encryption, but also to retrieve the original data back, called decryption. The purpose of cryptography is to transmit a message between a sender and receiver such that an eavesdropper is unable to comprehend it. To accomplish this, not only we need a strong algorithm, but a strong key and a strong concept for encryption and decryption process. We have introduced a concept of DNA Deep Learning Cryptography which is defined as a technique of concealing data in terms of DNA sequence and deep learning. In the cryptographic technique, each alphabet of a letter is converted into a different combination of the four bases, namely; Adenine (A), Cytosine (C), Guanine (G) and Thymine (T), which make up the human deoxyribonucleic acid (DNA). Actual implementations with the DNA don't exceed laboratory level and are expensive. To bring DNA computing on a digital level, easy and effective algorithms are proposed in this paper. In proposed work we have introduced firstly, a method and its implementation for key generation based on the theory of natural selection using Genetic Algorithm with Needleman-Wunsch (NW) algorithm and Secondly, a method for implementation of encryption and decryption based on DNA computing using biological operations Transcription, Translation, DNA Sequencing and Deep Learning.

Keywords Cryptography · DNA cryptography · Deep learning · DNA computing · Genetic algorithm · Needleman-Wunsch algorithm (NW) algorithm

This article is part of the Topical Collection on *Image & Signal Processing*

✉ Harleen Kaur
harleen.unu@gmail.com

Shruti Kalsi
shruti.kalsi88@gmail.com

Victor Chang
Victor.Chang@xjtlu.edu.cn

[1] Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi, India

[2] Xerox Technology Ltd., 5th-6th Floor, Vatika Business Park, Gurugram, Haryana, India

[3] International Business School Suzhou, Xi'an Jiaotong Liverpool University, Suzhou, China

Introduction

The developments in the area of information security have been ever emerging. All the efforts in the development of better strategies for information security aim towards three basic outcomes; information availability, integrity, and confidentiality [1–3].

Cryptography is not only a science of applying complex mathematics and logic to design strong methods to hide data called as encryption but also to retrieve the original data back, called decryption. Deep learning is the method which is incorporated for the encryption and decryption process for the cryptosystem. Deep learning is the approach where we encrypt and decrypt data through various layers [4]. Cryptography had been in a silent existence for thousands of years, however, systematic study of cryptology as a science just started around one hundred years ago and deep learning is the latest science come into existence. Fast forwarding to the present, a lot of symmetric and asymmetric ciphers are used for encryption and decryption today in the modern world [5–8]. However, researchers have been trying to find a new computing model,

Published online: 05 December 2017

in order to meet the requirements of the large amount of operation and storage, which can create an entirely new concept and method of information processing and DNA cryptography along with deep learning can play an effective role to meet these requirements [9–11].

DNA Cryptography with deep learning is considered as a next potential medium for data storage and protection, mainly due to its huge storage capabilities and vast parallelism. DNA computing is able to work in parallel to solve the computing problems and deep learning for the better performance. The cryptographic work using DNA computation can extend to various other operations like DNA annealing, DNA synthesis, so on and so forth. This illustrates that DNA molecules can be used for non-biological purposes as well, extending its scope to the digital world.

A new method for symmetric encryption and decryption of the text based on the structure, molecular properties, and biological operations on DNA is presented in the paper. The encryption method is conceptually based on the process of transcription and translation, which are the two biological operations for replicating the DNA and then translating DNA into protein respectively. Since there is no molecular process of converting protein to DNA, in this paper the decryption is logically done by reverse translation and reverse transcription respectively and the concept of deep learning is used for encryption and decryption of DNA sequence. DNA Cryptographic algorithm is designed and implemented to be practically used at digital level, but not at molecular level. In this paper we have created a cryptographic algorithm for key generation. Furthermore, deep learning has been used for better performance of the algorithm, to encrypt and decrypt the data which is in terms of a DNA sequence.

As discussed in Section 2, information represented using DNA uses four character nucleotide bases, $\sum$ = (A,C,T,G), which implies that for a given set of 100 words, there are $\sim 2^{22}$ combinations as compared to the use of $\sum$ = (0, 1) in traditional computers forming only $\sim 2^{13}$ combinations. Hence, the paper deduces a very strong encryption algorithm, whose results are nearly unbreakable and indecipherable. The computations in the classic cryptographic algorithms give rise to a lot of storage issues as billions of computers are working in processing a billion calculations per second. These huge calculations also maximize the power requirements. Also, the storage of keys used in these algorithms is a problem which has started to raise alarms. This may not be a fatal issue now, but it is going to be in the near future. In theory, a single strand of DNA contains ~12 $X10^9$ nucleotide bases. Digitally, 1 byte (8-bits) of binary data can be represented by 4 of nucleotide bases that is 1 nucleotide base for 2 bits of data. This implies a single strand of DNA can contain up to 3000 $X\ 10^6$ bytes or 3000 MB of data. In this paper, the string used for encryption "*eucalyptus is a plant*" requires just 84 nucleotide bases. Hence, DNA computing may become an answer to this in future.

In the proposed work, a totally random and nearly unbreakable key is generated from Genetic Algorithm, an algorithm based on the process of natural selection, coupled with Needleman Wunsch (NW) algorithm which is an algorithm for finding dissimilarity in various DNA strands. When this key and DNA cryptography algorithm are used along with deep learning, it creates a completely new cryptosystem which is completely designed on the biological operations and human behaviour, feasible to be logically implemented as shown in this paper.

DNA Cryptography and Deep Learning

Machine learning is the power given to a machine to understand, learn and resolve the problem in a particular situation without being explicitly programmed and helped. The major key feature of machine learning is to handle a huge amount of data and compute it efficiently and effectively [12–15]. In the past few years, we explicitly programmed the machine to do any specific task. But now in the era of artificial intelligence, we are working to teach the machine to learn by itself through its past experiences or work according to the current situation as a human does in real life.

Deep Learning is a subfield of the machine learning and application of the artificial neural network which is spurred by human brain neural network. In deep learning, the input layer and output layer are linked with a number of hidden layers intermediately [4, 16, 17].

