Cryptanalysis Using Genetic Algorithm

Manish Kumar Tiwari, Darren Pinheiro, Shivam Shukla, Sanjay Poptani, Dr. K Natarajan
Computer Science, Christ University, Bangalore, Karnataka, India-560074

Abstract— *A world full of security breaches, where we all thrive to secure our data and fight with data leak and data theft problems, we have come up with an idea of encryption and decryption with help of genetic algorithm and non-deterministic random numbers. In this paper we have generated mathematical model for encrypt and for decrypt. For encryption we are using three level of encryption. Ciphertext1 will be generated using the text given and it will be treated with genetic algorithm to produce Ciphertext2, now to make cipher more strong a random number as Ciphertext3 will be generated. All three of cipher text will be sum up to create complex cipher text and the addition of three cipher text will be the final cipher text. While treating ciphertext1 with genetic algorithm fitness function will be used to choose the best key to generate Ciphertext2, different test will be done on output of fitness function to make sure of the effective randomness of ciphertext2.To sum up, this paper demonstrates a new algorithm for encryption and decryption with its strength and effectiveness totally based upon genetic algorithm and random number. The results on the effect of the genetic algorithm are inconclusive. However, even if the genetic algorithm does off an improvement it has the major drawback of running very slowly. There are many possible ways of improving this system. And it seems to be an area where much research is needed.*

Keywords— *Cryptography: Decryption: Encryption: Genetic Algorithm: Non-deterministic random numbers.*

I. Introduction & Literature Study

In the present-day scenario, there is sudden increase in number of cases of data breaches, at any time any data/information is transfer/send it need to be safe and secure [1]. This article introduces the cryptanalysis approach of the RSA cryptosystem based on the application of genetic algorithms. This approach suggests a faster process aimed at reducing the number of simple patterns required for a successful time attack. Further research into advancing the idea of genetic algorithm technology in a working RSA system has yielded promising results [2]. This article suggests that it is possible to genetically reproduce a hard-to-write computer program, a randomizer that runs a continuous data sequence with the maximum entropy of pseudorandom bits [3]. Decryption using genetic algorithms has attracted great interest in recent years. This article introduces a two-round cryptographic analysis approach based on genetic algorithms. However, genetic algorithm cryptanalysis is usually a difficult task. In this article, we employ a simple unknown attack and develop various keys based on fitness functions. Experimental results suggest that this is a promising method and can be employed to handle other complex blockers [4]. In this experiment, we analyzed the effectiveness of genetic algorithms applied to detect computer interference and malicious computer behavior. Genetic algorithms are a way to solve the problem of artificial intelligence based on Darwin's theory of evolution applied to mathematical models [5].

Encryption -The process of encrypting data to secure it, it is process of converting data using different technique to transform data into unreadable/without any meaning form, encrypted data is known as cipher text.
Decryption - The process of decrypting the cipher text to get back the original data.
Mutation – It is a genetic operator used to maintain diversity from one generation of a population of genetic algorithm chromosomes to the next.
Crossover – It is a genetic operator used to combine the genetic information of two parents to generate a new offspring.
Selection – At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them to produce the children for the next generation.
Genetic Algorithm – It is a search heuristic based algorithm that reflects the process of natural selection where the fittest individuals are selected for reproduction of next generation.
Cipher text - Encrypted data, which is unreadable and makes no sense until converted to plain text.
Plain text -The original data that need to be encrypted before transfer.
Key – it is alphanumeric/numeric/text/special symbol which is used to encrypt/decrypt data.

