International Journal of Current Engineering and Technology

Research Article

A DNA Based Secure Data Hiding Technique for Cloud Computing

Neha Pallavi*, Archana Singh and Surya Prakash Dwivedi

Department of Computer Science and Engineering, SHIATS, Allahabad, India

Accepted 03 July 2016, Available online 11 July 2016, **Vol.6, No.4 (Aug 2016)**

Abstract

Cloud computing is a set of IT services that are provided to a customer over a network on a leased basis and with the ability to scale up or down their service requirements. Usually cloud computing services are delivered by a third party provider who owns the infrastructure. Cloud computing provides people the way to share private resources and services. Since data are share between distributed resources via the network in the open environment there is always security threat. Due lack of proper security control policy and weakness in saving private data that lead to many vulnerability in cloud computing. This paper proposes alternative security methods based on DNA. A mechanism has been proposed and simulated for secure data transfer. The main objective of this paper is to study the existing approaches of secure data transfer over network in Cloud using proper authentication and propose a solution using DNA technique.

Keywords: *Cloud Computing, DNA Computing, DNA Sequence, Data Hiding.*

1. Introduction

Cloud Computing: Cloud Computing is an evolving term, defined more by usage than by written documents. Cloud Computing has evolved over the past from utility computing, autonomic computing and grid computing through the sharing of resources, computation and storage capabilities. According to National Institute of Standards and Technology cloud is a model for enabling ubiquitous, convenient, on demand network access to shared pool of resources which can be provisioned rapidly with minimum management and service provider interaction. Cloud computing security is an apprehension that needs great courtesy. Cloud has been prone to various security issues like storage, computation and attacks like Denial of service, Distributed Denial of Service, Eavesdropping, insecure authentication, Man-in-the-middle attack or logging etc. That's why in this paper we propose a model using DNA strand and DNA technique in order to avoid these attacks.

DNA Computing: DNA was proposed for computation by Adelman in 1994. After that many approaches have been investigated. Theoretical consideration that it may simulate Turing machines and cryptography. It has been shown that DNA computing was suitable for some problems currently hard to resolve (intractable) and could work faster than electronic computer. A good background between the DNA molecule and computer engineering are required to develop the efficient algorithms of DNA computing .After Adelman solved the Hamilton Path Problem using a combinatorial molecular method, many hard computational problems were calculated by DNA computer. DNA cryptography is based on DNA computing where, message is encrypted in the form of DNA nucleotide sequence. DNA computing can be used as conceptual platform for data encryption and decryption by using symmetric or asymmetric key. In current scenario it is not much effective than traditional cryptography but it can provide a hybrid security by combining traditional cryptography with it. However DNA logic can be implemented with traditional cryptography. The ultimate target is to scramble data in the way that the person who doesn't know the key, can't read or modify data. DNA is introduced as a new technology for unbroken data. Genetic information is encoded as a sequence of nucleotides Guanine-G, Adenine-A, Thymine-T and Cytosine-C. Adenine, Thymine and Guanine, Cytosine are base pairs, which are attached to a sugar and a phosphate to maintain helical structure. DNA strands combined with hydrogen bond. A and T DNA sequences are combined with double hydrogen bond while C & G are combined with triple bond. Each nucleotide consists of the following three components, A Nitrogenous Base, A five carbon Sugar, A Phosphate Group.

2. Related Work

In 2004, Sabari Pramanik *et al*., presented the parallel cryptography technique by using DNA molecular

*Corresponding author **Neha Pallavi** and **Surya Prakash Dwivedi** are M.Tech Scholars; **Archana Singh** is working as Assistant Professor

structure, one time pad, DNA digital coding technique and DNA hybridization technique. One-time-pad technique is used for encryption key, which increases computational complexity.

In 2005, Kazuo Tanaka proposed the DNA cryptographic approach based on Public Key. Public key is used to encrypt message in DNA sequences and encrypted message sequence forwarded to the immobilization process and then for PCR amplification. Polymerase Chain Reaction (PCR) amplifications are used two primers to encode a message.

In 2006, Sherif T. Amin proposed the DNA cryptographic approach based on symmetric key, where key sequences are obtained from the genetic database and remain same at both ends (sender and receiver). Message/plaintext is first converted into binary format and then to DNA format using substitution.

Anil Kurmus *et al*. demonstrates comparison of two own multi-tenancy architectures defined at different levels, one at operating-system kernel level and second at hypervisor level. Both these architectures are analyzed as solution to security risks such as data integrity, malicious customer, unauthorized data access and confidentiality.

Sangdo Lee *et al*. have defined new term called as rain cloud system. In this model, libraries are used to manage different CSPs. Further, actual data storage using library interface is demonstrated.

In 2000 Leier and *et al*. brought a robust technique by utilizing a special key strand. They called it, primer. Primer has key role to decrypt a coded strand.

Bibhash Roy *et al* proposed a DNA sequencing based encryption and decryption process. This paper also proposes a unique cipher text generation procedure as well as a new key generation procedure. But the experimental result shows that the encryption process requires high time complexity.