We galvanized the proposed DNA cryptography algorithm with deep learning due to its ability to increase the security and to reduce the complexity of the mathematical equation and formula which are used in the aforementioned algorithm. Deep learning has been used to encrypt and decrypt the data as in Fig. 1.

Today, when the world has become a mesh of networks, data loss is a common occurrence. Deep learning may help to retrieve back such data which is lost in between the network communication [18, 19]. There are different approaches to deep learning like supervised, unsupervised and deep reinforcement learning [20]. When we merge DNA Cryptography and Deep Learning, the power and efficiency of the security automatically increase as both cryptography and deep learning are concrete approaches.

Cryptography, today in the modern world of electronic security technologies has become the foundation to protect valuable information and data. However, the roots of Cryptography go back to the eighteenth century, where the earliest known cryptography found, is some nonstandard hieroglyph [21, 22]. Fast forwarding to the current century, some of the earliest encryption methods included Substitution Ciphers, like the Caesar's Cipher and the Vigenere's Cipher, where the alphabets of the plaintext were

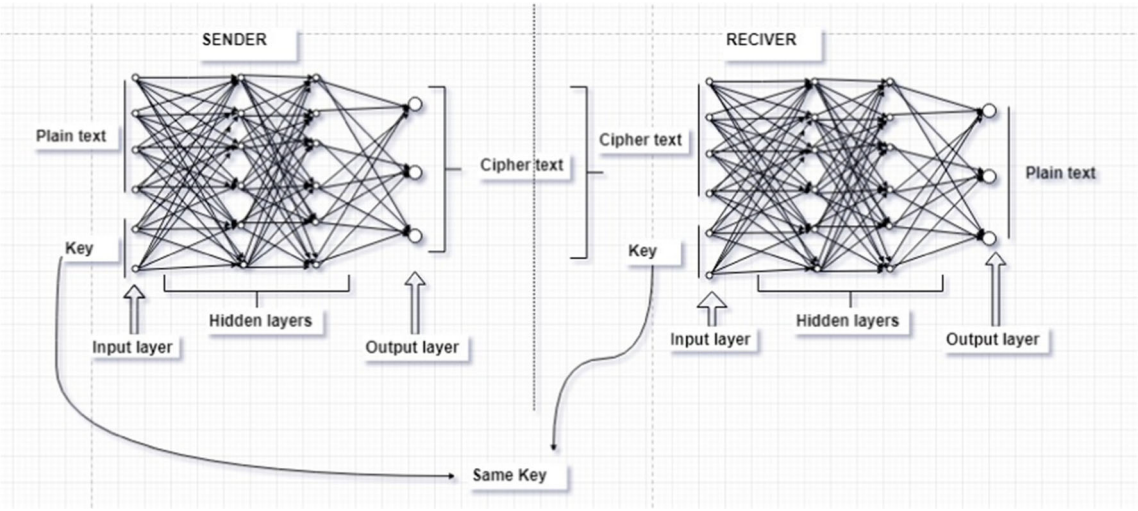


Fig. 1 Deep Learning DNA Cryptography for data encryption and decryption

substituted with some other alphabets. These ciphers were found to be weak and easily breakable, primarily because they contained characteristics of the plaintext language. The contemporary cryptography consists of the following concepts:

i) Plain Text: The original message that needs to be communicated is defined as plain text.
ii) Cipher Text: The message which can only be understood by intended person or system is defined as cipher text
iii) Encryption: the process of converting Plain text into cipher text using a key is defined as encryption.
iv) Decryption: the process of converting cipher text into plain text using the same or different key (as used during Encryption) is defined as decryption.
v) Key: Combination of numeric or alpha numeric text or any special symbol text is referred as key. Key is the most vital part of cryptography as the algorithm depends on the key.

Types of Cryptography Secret Key Cryptography (SKC): Uses a unique single key for both encryption and decryption; also called symmetric encryption.

Some of the examples of symmetric cryptography are the block ciphers like AES and DES. Public Key Cryptography (PKC): Uses one public key for encryption and a private for decryption; also called asymmetric encryption.

The inception of DNA Cryptography has resulted from the quest of finding a new and efficient computing model, in order to meet the requirements of the large amount of operation and storage, which can create an entirely new concepts and methods of information processing. DNA Cryptography is based on DNA computation which takes its inspiration from the biological operations taking place on a DNA molecule. Since its birth in 1994, when Leonard Max Adleman used it to find the solution for Hamilton Path Problem, it has attracted the eyes of many researchers. In the fields of cryptography DNA computing has been described as a possible technology that may bring forward a new hope for unbreakable algorithms [23–26].This was due to the fact that data could be encoded in DNA strands and biological operations can be used in place of cryptographic algorithms on the encoded data to further encrypt the data [27].

DNA, by essence, is an information-storing particle in the living organisms, the genes we pass from one generation to another. It transmits the blueprints for creating the living body. A single strand DNA consists of four different base nucleotides, including adenine (A), thymine (T), cytosine (C) and guanine (G). These are the building blocks of the DNA. Information is stored in strings of this four-lettered DNA code, which when attached to deoxyribose, generate long sequences of information [28–30].

Since in DNA computing, information is represented as four-character genetic alphabet $\sum$ = (A,C,T,G), rather than the binary alphabet used by traditional computers, $\sum$ = (0,1), DNA strands can store much more information than the storage systems used in the traditional computer.

Deep Learning DNA Cryptography is a process of hiding data as long DNA Sequence using DNA computation.

In this paper, the sequence of nucleotides in DNA has been used as.

A 00
C 01
G 10
T 11

In the traditional computer system, B is equivalent to 01100010, which in DNA computing would be equivalent to CGAG.

The biggest advantage of using DNA as the basis of computation is its storage capacity. A gram of DNA contains 1021 DNA bases = 108 Terabytes of data, which can be stored in a very compact form. Also, there is no power required for DNA computing while the computation is taking place. The chemical bonds that are the building blocks of DNA happen without any outside power source.

