Simple and Effective Data Encryption and Decryption Technique for Cloud Environment without using Reference Sequences

Shreenath Waramballi¹, Dr. Guruprakash C D², Rama Moorthy H³, Manjunath Raikar⁴, Sunil B N⁵

¹Assistant Professor, Computer Science Engineering Department, Sahyadri College of Engineering and Management, Mangalore
²Professor, Computer Science Engineering Department, Shree Siddarthha Institute of Technology, Tumkur
³Assistant Professor, Computer Science Engineering Department, Shree Madhwa Vadiraja Institute of Technology, Udupi
⁴Assistant Professor, Computer Science Engineering Department, Klaravali Institute of Technology, Mangalore
⁵Assistant Professor, Computer Science Engineering Department, Sahyadri college of Engineering, Mangalore

Abstract: To secure the data over the unsecured networks, various types of data protection methods are necessary. With advent of Cloud Computing idea, a common problem, confidentiality of data, was emerged. The main idea of this paper is to propose an algorithm to provide the data security in cloud computing by increasing the complexity in encrypted data without using the reference sequences. The given text data is converted into the binary form and each two bits of binary data is mapped into the first four alphabets. These alphabets are replaced by very next characters in the sequence. Now the sequence of characters formed would serve as encrypted characters. These are the characters to be sent to cloud and at the receiver’s end these encrypted characters are decrypted back by replacing each alphabet by the previous one and in turn the intermediate alphabets are converted back to the binary data to form the original data. Thus we provide a secured means of communication over the cloud.

Keywords: Data security, cloud computing idea

I. INTRODUCTION

One essential aspect for secure communications is that of Cryptography. The concept of securing messages through cryptography has a long history. Indeed, Julius Caesar is credited with creating one of the earliest cryptographic systems to send military messages to his generals. Cryptography is the science of using mathematics to encrypt and decrypt data. Cryptography enables you to store sensitive information or transmit it across insecure networks (like the Internet) so that it cannot be read by anyone except the intended recipient. While cryptography is the science of securing data, cryptanalysis is the science of analyzing and breaking secure communication. Classical cryptanalysis involves an interesting combination of analytical reasoning, application of mathematical tools, pattern finding, patience, determination, and luck. Cryptanalysts are also called attackers. Cryptology embraces both cryptography and cryptanalysis.

In today’s world, data protection and privacy are the two important issues on which clients and servers should concentrate upon with the daily growing population the requirement for digital data is increasing abruptly. Therefore it is very essential to provide the data security to the data sent on cloud. The data confidentiality is the very important factor to be considered. In order to provide the data security over the network different approaches have been developed. One of the popular way of providing the data security is by implementing data hiding technique.

To solve the raised difficulty, combining different ideas can help to achieve an acceptable level of confidentiality in Cloud Computing environments.

One of the recent data hiding technique was by using DNA reference sequences to encrypt and decrypt the text data.

In this technique a reference sequence will be chosen from NCBI database to implement the algorithm. However the disadvantage lies in the selection of sequences as there are multiple occurrence of same sequence.

Access control systems require time-trusted and reliable personal recognition. To overcome the problems faced by these processes individually, we can use a combination of two or more security processes.

Encryption is a security method in where information is in the form of plain text is encoded and converted in the form of a meaningless text format called cipher text. In such a way that only authorized user can read it. It uses encryption algorithm to generate cipher text that can only be read if decrypted. Decryption involves with the process of converting the cipher text back to the plain text.
There are two types of encryptions

1. Symmetric Key Encryption
2. Public Key Encryption

1. Symmetric key encryption: This algorithm uses same cryptographic keys for both encryption and decryption of cipher text as shown below.

2. Public key encryption: This algorithm uses two keys; they are private key and public key. Where public key is used for encryption and private key is used for decryption respectively. These two keys are mathematically linked with each other.

In the previous method of data hiding the reference sequence had been chosen from NCBI data base. The approach uses an interesting feature of DNA. In order to convert binary data into amino acids as a DNA sequence, the base pairing rules must be used. Synthesizing nucleotides in real environment (biology) is done in constant rules:

- Purine Adenine (A) always pairs with the pyrimidine Thymine (T)
- Pyrimidine Cytosine (C) always pairs with the purine Guanine (G)

Always, those rules are done naturally because the opportunities to synthesize hydrogen bonds between A and T (two bonds), and also between C and G (three bonds) is different, basically (hydrogen bonds have been shown with dotted lines in Fig 1). These concepts are named Watson-Crick base pairing rules when they discovered DNA’s fundamental structure as a Noble prize.