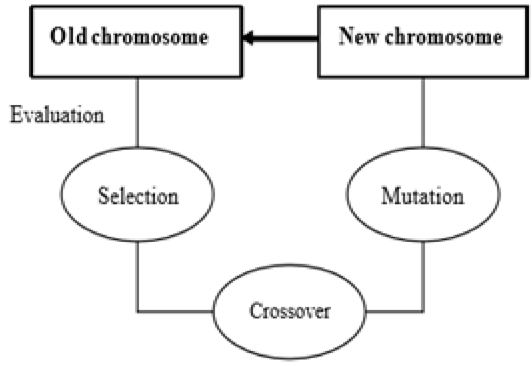

Fig. 1. The basic genetic algorithm cycle

In the work [6], the authors discussed the concept of basic genetic algorithm in substitution cipher and in the figure1 it has been shown to produce new chromosome we require old chromosomes. A method called genetic algorithm (GA) was first introduced by J. H. Holland [polak 2015]. Its functioning is based on evolution of living organisms. It uses phenomena such as natural selection and evolutionary operators like crossover and mutation. Genetic algorithm works on specific population which consists of finite number of individuals. Each individual is represented in a specific way suited best for considered task. Proper selection of GA’s functions and individual's representation is essential to obtain satisfying results. Genetic algorithm belongs to the set of probabilistic methods, which means that for every run of the algorithm different results could be obtained. It also gives approximate solution, very good one, but not necessarily the best one [Polak 2015].

Algorithm 1 Genetic Algorithm

```
1. Randomly generate initial population
2. Evaluate the fitness for every individual
3. While termination condition has not been reached to
   a. Apply for chosen individuals
   b. Crossover
   c. Mutation
   d. Replace old population with new one using selection
      and reproduction
   e. Evaluate the fitness for every individual
4. Return the best fitness solution found
```

Crossover: - The process in which two chromosomes or two attributes are taken and a resultant chromosome(new) is formed by taking some part of first chromosome and the rest by second chromosome. There are three types of crossover operation in genetic algorithm.

A. Single point crossover: the selected chromosome is broken into half and the new chromosome is formed by interchanging or swapping the half of the selected two chromosomes.

B. Two-point crossover: the selected chromosome is divided into three parts by taking two points and then one parts of each chromosomes are swapped to form new chromosome.

C. Multi point crossover: the selected chromosome is divided into 'n' number of parts and then the swapping is done in order to form new or resultant chromosomes.

Mutation: - It is a genetic operation which is similar to biological mutation and is used to create genetic diversity of its one generation from its successive generation. Mutation allows the algorithm to prevent the population of chromosomes from becoming similar to each other. In the genetic programming paradigm, the individuals in the population are compositions of functions and terminals appropriate to the particular problem domain. The set of functions used typically includes arithmetic operations, mathematical functions, conditional logical operations, and domain-specific functions. The set of terminals used typically includes inputs (sensors) appropriate to the problem domain and various constants. Each function in the function set must be well defined for any combination of elements from the range of every function that it may encounter and every terminal that it may encounter [Stanford]. If we try to change the rate of mutation, then the total effectiveness and strength of the algorithm will be increased and will may it difficult to break.

II. Methodology Used

We have come up with an algorithm which uses non deterministic random number to encrypt data. We have formulated a mathematical model to encrypt the data using 3 different cipher texts namely c1, c2, c3. In our mathematical model, the decryption is done using the three cipher text generated and the private key of the receiver. Algorithm of our work include: -

1------------- Alphabet to ascii
2------------- ascii to binary
3------------- Sum of each alphabet is C1
4------------- Genetic Algorithm c2
5------------- Selecting random number c3
6------------Encryption c1+c2+c3
7---------- Decryption step wise
7.1 c1[E-(c2+c3)], c2, c3
7.2 decrypt c2 using genetic algorithm

Different Modules of Work

i. Generating a mathematical model
ii. Selecting Plain Text
iii. Generating C1
iv. Generating C2
v. Generating a Non-deterministic random number (C3)
vi. Encryption
vii. Decryption
viii. Complexity Analysis
ix. Gaps test and Chi-Square test
x. Example problem
xi. Conclusion and future work

II. Mathematical Formulation

1 GENERATING A MATHEMATICAL MODEL

Encryption Equation

Here encryption is a single step process

$Encryption(C1, C2, C3) Encryption(C1, C2, C3)$

$E = C1 + C2 + C3$

$E = C1 + C2 + C3$

Decryption Equation

Here decryption is a two step process in first step,
C1, C2, C3 are obtained from the final encrypted cipher text and in the second step,
Genetic algorithm is used to get back the initial value of C2 into the plain text.