DNA computation based cryptography was proposed by Adelman, L. in 1994 [1] and DNA model had been used as a basis for cryptography along with present algorithms for encryption and decryption [3, 8, 27, 31]. However, cryptography based completely on DNA and its operations only is still a distant vision. DNA Cryptography definitely challenges the traditional cryptography, both theoretically and technologically. In this paper, we have made an effort to create DNA cryptography completely based on natural DNA processes like translation and transcription, which have been used for encryption. We have proposed an algorithm, logically and conceptually based on these processes in which the information is encrypted exactly the same way as it is encrypted, duplicated and passed on by DNA in human beings. Further, this method has been used along with deep learning to technically provide better security and efficiency to the algorithm.

Pre-processing of the data is indispensable for DNA deep learning cryptography based on DNA computing. This naturally provides a twofold security to the data. DNA provides a great storage medium hence solving both computational and storage problems. It represents how a cryptosystem can be completely designed and implemented on biological processes. In the proposed work, the key used for encryption and decryption, which has been generated through Genetic Algorithm and Needleman-Wunsch (NW) algorithm further ensures the efficiency of the algorithm to produce a strong cipher text.

Using Genetic Algorithm

The Genetic algorithm is based on the Charles Robert Darwin's theory of evolution. It is adaptive heuristic exploration algorithm based on mechanics of the theory of natural selection and natural genetics [32–37]. Based out on the theory of evolution, Genetic Algorithms belong to the family of evolutionary algorithms.

Steps Involved in Genetic Algorithm

Step I: A random generator is used to generate the initial population having P chromosomes. Survival and reproduction of an individual chromosome in the population is promoted by the elimination of useless features and by rewarding useful behavior.

Step II: Reproduction of the population is achieved by iterative application of a set of stochastic operators, which usually consists of Mutation, Cross-Over and Selection.

Step III: Selection is usually based on the fitness function, which calculates the fitness of each individual chromosome in the population, in terms of real number.

Cross-over Crossover is a genetic operator used to induce variation from one generation to another. Crossover rate indicates the likeness of the chromosomes picked up for crossovers [38]. For example, if the crossover rate is 0.7%, this means in a population of 100 chromosomes, chances of crossover is 0.7. This needs to be done since sometimes, crossover can break the symmetry and can fetch undesirable results. Hence, it has to be applied in moderation.

Mutation Mutation is a genetic operator, applied to a set of population to maintain its genetic diversity. This is usually done by altering one or two genes in an individual chromosome. Mutation rate is the chance that a gene within a chromosome will be altered or flipped.

Selection Selection is a process of selecting chromosomes which will mate and recombine to create offspring for the next generation. In order to select a chromosome of having a high chance mating and propagating its feature to the next generation, a selection strategy consisting of fitness function is applied. Therefore, a fitter chromosome over the others in the population is selected, evolving better individuals over time.

Key Generation Using Genetic Algorithm with DNA Computing

In this paper, Genetic Algorithm (GA) forms the basis of key generation which will generate a unique key, used later along with the encoded plain text for encryption [34]. A random number generator is used to generate the initial population of chromosomes. Since in Genetic Algorithm, only the fittest chromosome survives, there is a need to define a fitness function, which will calculate the fitness of the population generated by applying the stochastic operations [39]. In this research, randomness forms the basis of uniqueness for the key, hence, the fitness function to be used should be able to calculate the randomness of an individual chromosome [35, 40].

Initial Population A random number generator is used to generate an initial population of random binary strings, known as chromosomes. Binary strings of 64-bits keys are generated 56 bits (+8 parity bits). In the proposed solution, an initial population of 100 chromosomes is produced randomly. The

population is going to be transformed into a new generation of the population by applying naturally occurring genetic operations as shown in Fig. 2.

Crossover Two offspring chromosomes are created by combining the parent chromosomes at the crossover point. There are several ways of accomplishing the crossover. The one used here is the K-Point Crossover. K-Point crossover uses more than one crossover point to produce two offspring chromosomes.

Two parent chromosomes are selected and a number at random is generated as the numbers of crossover points as shown in Fig. 3.

In the above example, the crossover is at the 1st and 2nd, 3rd and 4th and lastly, 7th and 8th position.

The number of crossovers with a specific crossover rate is determined by Eq. 1:

$$NC = CR^{*}NB^{*}NK/100 \quad (1)$$

Where,

NC number of crossovers
CR crossover rate
NB number of bits
NK number of keys

Mutation Mutation may be defined as a small random tweak in the chromosome, to get a new solution. It is used to maintain and introduce diversity in the genetic population and is usually applied with a low probability. If the probability is high, it is going to leave the algorithm to a random search.

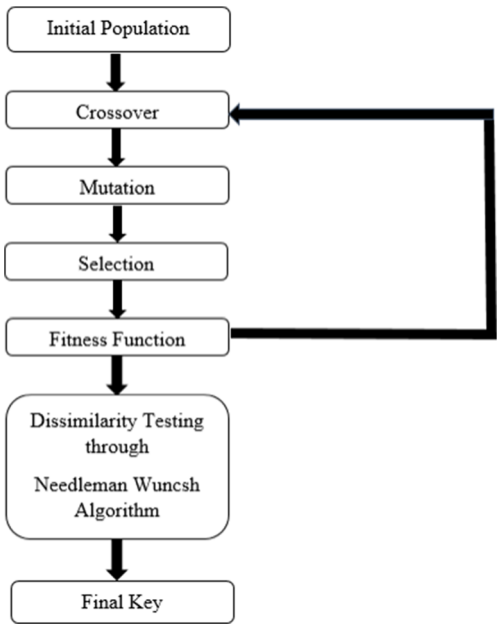


Fig. 2 Flowchart for Key Generation using Genetic Algorithm

A mutation point is selected at random and the bit is flipped, i.e. 0 becomes 1 and 1 become 0 leaving the other bits as it is as in Fig. 4.

The number of mutations with a specific mutation rate is determined by Eq. 2:

$$NM = MR^{*}NB^{*}NK/100 \quad (2)$$

Where.

