Enhanced Double Layer Security using RSA over DNA based Data Encryption System

Meettu Skariya
Department of Information Technology
Rajagiri School of Engineering and Technology
Kochi, India
meettuskariya@gmail.com

Mariam Varghese
Department of Information Technology
Rajagiri School of Engineering and Technology
Kochi, India
mariamv@rajagiritech.ac.in

Abstract— In this paper we propose an enhanced algorithm to communicate data securely for communication and information security. The DNA cryptography is a new and promising area in cryptography. Here we propose techniques that use the objectives of cryptographic encryption and steganography i.e. Encryptions to make the message unreadable to anyone except the recipient and steganography is to hide the existence of the message so that we can provide more security. This is based upon DNA and amino acids-based structure to the core of the ciphering process. In the proposed scheme we have three phases. In the first phase we encrypt the information using DNA and amino acids cipher text and the secure data is formed. And in the second phase we hide the formed DNA (cipher text) using reference DNA by insertion technique. In the third phase we encrypt the key generated in phase 1 with conventional RSA algorithm. This technique can be done on any type of data and for applying this technique we have to convert the data into binary form. And to get the original information decrypt it by reversing the process. With the proposed algorithm we can easily make a cipher text with less time complexity, provides more security, and as properties of DNA is used it can be used for storage of large data in small DNA(cipher text) that provide compactable storage. When we decrypt the data we get the exact input without any junk values. It also ensures double level of security by incorporating RSA algorithm along with that of DNA, RSA is a powerful encryption algorithm.

Keywords-*:* DNA, amino acid, ambiguity, RSA, steganography, encryption

I. INTRODUCTION

In this century of rapid advances, information explosion in which information has become a very important strategic resource, and the information security has become increasing important. [2] Cryptography is the most important component in the communication security and computer security. And these helps in providing security and confidentiality to data that prevents others from reading the data. [5] There are several techniques but here we propose a DNA based cryptography that uses the structure of DNA and amino acids which is a Modern cryptography encryption technology. These techniques depend on the high randomness of the DNA to hide any message without being noticed. [4] DNA has many characteristics which make it a perfect Data hiding media. DNA has tremendous information storage capacity [1]. In addition, any DNA sequence can be synthesized in any desirable length. In cryptography, encryption is the process of encoding data in such a way that eavesdroppers or hackers cannot read it, and only the authorized parties can read. Encryption is basically to scramble data so that it is unreadable to persons that don't know the key or methods by which it was made unreadable. And decryption is to get the original information using the reverse process. [6] Steganography is the art and science of writing hidden information in such a way that no one apart from the intended recipient knows of the existence of the information i.e. data hiding. [7]

In molecular biology DNA, i.e. deoxyribonucleic acid, is a nucleic acid that contains the genetic instructions. The main role of DNA molecules is the long-term storage of information. DNA is often compared to a set of code, since it contains the instructions needed to construct other components of cells, such as proteins. DNA has a double helix shape, twisted into a spiral. Each step of the ladder is a pair of nucleotides DNA is made of four types of nucleotide. Nucleotide are Adenine (A), Thymine (T), Cytosine (C), Guanine (G).Information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T).Genetic code by which DNA stores the genetic information consists of "codons" of three nucleotides. With

four possible bases, the three nucleotides can give 4^3 = 64 different possibilities, and these combinations are used to specify the different amino acids used by living organisms [1]. And the structure of DNA and amino acid is used in this paper.

DNA based cryptography is shown to be very effective. Currently, several DNA based algorithms are proposed for quite some cryptography and also steganography problems and they are very powerful in these areas. In this paper a significant modification to the DNA-based and amino acids based structure to the core of the ciphering process. We have to convert the original data into a binary form of data, such as plaintext messages, or images are transformed into sequences of binary data which is later on converted to DNA nucleotides. Subsequently, these nucleotides pass through an encryption process based on amino-acids structure along with an ambiguity, taken as secret key. Then this amino acid structure is again encrypted to DNA nucleotides provides more security [10] and convert the data into a desirable length. After the encryption we have to provide data hiding for that we use the insertion technique. Insertion technique me make a fake DNA (cipher text) by mixing up the formed DNA sequence with the reference DNA sequence. The proposed DNA- based cryptography encryption method combined with data hiding techniques for an increased level of security. Along with these we can use conventional RSA algorithm which provides high level of security and make the data more secure. For that we encrypt the ambiguity (secret key) that shows the position of amino acid with the generated public key and we form the encrypted secret key.