Step – 1

$C1 \ = E - (C2 + \ C3)$

$C2 \ = E - (C1 + \ C3)$

$C3 \ = E - (C1 + \ C2)$

Step -2
Getting complete C2 using genetic algorithm

1 Selecting Plain Text

As for now the plain text may consist of

i. Alphabets
ii. Numerals
iii. Special Characters
iv. Boolean
v. Or with any combination of the above

2 GENERATING C1

Here the plain text is divided into each and individual characters and based upon their character ascii value they will be converted into the Boolean numbers and all the different characters of the plain text are added to produce the final result C1.

$$Plain\ text = \boldsymbol{HELLO\ WORLD}$$

N = no of letters in plain text
N = 10

Here ascii value of different characters are considering all the letters are in capitol form

$H = 72$
$E = 69$
$L = 76$
$O = 79$
$W = 87$
$R = 82$
$D = 68$

$C1 = H + E + L + O + W + R + D$
$C1 = 72 + 69 + 76 + 79 + 87 + 82 + 68$
$C1 = 533$

Now converting the value of C1 into binary we get

$C1 = 100010101$

3 GENERATING C2

Here basically the genetic algorithm is run here since our plain text is HELLO WORLD
After running our genetic algorithm we got
Genes are “1” and “0”
So similarly the Boolean value is given below

$H = 1001000$
$E = 1000101$
$L = 1001100$
$L = 1001100$
$O = 1001111$
$W = 1010111$
$O = 1001111$
$R = 1010010$
$L = 1001100$
$D = 1000100$

And after performing the sum(+) operation on these number we get ,
Sum = 1011111100
Performing genetic algorithm,
Now we get **(e]L3S.?BPLy)**
for hello world when we applied genetic algorithm. Now converting **(e]L3S.?BPLy)** to ascii value.

101 93 76 51 83 46 63 66 80 76 121 = (e]L3S.?BPLy)
Now converting ascii value to its Boolean.

101 = 1100101
93= 1011101
76 = 1001100
51= 0110011
83 = 1010011
46= 0101110
63= 0111111
66= 1000010
80= 1010000
76= 1001100
121= 1111001

Final c2 will considered as
11001011011101100110001100111010011010111001111111
000010101000010011001111001

C2 is combination of binary of each character of output of genetic algorithm.

4 GENERATING A NON DETERMINISTIC RANDOM NUMBER (C3)

Here a random number is generated by any known method and then the value is stored in C3.

Selecting a random nuber in next step we convert random number also in Boolean

C3 = 99999

Converting into Binary we get,

C3=00000000000000000000000000000000000000000000
00000 0000000000011000011010011111

5 ENCRYPTION

For the Encryption process a simple arithmetic operation is used between C1, C2, C3 to increase the total effectiveness and total randomness

$Encryption(C1, C2, C3)Encryption(C1, C2, C3)$
$E = C1 + C2 + C3E = C1 + C2 + C3$

C1=00000000000000000000000000000000000000000000
000000000000000000000000100010101

C2=1100101101110110011000110011101001101011100111
1111000010101000010011001111001

C3=00000000000000000000000000000000000000000000
00000 0000000000011000011010011111

E = c1+c2+c3

Considering everything for 72 bits

E=110010110111011001100011001110100110101110011
11111000010101011010111000101101

6 DECRYPTION

Decryption here is a two step process

In the step 1,

We will get the c1, c2, c3 at each and every individual step

$C1 = E - (C2 + C3)$
$C2 = E - (C1 + C3)$
$C3 = E - (C1 + C2)$

C1 = e – (c2 + 11000011010011111)
C2 = e – (100010101+11000011010011111)
C3 = e – (100010101+C2)