NM number of mutations
MR mutation rate
NB number of bits
NK number of keys

Fitness Function Fitness function calculates the measures of how fit the resultant chromosome are. The fitness function increases as the algorithm proceeds, indicating that the results are getting better and better. In this paper, Run Test has been used to calculate the randomness in the generated chromosome, since these chromosomes serve the purpose of the cryptographic key. Following this, number of runs for each chromosome is calculated, which is an indicator of randomness.

Run Test Randomness is hard to identify, as it is very difficult to simply look at data and determine whether or not it is random. The runs test is a test of significance or hypothesis test. The procedure for this test is based upon the number of runs, which are sequences of data that have a particular trait. When the observations are more than twenty, then the distribution of the observed number of runs would approximately follow normal distribution. Hence, the fitness is calculated from the formula in Eqs. 3 and 4:

$$f = a - \mu_a/\sigma a \quad (3)$$

Where, a = number of runs.

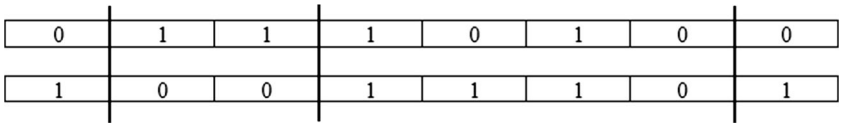
μ_a, σ_a are the mean and variance respectively., calculated by

$$\mu_a = 2N - 1/3 \; and \; \sigma_a = \sqrt{16N - 29/90} \quad (4)$$

Where, $N = 64$.

Final Selection: Needleman- Wunsch (NW) Algorithm Sequence alignment is a way of arranging two or more sequences of characters to identify regions of similarity. Needleman-Wunsch is an algorithm used in the field of biology and bioinformatics to compare biological sequences for similarity. It was developed in 1970 by Saul B. Needleman and Christian D. Wunsch. It uses a scoring system to calculate the degree of similarity or dissimilarity in the two DNA sequences [41]. Ideally, the greater the score, greater is the similarity. In this paper, Needleman- Wunsch Algorithm is used to

Fig. 3 Crossover Operation



calculate the dissimilarity in the two keys. The key with the highest fitness function is matched against each key in the repository. Keys with the minimum score are chosen and XOR-ed. The final key is stored in a final repository, ready to be taken as a key for encryption. The process is iterated over 100 times.

DNA Computing

DNA Sequencing DNA sequencing is the process of determining the sequence of nucleotide bases (As, Ts, Cs, and Gs) in a piece of DNA.

Data Encoding Each English alphabet and numerical digit used in the plain text data is encoded as a DNA sequence. A random sequence generator is used to assign a random sequence of four DNA nucleotides to the alphabets and numerical digits.

A random sequence generator is used to produce a random string of nucleotides. At a given point in time, a random sequence of DNA nucleotides for English alphabets and numeric digits as shown in Fig. 5.

For example, the text **P** = "*eucalyptus*" is encoded in DNA sequence as.

P` = TAGATAGGTTCAACAGCCAGAGGGCGTGCAGATAGGGTGG

After data gets encoded as a DNA sequence, this sequence is converted into binary form.

The sequence to convert DNA nucleotides in binary as shown in Table 1.

Transcription In Biochemistry, the process of transcription begins when an enzyme called RNA polymerase (RNA pol) attaches itself to the template DNA strand and begins to catalyse the production of the complementary RNA, called the mRNA. A copy of a single DNA strand is produced, ready for the process of Translation as shown in Table 2.

In this paper, P′ is the encoded text, representing the DNA strand and the transcript text T, representing the mRNA is produced as a result of complimenting the current encoded text **P`**

Complimentary strand T

P` = TAGATAGGTTCAACAGCCAGAGGGCGTGCAGATAGGGTGG
T = UTCUAUCCAACUUGUCGGUCUCCCGCACGUCUAUCCCACC

Applying the Binary Encoding:

P` = B =
0100110001001111010110000010001110100011001111
1101101111000110001 00
111111011111

T =
1011001110110000101001111101110001011100110000
0010010000111001110 11
0000001000001

This text is divided into blocks of 64-bits.

Translation In Biochemistry, translation is a process of converting the genetic code from its deoxyribonucleic acid form (DNA) consisting of a chain of four repeating letters to a final protein product consisting of amino acids. Protein complex molecules called "*ribosomes*" attach themselves to the modified mRNA strand, produced from transcription and translate the strand into a chain of protein molecule [42]. This is accomplished by transfer RNA (tRNA) molecules, which is the complementary strand to the mRNA, carry specific amino acids to the ribosomes where nucleotides are read and matched with specific amino acids [41, 43].

Steps-wise encryption method:

Step I: Complimentary strand (T) represents the mRNA, which has been divided into blocks of 64 bits each. It further gets divided into blocks of 8 bits each.

mRNA(T) =
1011001110110000101001111101110001011100110000
0010010000111001110 11
0000001000001
Block1 =
1011001110110000101001111101110001011100110000
00100100001110011
Block 2 = 10110000001000001

Step II: Complementary strand tRNA carry the amino acids.