II. RELATED WORK

Based on the investigation conducted by Yunpeng Zhang [2] on a conventional symmetric cryptographic encryption algorithm .Hence introduces the concept of using DNA structure in the field of cryptography in order to enhance the security of cryptographic algorithms.[2] Along with these we focus on the efforts of Sherif T[3] in using conventional Playfair cipher for encryption of the data. But using these Playfair cipher we cannot get the actual input because when decrypting the cipher text because it consist of the junked data in it as it cannot represent space and other symbols, it can be used only for alphabets. Above two are considered a pioneering idea behind this paper. Thus introduces a modification to the Playfair cipher encryption used by [3] and use Biological concepts such as DNA and amino acids structures to improve from Playfair encryption. So it helps in secure communication and will get the same data after decryption. RSA algorithms are conventional asymmetric algorithm that uses pair of keys and provides high level security and confidentiality for our data. And because of that we use the RSA algorithm to encrypt the secret key.

III. METHODOLOGY

Numerous algorithms have been identified for DNA based cryptography. The proposed algorithm for data security uses simple encryption technique based on DNA structure and instead of using conventional encryption we use the simple substitution only. Even if we use simple technique it strongly secures data without any complex computation which are time consuming [9] and also it help us to store data in a desired length. Then we hide cipher DNA using inserting reference DNA and for double layer protection we use RSA algorithm.

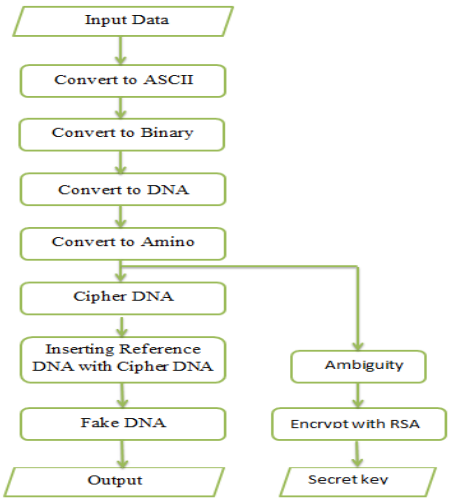

Figure 1: Overall Encryption process

In this algorithm, the encryption process starts by taking the data that has to be kept secure. And this data is converted into binary form, which is transferred to DNA by using table 1[17]. Then the DNA form is transferred to the Amino acids form based on another table which is a standard universal table of Amino acids and their codons representation in the form of DNA. [3] DNA form reduces the large input data into a desirable

length. Hence we form a DNA based cipher and there will be another text which act as a secret key (ambiguity) which is formed along with amino acid [13].

DNA	Bit 0	Bit 1
A	0	0
G	0	1
C	1	0
T	1	1

Table 1: DNA representation of bits

Figure 1 shows how we apply encryption in the first phase of the proposed technique. In this phase, the data "plaintext" is encrypted using DNA and Amino Acids concepts as follows: Convert plaintext to ASCII then this ASCII is converted to binary form such as 8-bits coding. After that, a binary coding scheme used to transform binary form into DNA alphabets A, C, G and T. Here we are considering the following as the binary coding used A -00, C- 01, G- 10, T- 11. For converting the data to DNA based alphabets we have to group the binary data into two and then replace it with the corresponding alphabets as shown in Table 1.

Next, the DNA form is transferred to the Amino acids form according to a standard table of Amino acids representation in the form of DNA and the new distribution of the alphabet with the corresponding codons. Now, English alphabets form of Amino Acids can go through another substitution, for that we have to group the three alphabets and form an amino acids. And replace each of this amino acid.

POSITION	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
0	GCT	TAA	TGT	GAT	GAA	TTT	GGT	CAU	ATT		AAA	TTA	ATG	AAT	TTA	CCT	CAA	CGT	TCT	ACT	AGA	GTT	TGG	AGT	TAT	TAC
1	GCC	TAG	TGC	GAC	GAG	TTC	GGC	CAC	ATC		AAG	TTG		AAC	TTG	CCC	CAG	CGC	TCC	ACC	AGG	GTC		AGC	TAC	
2	GCA	TGA					GGA		ATA			CTT				CCA		CGA	TCA	ACA		GTA				
3	GCG						GGG					CTC				CCG		CGG	TCG	ACG		GTG				
4												CTA						AGA	AGT							
5												CTG						AGG	AGC							

Table 2. Distribution of Alphabets corresponding to codons

Based on the Table 2 corresponding to each amino acid group there will be a row and column. The column represents the substitution for the group and the row represents the position of each group that is taken as secret key and is known as ambiguity. This reduced form of alphabet form the amino acid representation. And we also have an ambiguity produced. This is based on the genetic code by which DNA stores the genetic information consists of "codons" of three nucleotides. With four possible bases, the three nucleotides can give $4^3 = 64$ different possibilities, and these combinations are used to specify the different amino acids.

