ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

Double Layer Security by DNA Based Cryptography and RSA Algorithm

Smitha Mohan M

Assistant Professor, Dept of Information Technology, Toc H Institute of Science & Technology, Ernakulam, Kerala, India

ABSTRACT: DNA based cryptography is a upcoming branch in cryptographic research and has a wide perspective. In this study, biological concept can be taken into consideration. Two bits can be used to represent each nucleotides. The main purpose behind this work is to discover new fields of encoding the data in addition to the conventional used encryption algorithm in order to increase the concept of confusion and therefore increase security. The fundamental idea behind this encryption technique is to enforce other conventional cryptographic algorithms which proved to be broken, and also to open the door for applying the DNA and Amino Acids concepts to more conventional cryptographic algorithms to enhance double layer security features.

In my work, applied the conversion of character form or binary form of data to the DNA form and then to amino acid form. Then the resulting form goes through the RSA encryption algorithm.

KEYWORDS: cryptography; encryption; DNA; Amino Acids; RSA.

I. INTRODUCTION

The design of encryption algorithms is based on complex problems in order to ensure the security of the algorithm for at least a certain large period of time. The idea behind this process is to increase the complexity of the problem by augmenting its size and achieving this way a system requiring tremendous resources and efforts to attack it. The best way of achieving a robust system is to act on scalability that is, to reach a large scale complexity for the problem. Such task can be made possible and handled by the innovative DNA computing, which allows a very high degree of parallelism on one hand and a huge storage capacity on the other hand.DNA is a nucleic acid that contains the genetic instructions used in the development and functioning of all living organisms and some viruses.NA computing was born with the Adleman's pioneering work[1].With his DNA algorithm for solving the Hamiltonian problem, Adleman set the foundation of the research in the field of bio computing. The vast parallelism and the density of information inherent to DNA in addition to the results of Adleman's experience encouraged many researchers to exploit this molecule to solve hard problems in different axis in computer science. In the field of cryptography, Gehanietal [2] introduces the first algorithm of DNA based cryptography, followed by many others [3][4]. In this work, we are not intended to use real DNA strands to implement my cryptographic algorithm, but only to simulate some mechanisms of the process of the central dogma of molecular biology. This paper is organized as follows: section 3 describes the central dogma of molecular biology .The encryption and decryption algorithm is presented in section 4,section 5 describes implementation details Section 6 analyses the security features and section 7 conclusion.

II. RELATED WORKS

The DNA based encryption algorithms are often proposed, a very few studies simulating processes of central dogma have been conducted. The only works we have found on this subject is a simple cryptographic method based on the same idea and reported in [5], and a symmetric encryption DNA-based algorithm called YAEA that was proposed by Aminet al

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

In [6] and in which a binary form of data, such as plaintext messages, and images are transformed into sequences of DNA nucleotides. Then, efficient searching algorithms are used to locate the multiple positions of a sequence of four DNA nucleotides that represent the binary octet plaintext character within a Canis Familiaris genomic chromosome. Randomly selected pointers of the four DNA nucleotides for each plain text character are assembled in a file that constitutes the ciphered text. This technique can be used to enforce other Conventional cryptographic algorithms. An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

My work is not limited to propose a method simulating mechanisms of the central dogma of biology, but presents a symmetric encryption algorithm designed according the recommendations of experts in cryptography [7] and containing apart called DNA module that simulates critical processes of the central dogma in order to enhance its security.

The DNA double helix is stabilized by hydrogen bonds between the bases attached to the two strands. The four bases found in DNA are adenine abbreviated (A),cytosine (C),guanine (G) and thymine (T).These four bases are attached to the phosphate to form the complete nucleotide, as shown for adenosine monophosphate. The main role of DNA molecules is the long-term storage of information.

III.BIOLOGICAL BACKGROUND

The central dogma of molecular biology was articulated by Francis Crick in a paper published in nature in 1970[8].It states that information cannot be transferred back from protein to either protein or nucleic acid. It also explains the so called general transfers that describe the normal flow of biological information: DNA (DNA replication), DNA RNA (transcription) and RNA protein (translation).

