IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-9, Sept 2019 E-ISSN: 2347-2693

# Modified RSA Cryptosystem with Data Hiding Technique in the Terms of DNA Sequences

**Harsh Sahay**

Master of Technology (Information Technology) Institute of Engineering and Management, Kolkata, India

**Corresponding Author: Sahayharsh53@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i9.9194 | Available online at: www.ijcseonline.org

Accepted: 09/Sept/2019, Published: 30/Sept/2019

***Abstract***- RSA algorithm is an efficient algorithm for preventing unauthorized access over the network. But there are some drawbacks of RSA algorithm such as its high computational time. In this work we are reducing the computational time of RSA algorithm and increasing security of RSA algorithm. In this work we are modifying security of RSA algorithm by using three prime numbers instead of two as used in RSA algorithm. For reducing computational time of RSA algorithm to each character, multiple characters are merged together to form a merged unit. For merging each character cantor's pairing algorithm has been used. The merged unit is now encrypted to the network. To the receiver side cipher text is received. After decryption, the merged data unit is received to the receiver side. After going to cantor's unpairing algorithm individual characters of merged data unit are displayed to the receiver side .The highlight of this work is it increases efficacy of RSA cryptosystem. This modified work reduces computational time of RSA algorithm, even increases security of this algorithm. Even In this work we are hiding cipher text in the terms of DNA sequences. So that it is very difficult for intruders to get a real DNAsequence.

***Keywords***- DNA steganography, Cryptography, RSA, public key, private key, pairing and unpairing algorithm.

## I. INTRODUCTION

The word cryptography has Greek origin it is combination of two,"Kryptos" which means hidden and "logos" which means word and graph, means secret and writing. Cryptography is a science of converting a stream of text into coded form in such a way that only the sender and receiver of the coded text can decode the text. Cryptography plays a very important role in internet based commercial activities as many secret documents which include payment details, money transfer, contract documents, and business plans and other confidential information are to be transferred from one computer to another computer. Cryptography is a technique that allows a piece of information to be converted into cyptic form before being stored in a computer database or transmit over the secure channel. Encryption of message is done to provide extra protection in order to maintain confidentiality of documents. For example, if an unauthorized person succeeds in tapping the channel then information he has copied may not be of his use, if it is encrypted. Cryptography is primarily used to protect the confidentiality of information from intruders. [1]

There are two kinds of cryptography Asynchronous Key Cryptography Synchronous Key Cryptography In synchronous key cryptography one key is shared between sender and receiver. While in Asynchronous key cryptography two key is shared between sender and receiver. One is called public key which is publically available while another is called private key, which is kept secret. Steganography is the process of hiding data into a medium such that medium appears to be unsuspicious. Combination between cryptography and Steganography is done by first encrypting data using encryption techniques and hiding it into transportation medium using an Steganography techniques (Atito et. al. 2012).The most common transportation media for Steganography are Images, audio, video and DNA.DNA Steganography is proved to be most promising one because its huge storage capacity, complexity and randomness. These features provides great uncertainty which makes encoding data into a DNA format to hide it within a DNA medium is far better than any other Steganography mechanism.[1] In this paper I have modified RSA Cryptosystem. And I have combined this modified RSA cryptosystem technique with DNA Steganography. In this paper data is firstly enciphered using modified RSA cryptosystem than the resultant enciphered data is encoded into a DNA format and hide it into a real DNA sequence using modified substitution Steganography technique. [1]

## II. RELATED WORK

1. Vivek Choudhary and Mr. N. Praveen have proposed modification of RSA algorithm by the use of third prime number in their work, which increases security of RSA algorithm.[4]

© 2019, IJCSE All Rights Reserved 

2.Samiha Marwan, Ahmed Sawish, Khaled Nagaty developed DNA based cryptographic methods for data hiding in DNA Media. [1]

3.Rivest, Adi Shamir and Adelman has invented RSA algorithm which it is widely most used public key cryptosystem, this algorithm used to encrypt the data to provide security [3].