And these 64 combinations of amino acid is distributed in different positions in table and the table column corresponds to alphabets and the 64 amino acids are distributed in all these 26 locations and each location has more than one amino acid and the row corresponds to the position of these amino acids(0,1,2,3,4...). And the number of amino acid under an alphabet is not fixed; it can be 0 or more. And space is represented at alphabet j, and j contains space only. So ambiguity represents the position of each amino acids and using these only we can decrypt the original data so it is taken as secret key. At this phase we get encrypted DNA sequence and we take the DNA sequence as the cipher text and the ambiguity as the secret key. [18] And to provide more security we use RSA algorithm to encrypt the secret key i.e. ambiguity .Finally we get our data in two different files, one the cipher text and the other the encrypted secret key.

Here we consider simple substitution for encryption and by these methods we can encrypt the data in simple steps and is secure. Then when the original data is needed we can decrypt the data as we are using simple substitution. When we decrypt the data we get the original data without any junk data. We can also represent the space and other symbols. But if conventional Playfair cipher is used it cannot represent the space and symbols and also when we decrypt it these spaces will be occupied with the junk data and we won't be able to get the actual input data when decrypted.

Phase II - Hiding cipher DNA with Reference DNA Sequence

In this phase we will use the encrypted DNA sequence and reference DNA sequence as input of insertion technique.[14] The insertion technique is introduced to deal with DNA alphabets, so by insertion method we can hide the original cipher text to a faked DNA sequence.[8][16] First, divide both encrypted DNA sequence and reference DNA sequence into segments where each segment contains a random number of DNA nucleotides so the segments are not fixed in length.[11] Next, insert each segment of encrypted DNA sequence before the segments of reference DNA sequence respectively. Finally, we get a faked DNA sequence with the encrypted

DNA sequence hidden [15] as shown in Figure 2. Using insertion of reference DNA i.e. DNA sequence of any living things e.g. - human beings.

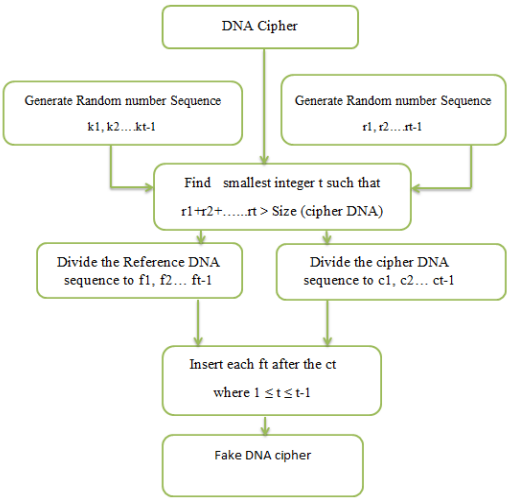

Figure 2: Data Hiding Using Insertion

Phase III - Encrypting Secret Key using RSA algorithm

In the third phase we encrypt the ambiguity which contains the position of amino acids with RSA algorithm. For that we need a pair of keys i.e. one public key and private key. And using the public key we encrypt the ambiguity and secret key is formed as shown in figure 3. And in order to decrypt data we have to first decrypt the ambiguity which has been encrypted using RSA and then using these ambiguity we can decrypt the actual data.