Fig. 4 Mutation Operation

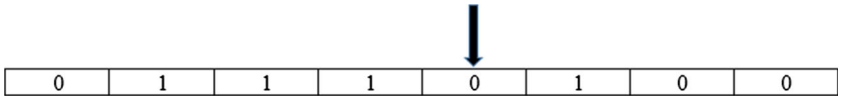


Fig. 5 Nucleotide Encoding Table

Encoding Text using DNA Computation

Assigning Random Sets of Nucleotides to Alphabets(a-z) And Numbers(0-9)

Value of a is CTGC	Value of m is AGAG	Value of y is GGCC
Value of b is TCCT	Value of n is TCAG	Value of z is TGGT
Value of c is TACG	Value of o is TCAC	Value of 0 is AGTG
Value of d is CGGG	Value of p is TAGC	Value of 1 is CGAC
Value of e is GCGG	Value of q is GGAA	Value of 2 is TGCT
Value of f is TGTG	Value of r is TCGA	Value of 3 is GCGG
Value of g is CGGT	Value of s is GTCG	Value of 4 is TTCT
Value of h is AATC	Value of t is ACGC	Value of 5 is CGCT
Value of i is GCAT	Value of u is GTAC	Value of 6 is ATAG
Value of j is ATGC	Value of v is CACG	Value of 7 is TCCG
Value of k is AGCT	Value of w is CGTC	Value of 8 is ACCA
Value of l is TGAG	Value of x is AGAG	Value of 9 is CGCG

Amino Acid represents the cryptography key (K), generated previously, to the mRNA (T).

tRNA =
0100110001001111010110000010001110100011001111111011011100011000100
1111110111110
AminoAcid (K) =
0110110101111001000010010101010111001001110010001010010110101001
Block1 =
0100110001001111010110000010001110100011001111111011011100011000
Block 2 = 01001111110111110

Working with block 1 in rest of the steps. The blocks with less than 64 bits are padded.

Step III: 64 bits block of mRNA, tRNA and amino acid gets divided into blocks of 8 bits each.

mRNA : 10110011 10110000 10100111 11011100 01011100 11000000 01001000 01110011

tRNA : 01001100 01001111 01011000 00100011 10100011 00111111 10110111 10001100

Amino Acid : 01101101 01111001 00001001 01010101 01110010 01110010 00100101 10101001

Step IV: Each byte of the mRNA and amino acid gets converted into decimal form.

mRNA : 10110011 10110000 10100111 11011100 01011100 11000000 01001000 01110011

Decimal Conversion: 179 176 167 220 92 192 72 115

tRNA : 01001100 01001111 01011000 00100011 10100011 00111111 10110111 10001100

Table 1 Random Sequence with DNA nucleotide

Nucleotide	Random Sequence
A	01
C	00
T/U	10
G	11

Table 2 Original Nucleotide with Complimentary Nucleotide

Original Nucleotide	Complimentary Nucleotide
A	U
T	A
C	G
G	C

Right Shift the bits by 2 in each byte

tRNA : 00110001 00111101 01100001 10001100 10001110 11111100 11011110 00110010

Decimal Conversion: 49 61 97 140 142 252 222 50

Step V: Matrix

A Matrix is formed by subtracting the decimal conversion of mRNA and tRNA (whichever is large) and permuting it.

Subtractive Matrix (SM)

130	115	70	80	50	60	150	65

Permuted Matrix (PM)

65	150	60	50	80	70	115	130

This means that 65 becomes the first byte, 150 becomes the second byte and so on.

The bits in the permuted table are again converted back to binary.

The tRNA now becomes:

tRNA : 01000001 10010110 00111100 00110010 01010000 1000110 01110011 10000010

01000001	10010110	00111100	00110010	01010000	01000110	01110011	10000010

+

01101101	01111001	00001001	01010101	01110010	01110010	00100101	10101001

Step VI: The tRNA obtained in Step IV is then bitwise XOR-ed to the amino Acid (K) as per Vernam Cipher. The result is stored as M

Step VII: M is then converted to its hexadecimal equivalent, Cipher Text.

Results

A detailed description and feasibility of the algorithm has been demonstrated and implemented using PHP for the key generation as well as DNA computation. For the key generation, a random initial population of 100 chromosomes was generated and various operators were applied. After applying the crossover operator with the crossover rate as 2.5, the population became 260. Mutation operator was applied to this population with a mutation rate of 0.5 and the final population became 426. The fitness values of all chromosomes were calculated using run test, which is a test for finding the randomness. The mean and variance for each chromosome were calculated and hence, fitness values were obtained. The chromosome with the highest fitness value was tested for dissimilarity against rest all of the chromosomes using the Needleman-Wunsch algorithm. The chromosome with the least score was the most dissimilar with the chromosome having the highest value. These two chromosomes were XOR-ed and a final chromosome was obtained, that was added to the final repository. This was taken as one of the keys fit to be used for encryption and decryption in the algorithm. The whole process was iterated 100 times to get a repository of the most random and non-repeatable cryptographic keys. A maximum of two chromosomes came out to be having the same score of dissimilarity with the chromosome having the highest fitness value. This proves the randomness of the population.

In the DNA computing, a random table of a combination of four DNA nucleotide bases; A, T, C, G was created for A-Z and 0–9. Plaintext was encoded into long sequences of DNA nucleotide bases. This sequence was further converted into binary using a table for nucleotide to binary conversion. This long binary string was divided into blocks of 64 bits. The blocks which were short of 64 bits were padded with NULL. Transcription and Translation operations, with amino acid digitally taken as Key for cryptography, were applied and the plain text was encrypted as shown in Fig. 6. Decryption process followed the exact reverse of the encryption process as in Fig. 7 and the complete process implementation as shown in Fig. 8.

The experimental analysis done here is the analysis of the number of characters/ alphabets taken for encryption and decryption process and the time taken by the algorithm in doing as shown in Table 3.

Discussions

In the investigated study many efforts are being made to find a new computing model, in order to meet the requirements of the large amount of computational operation and storage, which is able to create an entirely new concept and method of information processing and storing. Further, DNA has been identified as a medium with potential for vast storage capacity and sustainability in adverse conditions. Although this is still relatively new field of cryptology, a lot has been researched in this filed which suffices the need to justify that someday it