The genetic code consists of 64 triplets of nucleotides. These triplets are called codons. Each codon encodes for one of the 20 amino acids used in the synthesis of proteins.

In my cryptographic algorithm, the principle ideas of the processes of transcription and translation are used. We briefly explain them below.

•*Transcription.* It's the process that creates a new RNA piece by transferring information from a section of DNA strand. A DNA segment is read, non coding areas are removed, and the remaining ones are rejoined and transcribed into a single strand of RNA.

•*Translation.* The RNA sequence is translated, first,into a sequence of amino acids, then into a protein, according to the genetic code table.

The genetic code can be expressed as either RNA codons or DNA codons. RNA codons occur in messenger RNA (mRNA).

In the process of translation, mRNA read the data and it acquires sequence of nucleotides by transcription from the corresponding gene. A gene is a sequence of DNA that contains genetic information and can influence the phenotype of an organism. Within a gene, the sequence of bases along a DNA strand defines a messenger RNA sequence, which then defines one or more protein sequences. The relationship between the nucleotide sequences of genes and the amino-acid sequences of proteins is determined by the rules of translation, known collectively as the genetic code. The genetic code consists of three-letter 'words' called codons formed from a sequence of three nucleotides (e.g. ACT, CAG, TTT).

In transcription, the codons of a gene are copied into messenger RNA by RNA polymerase. This RNA copy is then decoded by a ribosome that reads the RNA sequence by base-pairing the messenger RNA to transfer RNA, which carries

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

amino acids. Since there are 4 bases in 3-lettercombinations,there are 64 possible codons . These encode the twenty standard amino acids, giving most amino acids more than one possible codon.

IV DNA – BASED RSA ALGORITHM

A. RSA encryption algorithm of DNA-based cryptography

RSA used to be applied to English alphabet characters of plaintext. Some algorithm does not encode special characters like symbols and equations. In my algorithm, we can use any numbers, special characters or even spaces in my plaintext. The encryption process starts by the binary form of data like message or image, which is transferred to DNA form according to Table 1. Then the DNA form is transferred to the Amino acids form according to Table 2 which is a standard universal table of Amino acids and their codons representation in the form of DNA.

Note that each amino acid has a name, abbreviation, and a single character symbol. This character symbol is what we will use in my algorithm.

TABLE I [9]

DNA Representation of bits

Binary Value		**DNA Coding**
0	0	A
0	1	C
1	0	G
1	1	T

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

B. Constructing the alphabet table

TABLE III [9]

Standard universal Amino acids

Ala/A	GCU,GCC, GCA,GCG	Leu/L	UUA,UUG, CUU,CUC, CUA,CUG
Arg/R	CGU,CGC, CGA,CGG, AGA,AGG	Lys/K	AAA,AAG
Asn/N	AAU,AAC	Met/M	AUG
Asp/D	GAU,GAC	Phe/F	UUU,UUC
Cys/C	UGU,UGC	Pro/P	CCU,CCC, CCA,CCG
Gln/Q	CAA,CAG	Ser/S	UCU,UCC, UCA,UCG, AGU,AGC
Glu/E	GAA,GAG	Thr/T	ACU,ACC, ACA,ACG
Gly/G	GGU,GGC, GGA,GGG	Trp/W	UGG
His/H	CAU,CAC	Tyr/Y	UAU,UAC
Ile/I	AUU,AUC, AUA	Val/V	GUU,UAC, GUA,GUG
START	AUG	STOP	UAA,UGA, UAG

In the table2, we have only 20 amino acids in addition to 1 start and 1 stop. We have to use only 20 letters from this table. Remaining letters such as B,J,O,U,X,Z can be constructed from the existing table. I and J assigned to one cell. The start codon is repeated with amino acid (M) so we will not use it. We will assign to (B) the 3 stop codons. We have 3 amino acids (L,R,S) having 6 codons. By noticing these sequence of DNA of each, we can figure out that each has 4 codons of the same type and 2 of another type. Those 2 of the other type are shifted to the letters (O,U,X) respectively. Letter(Z) will take one codon from (Y), so that Y: UAU, Z:UAC. Now the new distribution of codons is illustrated in Table 3.