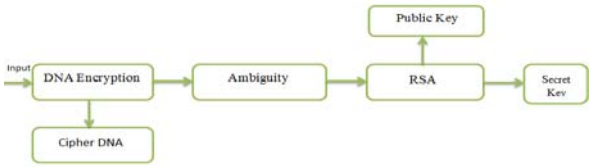

Figure 3: Encrypting the key

IV. EXPIREMENTAL ANALYSIS

We all know the e-mails and chatting are mostly prone to several types of attacks and also it is widely hacked by hackers. So we use the application mainly for sending and receiving data securely through e-mails and chat. In the experiment we are considering a client–server enviornment and server will contain the application with the above mentioned techniques. And for using this application in server, user has to register. So the registered users will be the clients and these users have the provision for sending mail and when a client sends mail using this application the data send will be in encrypted form i.e. The encrypted data contains cipher text and key (amino and ambiguity).Both are send as attachments.Then using the public key of intended recipient we encrypt the ambiguity. And whoever is the intended recipient use their private key for decryption and get the actual ambiguity .Use the produced ambiguity and the fake DNA in another file for decryption to get the actual data And in chatting the intended recipient will get message and the have to decrypt it and read.

V. CONCLUSION

As there are many algorithms for providing security efficiently. We use DNA based cryptography which is based on the fundamental idea of using the structure of DNA and amino acids .Using the concept of DNA we encrypt the data to enhance their security features. In this encryption instead of using conventional cryptographic algorithms we just use DNA and amino acid structure and the algorithm succeeded in overcoming problems in using the conventional algorithms which had already prone to attacks and had been broken. And also "Playfair cipher based DNA based encryption" like junk values in data after decryption. As in our algorithm the plaintext is to be converted to its binary value before encryption, it is now clear that the plaintext message can be written in upper or lower case, with any punctuation, and numerical values. As DNA concept is used, it helps in high information storage. And we are using simple substitution for encrypting that can reduce the time complexity and is secure. And in the second phase we insert data hiding by insertion technique which also provides a high level of security. And the DNA based encryption is best for using with the conventional algorithms to provide more security. We consider powerful asymmetric RSA algorithm in third phase for encrypting the key so it provides a double layer security.

VI. REFERENCES

[1] R. R. Sinden, DNA Structure and Function. New York: Academic, 1994.
[2] Yunpeng Zhang and Liu He Bochen Fu. Research on DNA Cryptography, Applied Cryptography 2012
[3] Sherif T. Amin, Magdy Saeb, Salah El-Gindi, "A DNA- based Implementation of YAEA Encryption Algorithm," IASTED International Conference on Computational Intelligence (CI 2006), San Francisco, Nov. 20, 2006.
[4] A. Atito , A. Khalifa , S. Z. Rida, DNA-Based Data Encryption and Hiding Using Playfair and Insertion Techniques(2011)
[5] William Stallings, Cryptography and Network Security Principles and Practices, Fourth Edition, 2005
[6] Hook, D., Beginning Cryptography with Java, Wrox Press, (2005)
[7] K. RAMA et al., SURVEY AND ANALYSIS OF 3D STEGANOGRAPHY, International Journal of Engineering Science and Technology (IJEST), Vol. 3, 2011.
[8] H. J. Shiu et al, Data hiding methods based upon DNA sequences, Information Sciences, 2010.
[9] Mona Sabry, Mohamed Hashem, Taymoor Nazmy, Mohamed Essam Khalifa A DNA and Amino Acids-Based Implementation of Playfair Cipher International Journal of Computer Science and Information Security, Vol. 8, No. 3, 2010
[10] Cui G et al. DNA computing and its application to information security field [C]. IEEE Fifth International Conference on Natural Computation, Tianjian, China, Aug. 2009. Xiong Fuqin, Cryptography Technology and Application [J]. Science, 2010
[11] H. J. Shiu et al., Data hiding methods based upon DNA sequences, Information Sciences, Vol. 180, 2010.
[12] Watson JD, et al., Molecular Biology of the Gene, 2004.
[13] Ashish Gehani, Thomas LaBean and John Reif. DNA-Based Cryptography. DIMACS DNA Based Computers V, American Mathematical Society, 2000.
[14] Voloshynovskiy S, Pun T, Fridrich JJ, Pérez-González F, Memon N. Security of data hiding technologies. Signal Process 2003; 83(10):2065–7.
[15] Hayam Mousa et al. , Data Hiding Based on Contrast Mapping Using DNA Medium, The International Arab Journal of Information Technology, Vol. 8 No. 2 , 2011.
[16] 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
[17] Heider and Barnekow; licensee BioMed Central Ltd. R. R. Sinden, DNA Structure and Function. New York: Academic, 1994.
[18] Leonard Adleman. "Molecular Computation of Solutions to Combinatorial Problems". Science, 266:1021-1024, November 1994.