**DNA Stenography**

DNA stenography is the science of hiding data into a DNA sequence as a hiding medium. Data must be encoded into a DNA format first in order to be merged within a DNA sequence, where resultant sequence looks like a real DNA sequence. There are different encoding techniques that can be used for encoding data into DNA format. For example data can be encoded into DNA format by matching each pair with a single bit, for example, "A-T" pair is encoded into 0, and "G-C" pair is encoded into 1.The most famous and simple way for encoding data into DNA sequence is by converting each two binary bits to a DNA nucleotide, if we have a binary message"01110010" it will be encoded to "GTAC" as shown in table below. [1]

DNA letter represented by binary bits

| DNA letter | Binary representation |
|---|---|
| A | 00 |
| G | 01 |
| C | 10 |
| T | 11 |

This method uses a random suitable real DNA sequence-reference sequence to hide data though it. There are almost 1.6 * 10^8 real DNA sequences available on online database(NCBI Database),which makes it computatitionaly so hard to detect real DNA sequence.Morever data is firstly encrypted by well- suited encryption technique before aforementioned hiding process. This process makes it virtually impossible to detect the original hidden message.[1]

## II. METHODOLOGY

**RSA cryptosystem**

Although there are several asymmetric key cryptosystem, one of the common public key algorithm is the RSA cryptosystem, named for his inventors (Rivest, Shamir and Adleman).RSA uses two exponents, e and d, where e is public and d is private. Suppose P is the plaintext and C sis the cipher text Alice uses C=P^e mod n to create cipher text C from plaintext P; Bob uses P=C^d mod n to receive the plaintext sent by Alice.Modulas n is a very large number is created during key generation process.

**RSA Algorithm**

1. Choose two large prime numbers P and Q. Let it be P=7 and Q=17.

2. Calculate N=P*Q.
We have N=7*17= 119

3. Select the public key (i.e. encryption key) E such that it is not the factor of (P-1) and (Q-1).

Let us find (17-1)*(7-1) =96
Factor of 96 are 2, 2,2,2,2 and 3(96=2*2*2*2*2*3).Thus, we have to choose E none of the factor of E is 2 and 3.As a few example we can't choose E as 4(because it has 2 as a factor), 15(because it has 3 as a factor),6(because it has 2 and 3 both as Factor).Let us choose E as 5 (it could have been any other number that does not its factors as 2 and 3).

4. Select a private key (i.e. decryption key) D such that the following equation is true (D*E) mod (p- 1)*(q-1) =1

Let us substitute the value of E, P and Q in the equation.
We have: (D*5) mod (7-1)*(17-1) =1. That is, (D*5) mod (6 * 16) =1.
That is, (D*5) mod (96) = 1.
After some calculation let us take D=77.Then
the following is true:
(77*5) mod (96) = 385 mod 96 = 1.Which is what we wanted.

5. For encryption, calculate the cipher text CT from plain text PT as follows:
CT= (PT^E) mod N

Let us assume that we want to encrypt plain text 10.Then we have, CT= (10^5) mod 119=100000 mod 119 = 40. Send CT as a cipher text to the receiver. Send 40 as a cipher text to the receiver.

PT= (CT^D) mod N.

We perform the following
PT= (CT^D) mod N. That is, PT= (40^77) mod 119=10 which is original plain text. [2]

6. For decryption, calculate plain text PT from cipher text CT as follows,

PT= (CT^D) mod N.

We perform the following
PT= (CT^D) mod N. That is, PT= (40^77) mod 119=10 which is original plain text. [2]

© 2019, IJCSE All Rights Reserved 

**Problem concerning RSA Cryptosystem:**
Security of RSA Cryptosystem based on the assumption that it is easy to multiply two large prime number together but it is extremely difficult to factor their product. In RSA if any one factored its product of two prime numbers, then private key can be detected and the security of RSA can be broken. So we need to increase the security of this algorithm. Even the limitation of RSA cryptosystem is its time of computation i.e. it takes more time to compute the mathematical operation of RSA algorithm.

**Solution Methodology**
The security of RSA algorithm can be compromised in the network. To increase the security of RSA algorithm we need to modify RSA algorithm. We are increasing security of RSA algorithm by using three large prime numbers instead of two as used in RSA algorithm.