Counting the number of codons of each character, we will find the number varies between 1 and 4 codons per character. We will call this number 'Ambiguity' of the character[AMBIG].Now we have the distribution of the complete English alphabet, so a message in the form of Amino Acids can go through Traditional RSA cipher process using the secret key.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

The output form is the amino acid form of cipher text. DNA form of cipher text can be demonstrated also from Table 4 choosing random codons accompanied to each character and Fig.1 shows all working of encryption algorithm. The concept that one character can have more than one DNA representation is itself an addition to confusion concept that enhances the algorithm strength. Table 4 shows the new distribution of codons on the amino acids and additional by alphabetical English letters according to my algorithm.

C. Decryption and Ambiguity problem

The decryption process is simply the inverse of the encryption process .We will find a problem in constructing the DNA form of plaintext from the amino acid form which is of length (L). The problem is that we are unable to choose which codon to put in accordance to each amino acid character. This is simply the problem of codon-amino acid mapping problem arised with other algorithm based on the concept of Central Dogma like [4].

The solution in my algorithm is located in two additional bits for each amino acid character to demonstrate which codon to choose. We said before that each amino acid has 1, 2, 3 or 4 codons to represent it. This is a number that can be put in 2 bits from 0 3.These 2 bits can be converted to DNA form from Table 1.That is why the final cipher text is both the DNA form of cipher text of length (3L) and the array carrying the ambiguity of length (L).In decryption, the amino acid form of plaintext with the assistance of the ambiguity array can construct the correct formof plaintext in DNA form which can be transferred to binary form and then the final character form.

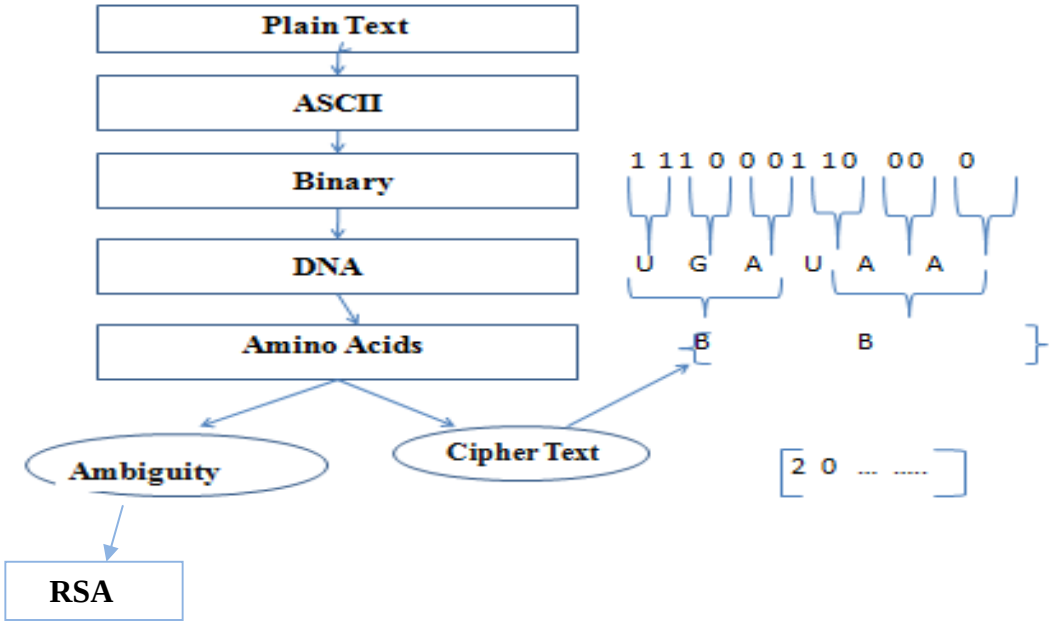

Fig.1 Flowchart of the DNA based RSA algorithm

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

D. Pseudo-code

Input:

[P] Plaintext (characters with spaces, numbers or any special characters).
[K] Secret key (English characters without any number or special characters).