In binary computing area, it is possible to change the natural rules by own decision. For example, in biology A is synthesized to T while we can assume A to C or A to G, and so on, as we prefer. Increasing the complexity of the algorithm is the main purpose of the changing the rules. In this paper, the authors consider A=00, T=01, C=10, and G=11 to convert binary data to DNA sequences. A way to increase the complexity is complementary pair rule. Complementary pair rule is a unique equivalent pair which is assigned to every nucleotides base pair:

![Fig: DNA Reference Sequence](image)

M=100111000011

Sub-phase1 ((A= 00, T= 01, C= 10, G= 11)): M′= CTGAAG

Sub-phase2 ((AC) (CG) (GT) (TA)): M′′= GATCCT

Sub-phase3 (Indexes): M′′′=8137

Now, embedding phase is finally completed. Then, sender sends 8,13,7 to the cloud. In the next section, the client 2 will apply the extracting phase.

DNA Reference Sequence:

M′′′=8137

Sub-phase1 (Indexes): M′= GATCCT

Sub-phase2 ((AC) (CG) (GT) (TA)): M′= CTGAAG

Sub-phase3 (A= 00, T= 01, C= 10, G= 11): M=100111000011

So, the receiver extracted the original data, accurately by using a simple algorithm.

The above method has some limitations:

1. Lengthy genome sequence.
2. Possibility of occurrence of repeated sequences.
3. One has to choose the sequence intelligently.

In order to overcome this in our paper we present an effective way of hiding the data without using the reference sequence. First, client1 must apply the method of data hiding on its data which it wants to hide to the cloud computing environments. This section (hiding data) is divided into two phases. The first one is, embedding data and the second one is, extracting the original data.
**STEP 1: Data Encryption**

In this step, the given text data is converted into respected binary form using ASCII to binary conversion method. Here the client decides to upload the file via a network to cloud computing environments. So, there are three sub-phases to provide the final form of data which is \( D''' \) and upload it to cloud.

The plain data is \( D \) which is in binary format is converted into alphabet sequence using the first 4 alphabets in the sequence A,B,C,D now the form of data is represented using \( D' \). Now the \( D' \) contains the alphabetic sequences of binary data. Now the modified data \( D' \) is made complex by replacing each alphabet by its very next alphabet in the sequence this will form the next set of data represented by the actual alphabet sequence i.e \( D'' \). The same data is sent to the cloud as secret data as shown below.

Example: \( D = 1000100011 \) Let \( A = 00, B = 01, C = 10, D = 11 \)

Now \( D' = \text{CACAD} \) and \( D'' = \text{DBDBE} \) now the secret data \( D''' = \text{DBDBE} \) will be sent to the cloud.

**STEP 2: Data Decryption**

In this step, the secret data is converted into the altered alphabet sequence and this altered alphabet sequence is converted back to the alphabet sequence in its generated form and finally. Generated alphabet sequence is mapped into the original plain text. In other words the encryption algorithm steps are reversed to implement the decryption process. The secret data \( D''' \) which is sent to cloud is converted back to \( D'' \) altered sequence by replacing each alphabet by the previous alphabet of actual alphabet sequence.

The plain data is \( D \) which is in binary format is converted into alphabet sequence using the first 4 alphabets in the sequence A,B,C,D now the form of data is represented using \( D' \). Now the \( D' \) contains the alphabetic sequences of binary data. Now the modified data \( D' \) is made complex by replacing each alphabet by its very next alphabet in the sequence this will form the next set of data represented by \( D'' \). now the alphabet sequence is replaced by the alphabet sequence using the alphabets A B C D. Finally the text data sequence is replaced by binary numbers to form the plain text..

Example: \( D''' = D'' = \text{DBDBE} \) Now \( D' = \text{CACAD} \) in order to convert the \( D' \) to \( D \) consider \( A = 00, B = 01, C = 10, D = 11 \) Hence the plain text will be \( D = 1000100011 \)

**Advantages**

1. The users are allowed to select 4 letters A,B,C,D with the binary values 00,10,10,11 respectively. Hence all the binary coding rules are \( 4 \times 3 \times 2 \times 1 = 24 \). So, the likelihood of making correct guess by attacker is \( 1/24 \).

2. Pairing rule: like binary coding rule, there is \( 4 \times 3 \times 2 \times 1 = 24 \) alphabet Therefore, the possibility of making successful attack is \( 1/24 \).

3. Easily implementable algorithm and no complexity in pairing rules.

**II. CONCLUSION**

As security is a major concern in cloud environment. Our paper proposes an easy, efficient and secure way of data transfer without using any reference sequence by using simple pairing rules. It is easy for the sender to implement the encryption and decryption steps on the data to be sent and received. The algorithm is also effective in case of security issues.

**ACKNOWLEDGEMENT**

I would like to thank my guide Prof. Dr. Guruprakash C D for his kind support to come up with this paper. I would also like
to thank all other coauthors for their support to publish this paper

REFERENCES

[9]. A Symmetric Key Cryptographic Algorithm Ayushi Lecturer, Hindu College of Engineering H.No:438, sec-12, sonipat, Haryana