Another limitation of RSA algorithm is its time of computation. We are reducing the time of computation of RSA algorithm by using cantor's pairing and unpairing algorithm.

**Cantor's Pairing Algorithm**
A pairing algorithm on set A associates each pair of members from A and generates a single integer number. Here is a classic example of a pairing algorithm. When x and y are nonnegative integers, Pair (x_, y_) outputs a single non-negative integer that is uniquely associated with that pair.

Pair@[x_,y_]:= Z=(x^2+3x+2xy+y+y^2)/2;

Here Z is a single integer number. This is generated by above equation. Which is a paired value of x and y.
The inverse function- Unpair@[Z_]:= i=
-1+sqrt (1+8*Z))/2);
x= Z-i(1+i)/2 ,
y=i(3+i)/2-Z

Above equations are way to evaluate value of x and y from Z. This is called unpaired value generated of Z

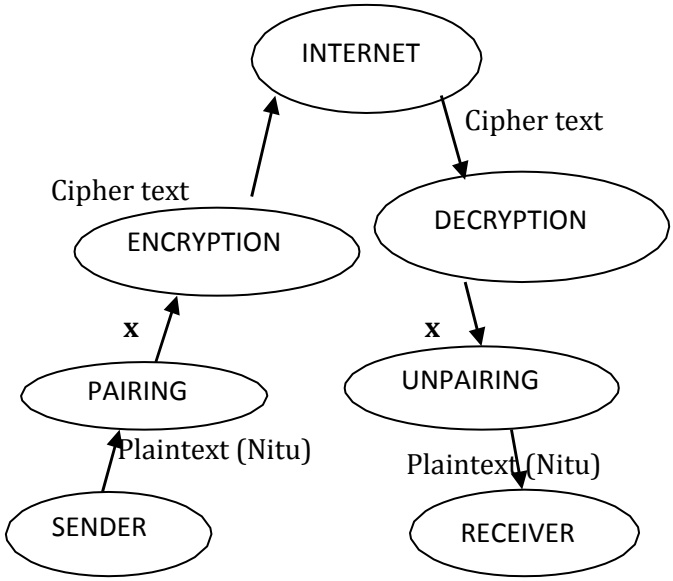
**Modified RSA algorithm**
1. Generate a single integer number of sent messages by cantor's pairing algorithm.
2. Choose three large prime numbers.
3. Calculate N=P1 *P2 *P3.
Here P1, P2 and P3 are prime numbers. 4. Calculate Q= (P1-1)*(P2-1)*(P3-1)
4. Select the public key (i.e. encryption key) E such that it is not the factor of Q.
5. Select the private key (i.e. the decryption key) D such that the following equation is true:

(D * E) mod Q =1.

6. For, encryption calculate cipher text CT from the plain text PT as follows
CT= (PT^E) mod N
Send CT (Cipher Text) as a secret code to the receiver from sender.
7. For decryption, calculate the plain text from the cipher text CT as follows
PT= (CT^D) mod N.
8. Displaying each character by using cantor's unpairing algorithm (Reverse process of pairing)
Let Z is a plaintext

i= -1+sqrt (1 + 8 * Z) / 2;

x= Z-i (1+i)/2
y=i(3+i)/2-Z [5]