C1 = 100010101
C2=1100101101110110011000110011101001101011100111
111100001010100001001100111100 1
C3=0000000000000000000000000000000000000000000000
000 0000000000011000011010011111

In the step 2,

We will decrypt C2 Completely
Now we will divide c2 in N equal parts.
N = 10 , numbers of letters in plain text
Now we will get ,
1100101
1011101
1001100
0110011
1010011
0101110
0111111
1000010
1010000
1001100
1111001

Now converting the Boolean to ascii values, we get :

101 = 1100101
93= 1011101
76 = 1001100
51= 0110011
83 = 1010011
46= 0101110
63= 0111111
66= 1000010
80= 1010000
76= 1001100
121= 1111001

Now converting ascii value to characters ,
We get
101 93 76 51 83 46 63 66 80 76 121 = (e]L3S.?BPLy)

Now we have retrieve c2 .
We will compare our retrieved c2 with the data stored in file which was uploaded in cloud while encrypting and get the original plain text back.

III. Complexity Analysis

This work is free from side channel attacks as a random number is used in encryption process. Since the genetic algorithm is used therefore it makes it safe from various dictionary as well as brute force attack.

IV. Conclusion and Future Work

The work considers exponential operations with random numbers for encryption process; the random number used in the encryption process makes the process to be free from side channel attack. The strength of the algorithm lies with strength of Discrete Logarithm problem which is a hard problem. Since the genetic algorithm is used

The work can be extended to be applications in somewhat and Fully Homomorphic encryption process and can also be extended to digital signature standard. The work can be further developed as authentication technique for various users. Since the algorithm takes less space this work can also be implemented into various internet of things (iot) devices.

Acknowledgement

I would like to thank CHRIST (Deemed to be University) Vice Chancellor, Dr. Rev. Fr. Abraham V M, Pro Vice Chancellor, Dr. Rev. Fr. Joseph C C, Director of School of Engineering and Technology, Dr. Rev. Fr. Benny Thomas and the Dean Dr, Iven Jose for their kind patronage.

I would like to express my sincere gratitude and appreciation to the Head of Department of Department of Computer Science and Engineering, School of Engineering and Technology, Dr. K Balachandran, for giving me this opportunity to take up this comprehension.

I would like to offer our sincere thanks to coordinator of comprehension for their immeasurable help and guidance without whom the completion of this comprehension would have been challenging.

I would like to thank my parents for their support and blessing without which we cannot succeed in life. I also take this opportunity to thank them for being the guiding light in our life.

References

[1] Public Key Exchange using Matrices over group rings, Delaram Kahrobaei ,Charalambos Koupparis & Vlendimin Shpilrain , Lecture Notes.
[2] Krishna A.V.N, Pandit,S.N.N (2004). A new Algorithm in Network Security for data transmission ,Acharya Nagarjuna International Journal of Mathematics and Information technology . 1(2), 2004,97-108.
[3] Krishna ,A.V.N(2005).A simple algorithm for random number generation, Journal for Scientific & Indsutrial research,(64).
[4] Krishna A.V.N(2010), Role of statistical tests in terms of security of a new encryption algorithm, International Journal of Advancements in Technology, Vol1, No.1, 2010.
[5] Marton, K(2010), Randomness in Digital Cryptography.
[6] J. L. Massey, "Cryptography—A selective survey," Digital Communications, vol. 85, pp. 3-25, 1986
[7] O. O. Khalifa, M. R. Islam, S. Khan and M. S. Shebani, "Communications cryptography," in RF and Microwave Conference, 2004. RFM 2004. Proceedings, Selangor, 2004.
[8] A Public key cryptosystem based on Algebraic Coding theory , DSN progress report 42-44,114.