Algorithm body:

Preprocessing:

1- Prepare the secret key:
- Remove any spaces or repeated characters from [K].
- Put the remaining characters in the UPPER case form. [K]⟶UPPER[K].
2- Prepare the plaintext:
- Remove the spaces from [P] (done to avoid attacker's trace to a character which is repeated many times within the message)

Processing:

1- Binary form [BP] = BINARY [P] (Replace each character by its binary representation-8 bits-)
2- DNA form [DP] = DNA [BP] (Replace each 2 bits by their DNA representation)
3- Amino acids form [AP] = AMINO [DP] (Replace each 3 DNA characters by their Amino acid character keeping in track the ambiguity of each Amino acid [AMBIG].
4- Do RSA encryption process in [AMBIG] and add [K] .
5- DNA form of cipher text [DC] = DNA [AC]+RSA[AMBIG].

Output:

Add [DC] and RSA [AMBIG] together in the suitable Form ⟶ final cipher text [C].

V. IMPLEMENTATION

Experiment preprocessing:

1- Loading the table of the 64 amino acids with their DNA Encodings and number of ambiguous encodings.
2- Formatting the secret key by removing spaces, repeated characters and non English letters.
3- Formatting the plaintext by removing spaces between words and separating the repeated doubles by the character '~' which chosen to be a rarely used character.

Processing:

This includes:

1- Converting characters to binary form.
2- Converting binary to DNA
3- Converting the DNA to amino acids and recording ambiguity.
4- Do RSA encryption.
5- Convert the amino acid form of cipher text to DNA form in addition to embedding the ambiguity in the DNA format.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

TABLE IIIV [9]

New Distribution for codons on English alphabet

A	GCU,GCC, GCA,GCG	L	UUA,UUG, CUU,CUC
R	CGA, CGG, AGA,AGG	K	AAA,AAG
N	AAU,AAC	M	AUG
D	GAU,GAC	F	UUU,UUC
C	UGU,UGC	P	CCU,CCC, CCA,CCG
Q	CAA,CAG	S	UCU,UCC, AGU,AGC
E	GAA,GAG	T	ACU,ACC, ACA,ACG
G	GGU,GGC, GGA,GGG	W	UGG
H	CAU,CAC	Y	UAU,UAC
I/J	AUU,AUC, AUA	V	GUU,UAC, GUA,GUG
B	UAA,UGA, UAG	O	CUA,CUG
U	CGU,CGC	X	UCA,UCG
Z	UAU,UAC		

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

TABLE IVII

New Distribution of English alphabet with the corresponding new codons

	STOP										FROM			TO			FROM	FROM		TO			TO	FROM	TO
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
4	3	2	2	2	2	4	2	3		2	6	1	2		4	2	6	6	4		4	1		2	
GCU	UAA	UGU	GAU	GAA	UUU	GGU	CAU	AUU	AUU	AAA	UUA	AUG	AAU	CUA	CCU	CAA	CGU	UCU	ACU	CGU	GUU	UGG	UCA	UAU	UAU
GCC	UAG	UGC	GAC	GAG	UUC	GGC	CAC	AUC	AUC	AAG	UUG		AAC	CUG	CCC	CAG	CGC	UCC	ACC	CGC	GUC		UCG	UAC	
GCA	UGA					GGA		AUA	AUA		CUU				CCA		CGA	UCA	ACA		GUA				
GCG						GGG					CUC				CCG		CGG	UCG	ACG		GUG				
											CUA						AGA	AGU							
											CUG						AGG	AGC							
4	3	2	2	2	2	4	2	3	3	2	4	1	2	2	4	2	4	4	4	2	4	1	2	1	1