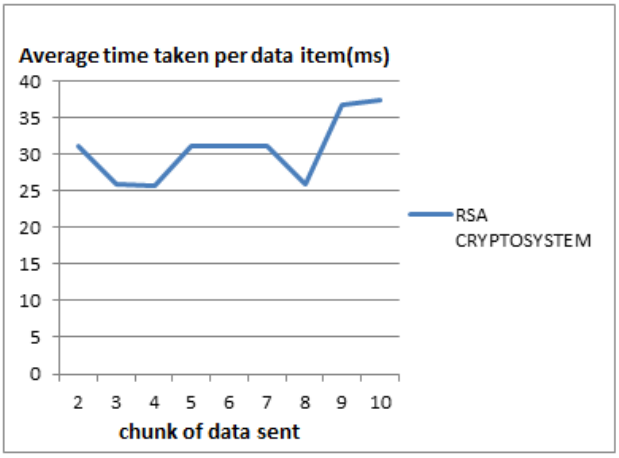

## III. RESULT AND DISCUSSION

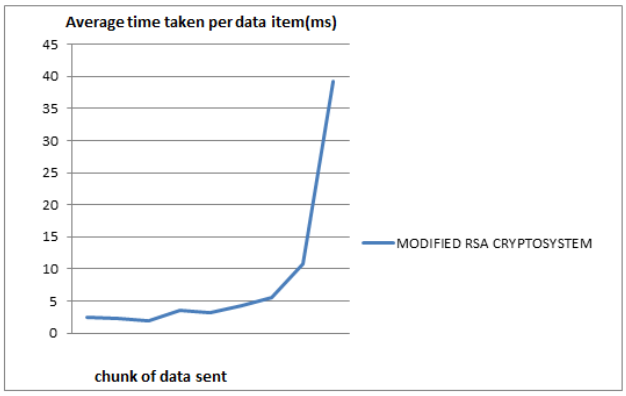
Let as shown in above diagram a Plain text message Nitu is send by the sender to the receiver by pairing function this message is converted into an integer x and it is encrypted into a cipher text and goes to networks. After Decryption encrypted message is converted into the integer number x which was actually sent by the sender and after passing to the unpairing function the Plaintext message Nitu is displayed to the receiver side.
According to graph of figure 1 and figure 2 , the following detail have found experimentally.

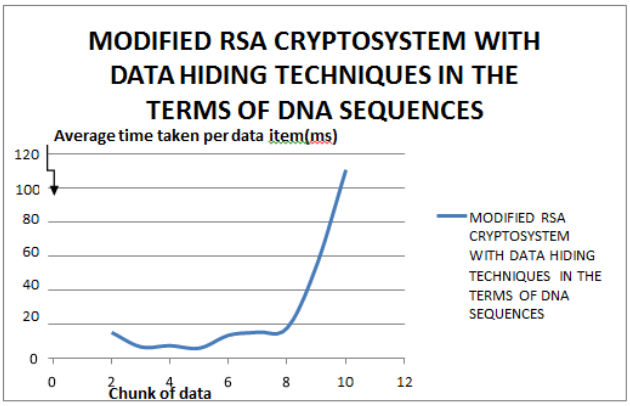
A a lot of messages were sent by RSA algorithm and modified RSA algorithm. If number of characters is less than or equal to nine Modified RSA algorithm performs very well in compared to RSA algorithm. Modified RSA algorithm reduces computational time of RSA algorithm.

© 2019, IJCSE All Rights Reserved 

Even security is excellent in terms of modified RSA cryptosystem because we are using three prime numbers instead two as we used in RSA algorithm.




## IV. CONCLUSION

Now we are hiding cipher text of modified RSA cryptosystem in the terms of DNA sequence we get graph experimentally shown in figure 3 below. It has been shown modified RSA performs quite well in terms of DNA sequences data hiding techniques if chunk (up to nine) of data sent by it. it is very difficult for intruders to detect a real DNA sequences.

## REFERENCES

[1]. DNA based cryptographic technique for data hiding in DNA media Samiha Marwan, Ahmed Shawish, Khaled Nagaty.
[2]. Atul kahate, Cryptography and network security (TMH) RSA algorithm.
[3]. R. L. Rivest, A. Shamir, L. Adelman, "On Digital Signatures and Public Key Cryptosystems," MI Laboratory for Computer Science Technical Memorandum 82, April 1977.
[4]. Vivek Choudhary1 and Mr. N. Praveen2 "Enhanced RSA Cryptosystem Based on Three Prime Numbers" 1 Post Graduate Scholar, Department of Computer Science & Engineering, SRM University, Chennai, Tamilnadu, India 2 Assistant Professor, Department of Computer Science & Engineering, SRM University, Chennai, Tamilnadu, India
[5]. An Elegant Pairing Function", Matthew Szudzik, Wolfram Research, Inc. NKS 2006 Wolfram Science Conference

© 2019, IJCSE All Rights Reserved 