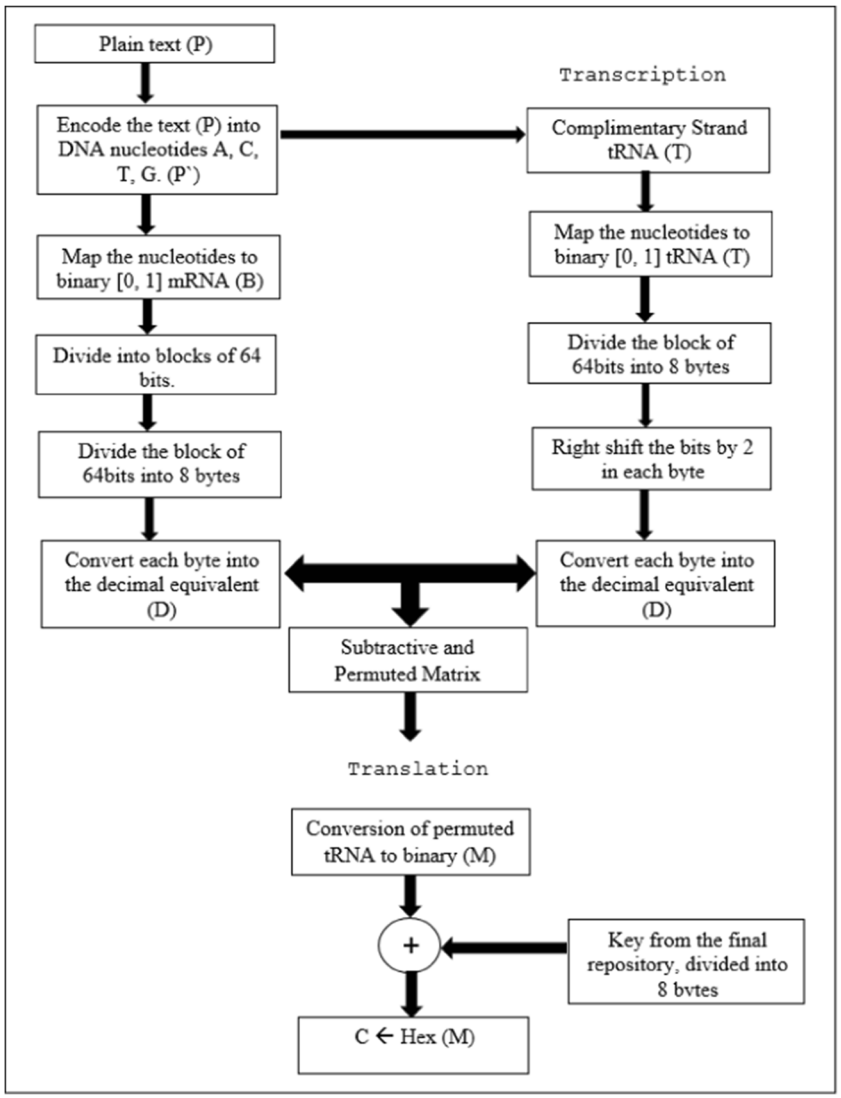


Fig. 6 The Encryption Process

may become the future of cryptography. In this paper which presents an encryption and decryption algorithm based on DNA Computing is performed. The work has been implemented and analysed. The implementation has been done in PHP language. A symmetric block cipher has been used for encryption and decryption based on DNA computing. The data is first encoded into DNA nucleotides and then encrypted. The algorithm is conceptually based on the biological operation, namely; transcription and translation, taking place on the DNA molecule with conjunction with some logical and computational operations to represent it on a digital level. This increases the complexity of the algorithm, making it secure. Detailed description and feasibility of the algorithm has been demonstrated and implemented. Time analysis graph of the algorithm shows its efficiency.

Also, the key used for encryption and decryption has been generated through genetic algorithm. An initial population of 100 chromosomes was taken. The genetic operators involved, when used with fitness function like run test and Needleman – Wunsch algorithm, ensured that the key produces is most random and non-repeating. The whole process was iterated 100 times to get a repository of the most random and non-repeatable cryptographic keys.

Conclusion

DNA cryptography and deep learning are still in its infancy stages with many unsolved problems, but the research on this newborn technology has been rapid and fast. In the proposed work, DNA cryptography and deep learning method is used to encrypt and decrypt the given data. Also, it is considered as a next potential medium for data storage and protection, mainly due to its huge storage capabilities and vast parallelism. Further, DNA computing is able to work in parallel to solve the computing problems. The cryptographic work using DNA