VI. ADDITIONAL SECURITY FEATURES

By various modes of operation happen in serial fashion and provide double layer (Bi-layer) security and therefore it is called as Multi-layer DNA Encryption Algorithm shown in Fig.2. Firstly the plain text is encrypted to DNA CODE using DNA Cryptographic Algorithm then it is encrypted using convention RSA algorithm.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

One of the advantages of this algorithm is the variety of ways we can use to write down the cipher text. It can be written in DNA form, binary form or even character form which is more confusing. The advantage of DNA form is that it can make use of several steganography techniques developed for DNA messages [3]. It can also be prepared in biological labs like in [2] in which DNA message goes through a biological DNA encryption process.

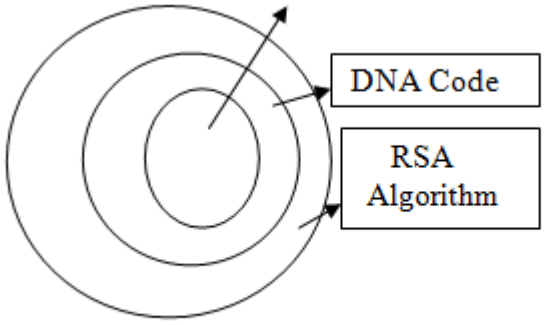

Fig.2 Bi-layer security

Some characters in table 2 can have 6 codons representing the problem of ambiguity. The way we handled the preparation of table 3 made each character have in maximum 4 codons. The number 4 can be represented by 2 bits and therefore ban be represented by one DNA character. That was a benefit that made us able to write the cipher text with ambiguity in the form of DNA.

VII. CONCLUSION

The fundamental idea behind this technique is to open the door for the idea of applying the DNA and Amino Acids encoding concepts to other conventional cryptographic algorithms to enhance their security –vulnerability features. My algorithm initially succeeded in overcoming some main problems in "RSA". As in my algorithm the plaintext is to converted to its binary value before encryption, it now clear that the plaintext message can be written in upper or low case, with any punctuation, and numerical values

REFERENCES

[1] Sherif T. Amin, Magdy Saeb, Salah El-Gindi, "A DNA-based Implementation of YAEA Encryption Algorithm," IASTED International Conference on Computational Intelligence (CI 2006), SanFrancisco, Nov. 20, 2006.
[2] Ashish Gehani, Thomas LaBean and John Reif. DNA-Based Cryptography. DIMACS DNA Based Computers V, American Mathematical Society, 2000.
[3] TAYLOR Clelland Catherine, Viviana Risca, Carter Bancroft, 1999,"Hiding Messages in DNA Microdots". Nature Magazine Vol.. 399, June 10, 1999.
[4] KANGNing "A Pseudo DNA Cryptography Method", Independent Research Study Project for CS5231, October 2004. Leonard Adleman. "Molecular Computation of Solutions to Combinatorial Problems". Science, 266:1021-1024, November 1994.
[5] Leonard Adleman. "Molecular Computation of Solutions to Combinatorial Problems". Science, 266:1021-1024, November 1994.
[6] Dan Boneh, Cristopher Dunworth, and Richard Lipton. "Breaking DES Using a Molecular Computer". Technical Report CS-TR-489-95, Department of Computer Science, Princeton University, USA, 1995.
[7] *Dominik.* Heider and Angelika Barnekow, "DNA-based watermarks using the DNA-Crypt algorithm", Published: 29 May 2007 BMC Bioinformatics 2007, 8:176 doi:10.1186/1471-2105-8-176, http://www.biomedcentral.com/1471-2105/8/176 ,© 2007 Heider and Barnekow; licensee BioMed Central Ltd.
[8] *William Stallings. "Cryptography and Network Security", Third Edition, Prentice Hall International, 2003.*
[9] *Mona Sabry,Mohammed Hashem,Taymoor Nazmy, " A DNA and Amino Acid –Based Implementation of Playfair Cipher",IJCSIS ,Vol.8.NO.3.2010.*