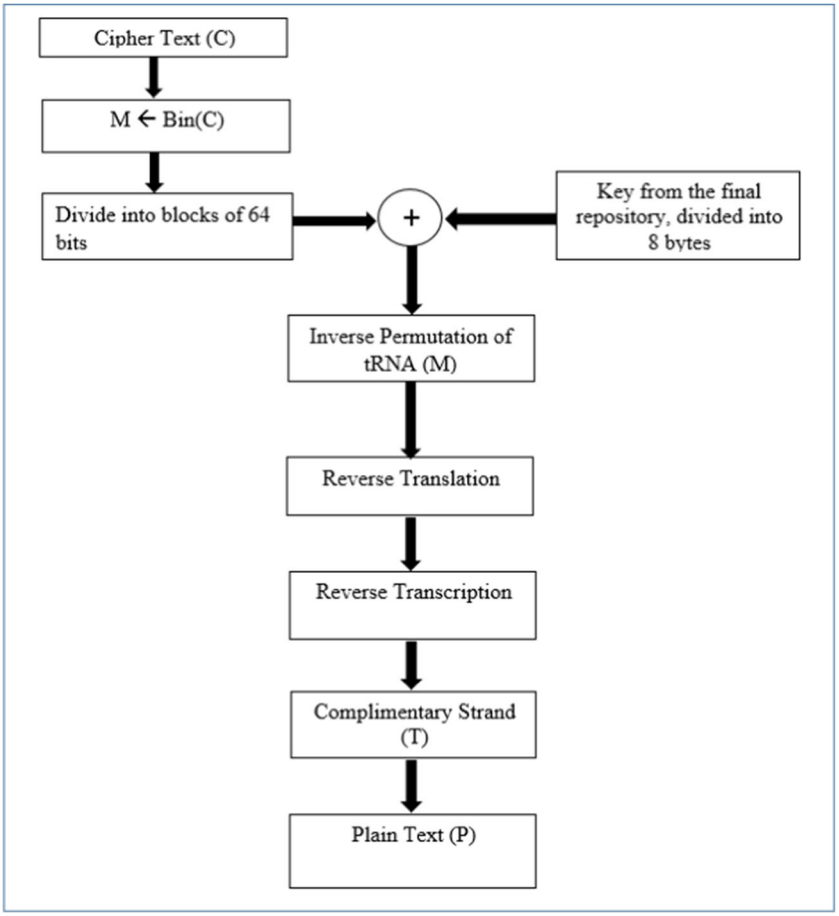


Fig. 7 The Decryption Process

computation can extend to various other operations like DNA annealing, DNA synthesis, so on and so forth. This illustrates that DNA molecules can be used for non-biological purposes as well, extending its scope to the digital world. However, much work is still impending on the ease of use DNA at a scalable laboratory level. At the digital level, this conceptual design is also conjugated with logical operations like XOR and subtraction to create a strong and secure cryptography algorithm. It is designed and implemented to be practically used at digital level, but not at molecular level. A new method for symmetric encryption and decryption of the text based on the structure, molecular properties, and biological operations

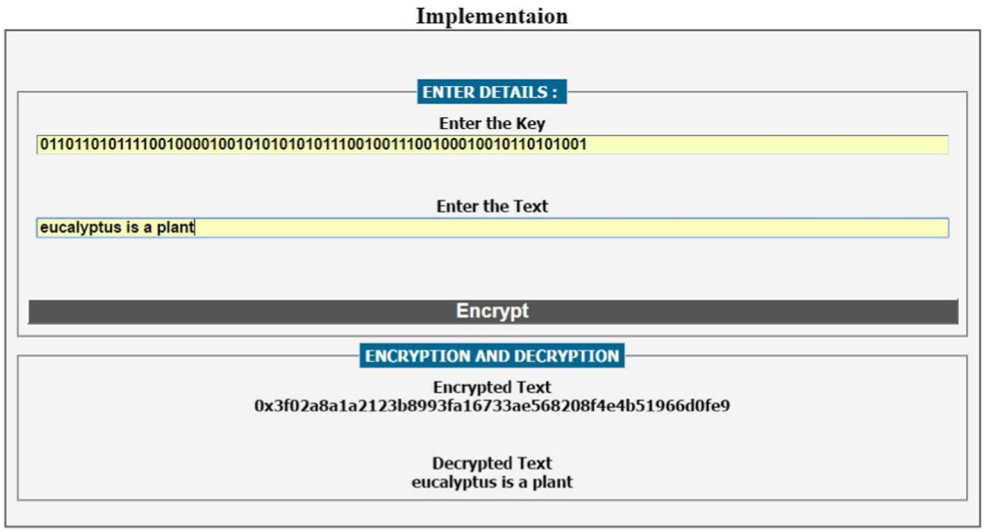


Fig. 8 DNA Encryption and Decryption process implementation

Table 3 Time Taken for DNA Encryption and Decryption process

No. of Characters	Encryption (ms)	Decryption (ms)
500	0.00016401500	0.00026401450
1000	0.00035178900	0.00048178658
1500	0.00067315500	0.00077815800

on DNA and deep learning methods is presented in the paper. The encryption method is conceptually based on the process of transcription and translation, which are the two biological operations for replicating the DNA and then translating DNA into protein respectively. Since there is no molecular process of converting protein to DNA, decryption is logically done by reverse translation and reverse transcription respectively with promising results. The future of DNA Deep Learning Cryptography is pretty much considerable. The proposed work has opened new avenues to research further on amalgamating different fields of Artificial Intelligence like Neural Networks, deep learning with DNA cryptography to succeed over the conventional cryptography algorithms. Moreover, to add another layer of security, the proposed algorithm may be used in conjunction with present cryptographic algorithms. As in the present world of hackers, handling big data of the huge population and securing this big data is a challenge. Hence, emerging techniques like DNA Cryptography and deep learning can play an indispensable role in it. However, there is still a lot needed to be done with regards to the cost and time effectiveness of this research. The research study in this paper is done conceptually. It requires specific skills, cost and time to actually implement this practically in the lab.

Acknowledgements This research work is catalyzed and supported by National Council for Science and Technology Communications (NCSTC), Department of Science and Technology (DST), Ministry of Science and Technology (Govt. of India) for support and motivation [grant recipient: Dr. Harleen Kaur]. The authors gratefully acknowledge financial support from the Ministry of Science and Technology (Govt. of India), India.

Compliance with Ethical Standards

Conflict of Interest None.

References

1. Adleman, L., Molecular computation of solutions to combinatorial problems. *Sci.* 266:1021–1024, 1994.
2. Stallings, W., Network security essentials, Prentice Hall, Fourth edition, 2011
3. Delman, B.,Genetic Algorithms in Cryptography, MS Thesis 2004.
4. Mislovaty, R., Klein, E., Kanter, I. and Kinzel, W. Security of neural cryptography, Proceedings of the 2004 11th IEEE International Conference on Electronics, Circuits and Systems, 2004. ICECS 2004, 2004, pp. 219–221.
5. Anurag Roy and Asoke Nath, "DNA Encryption Algorithms: Scope and Challenges in Symmetric Key Cryptography", IJIRAE 2016
6. John H Reif, Michael Hauser, Michael Pirrung and Thomas LaBean, "Application of Biomolecular Computing to Medical Science: A Biomolecular Database System forStorage, Processing & Retrieval of Genetic Information & Material", Duke University, 2006
7. Junling Sun, "Sequence Splicing Techniques and Their Applications For Information Encryption", International Conference on Advanced Mechatronic Systems, Tokyo, Japan, September I S-21, 2012
8. V. M. M. Shyam, N. Kiran, "A novel encryption scheme based on DNA computing," In 14th IEEE International Conference, Tia, India, Dec. 2007
9. Yunpeng Zhang and Liu He Bochen Fu, "Research on DNA Cryptography", College of Software and Microelectronics, Northwestern Polytechnical University, Xi'an,China
10. Risca, V.I., DNA-based steganography. *Cryptologia, Tylor and Francis*. 25(1):37–49, 2001.
11. Kaur H, Ahmed J, Scaria V, Computational analysis and In-silico predictive modeling for inhibitors of PhoP regulon in S. typhi on high-throughput screening bioassay dataset., *Interdisciplinary Sciences: Computational Life Sciences (a Springer SCI Journal),*2016.
12. Kaur, H., Chauhan, R., Wasan, S. K. A Bayesian network model for probability estimation, Encyclopaedia of Information Science and Technology, IGI Global, Third Edition, 1551–1558, 2015.
13. Kaur, H., Chauhan, R., and Ahmed, Z., Role of data mining in establishing strategic policies for the efficient management of healthcare system–a case study from Washington DC area using retrospective discharge data. *BMC Health Services Research*. 12(S1):P12, 2012.
14. Chauhan, R., Kaur, R. Predictive Analytics and Data Mining: A Framework for Optimizing Decisions with R Tool, Advances in Secure Computing, Internet Services, and Applications, Springer, 73–88, 2014.
15. Kaur, H., Chauhan, R., and Alam, M.A., Spatial Clustering Algorithm using R-tree. *Journal of Computing*. 3(2):85–90, 2011.
16. Hermans, M. and Schrauwen, B. Training and analysing deep recurrent neural networks. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 26, pages 190–198. Curran Associates, Inc., 2013.
17. Chen, C., Xiang, H., Qiu, T., Wang, C., Zhou, Yang., Chang, V. A rear-end collision prediction scheme based on deep learning in the Internet of Vehicles, Journal of Parallel and Distributed Computing, 2017.
18. Liao, X., Yin, J., Guo, S., Li, X., & Sangaiah, A. K. (2017). Medical JPEG image steganography based on preserving interblock dependencies. *Computers & Electrical Engineering*.
19. Zheng, H. T., Wang, Z., Ma, N., Chen, J., Xiao, X., & Sangaiah, A. K. (2017). Weaklysupervised image captioning based on rich contextual information. *Multimedia Tools and Applications*, 1–17.
20. Zhang, R., Shen, J., Wei, F., Li, X., & Sangaiah, A. K. (2017). Medical image classification based on multi-scale non-negative sparse coding. *Artificial Intelligence in Medicine*.
21. Diffie, W., and Hellman, M., New directions in cryptography. *IEEE Transaction on Information Theory*. 22(6):644–654, 1976.
22. EI Gamal T., A public-key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory*. 31(4):469–472, 1985.
23. Borda M. & Tornea O. DNA secret writing techniques [C]. In COMM(2010), Chengdu: IEEE, June 10-12, 2010: 451–456

24. Hongjun Liu, Xingyuan Wang and Abdurahman Kadir, "Image encryption using DNA complementary rule and chaotic maps", ScienceDirect, 2012
25. Martin JAVUREK and Marcel HARAKAĽ, "Cryptography And Genetic Algorithms", Science & Military, 2016
26. Tornea, O., and Borda, M.E., DNA Cryptographic Algorithms, MEDITECH 2009. *IFMBE Proceedings*. 26:223–226, 2009.
27. U.Noorul Hussain, T. Chithralekha and A.Naveen Raj, G.Sathish, A.Dharani, "A Hybrid DNA Algorithm for DES using Central Dogma of Molecular Biology (CDMB)", International Journal of Computer Applications, 2012
28. K. Li, S. Zou, and J. Xv, Fast parallel molecular algorithms for DNAbased computation: Solving the elliptic curve discrete logarithm problem over gf(2 n), Journal of Biomedicine and Biotechnology, Hindawi., vol. 2008, pp. 1–10, Apr. 2008
29. Fastest DNA Computer. Science, 2005, 308: 195
30. Roweis, S., Winfreel, E., Burgoyne, R., et al., A sticker based model for DNA computation. *Journal of Computational Biology*. 5(4): 615–629, 1998.
31. Tornea, O., and Borda, M.E., DNA Cryptographic Algorithms. *IFMBE Proceedings*. 26:223–226, 2009.
32. Goyat, S.: Cryptography Using Genetic Algorithms (GAs). In: IOSR Journal of Computer Engineering (IOSRJCE), 1(5), pp. 06-08 Identification of Common Molecular Subsequences. *J. Mol. Biol.* 147, 195-197. 2012.
33. Mishra, S., and Bali, S., Public key cryptography using genetic algorithm. *Int. J. Recent Technology and Engineering*. 2(2):150–154, 2013.
34. A. J. Bagnall, "The Applications of Genetic Algorithms in Cryptanalysis", School of Information Systems, University Of East Anglia, 1996.
35. Kaur, H., & Tao, X. (Eds.). ICTs and the millennium development goals: A United Nations perspective. New York, Springer, US, 2014.
36. Kaur H, Lechman E and Marszk A (Eds.), Catalyzing Development through ICT Adoption: The Developing World Experience, Springer Publishers, Switzerland, 2017.
37. A. Tragha, F. Omary, A. Kriouile, "Genetic Algorithms Inspired Cryptography", A.M.S.E Association for the Advancement of Modeling & Simulation Techniques in Enterprises, Series D: Computer Science and Statistics, November 2007.
38. A.J. Umbarkar and P.D. Sheth, "Crossover Operators In Genetic Algorithms: A Review", Ictact Journal On Soft Computing, October 2015
39. Watson J D, Hopkins N H, Roberts J W, et al. Molecular Biology of the Gene. 4th ed. Menlo Park, CA: The Benjamin/Cummings Publishing Co., Inc., 1987
40. Taylor, C., Risca, V., and Bancroft, C., Hiding messages in DNA microdots. *Nature*. 399:533–534, 1999.
41. Needleman, S.B., and Wunsch, C.D., A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48:443–453, 1970.
42. Qin Limin. The Study of DNA - Based Encryption Method [D].Zheng Zhou: Zheng Zhou University of Light Industry, 2008.
43. Zhi-min Zhou and Zhong-wen Chen, "Dynamic Programming for Protein Sequence Alignment", International Journal of Bio-Science and Bio-Technology Vol. 5, No. 2, April, 2013