Pankaj Rakheja designed a new method by integrating DNA computing in IDEA. Such conceptual works can be useful in the development of this new born technology of cryptography to fulfill the future security requirements.

3. Materials and Methods

3.1 Deoxyribonucleic Acid (DNA) Bio-Logical Theory

DNA is a molecule that caries most of the genetic instructions used in the development functioning and reproduction of all known living organisms and many viruses. DNA is a nucleic acid; alongside proteins and carbohydrates, nucleic acids compose the three major macromolecules essential for all forms of life. Most DNA molecules consist of two biopolymer strands coiled around each other to form a double helix. The two DNA strands are known as polynucleotides since they are composed of simpler units called nucleotides. Each nucleotide is composed of a nitrogen-containing nucleobase- either cytosine(C), guanine(G), adenine(A) or thymine(T) as well as a monosaccharide sugar called deoxyribose and a phosphate group. The nucleotides are joined to one another in a chain by covalent bonds between the sugar of one nucleotide and phosphate of the next, resulting in an alternating sugar-phosphate backbone. According to base pairing rules (A with T and C with G), hydrogen bonds bind the nitrogenous bases of two separate polynucleotide strands to make double stranded DNA. A DNA molecule is composed of two single strands which form a double helix structure shown in figure 1.

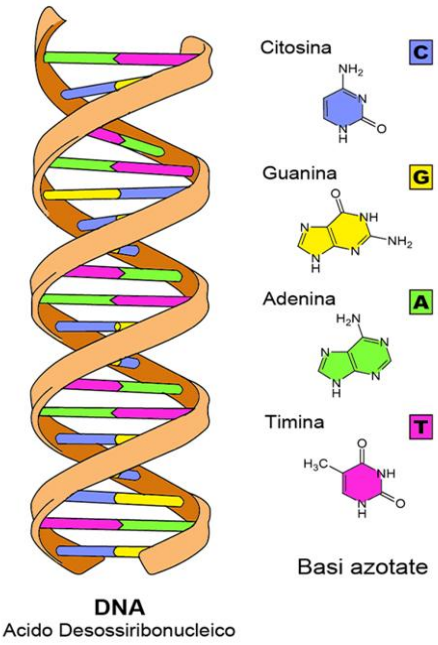

Figure 1 Basic DNA Structure

3.2 Complementary Rule of DNA

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.

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.

For Example:

Complementary rule: ((AC) (CG) (GT) (TA))
DNA strand: AATGC
Applying complementary rule on DNA strand:
Original Strand: AATGC
Complementary Strand: CCATG

Nucleotide	Decimal Value	Binary Value
A	0	00
T	1	11
C	2	01
G	3	10

3.3 Biological View of DNA Encryption

DNA encryption is a next generation security mechanism, storing almost a million gigabytes of data inside bacteria. Research from two prominent universities indicates that it is not only possible but also practical to store digital data in the genome of a living organism and retrieve that data hundreds or even thousands of years later, after the organism has reproduced its genetic material through hundreds of generations.

DNA Computing in Cryptography

DNA cryptography is far away from realization because in current time it can be performed only in labs using chemical operations. In order to provide better security and reliable data transmission an effective method of DNA based cryptography is proposed here. In this method the mixture of mathematical and biological concepts are used to get the encrypted data in the form DNA sequences.

To understand the data hiding through DNA complementary rule, separating this process into different steps.

Step 1: Embedding Secret Data in order to explain embedding phase, separating the phases is the best way of proposing current method. In below, sub-phases have been shown, respectively.

Firstly convert the message into the binary form---
For example the original data is to be send is M = 110111000010
DNAreferencesequence=$AT_1CG_2AA_3TC_4CG_5CG_6CT_7GA_8GT_9CA_{10}CA_{11}AT_{12}TA_{13}GC_{14}GC_{15}TG_{16}AG_{17}TG_{18}AA_{19}CC_{20}$
Encryption of the original message is as follows
M=110111000010
Sub $Part_1$ (T = 00, A = 01, G = 10, C = 11)
M' = CTGAGC
Sub-$part_2$ ((Ag), (CA), (GT), (TC)).
M= ACTGTA
Sub-Part3 (Picking Indexes);M''' =81614
Thus, embedding phase is completed; client sends 81614 to the cloud. In the next section, the receiver will apply the extracting phase for extracting the original data by using three consecutive phases.
Step 2: Extraction of Original Data
DNAreferencesequence=$AT_1CG_2AA_3TC_4CG_5CG_6CT_7GA_8GT_9CA_{10}CA_{11}AT_{12}TA_{13}GC_{14}GC_{15}TG_{16}AG_{17}TG_{18}AA_{19}CC_{20}$
M'''= 81614
Sub-$Part_1$ (Indexes from reference sequence);
M = ACTGTA
Sub- $Part_2$ ((AG), (CA), (GT), (TC)):
M' =CTGAGC.
Sub-$Part_3$ (T = 00, A = 01, G = 10, C = 11):
M = 110010011011.
So, the receiver extracted the original data, accurately by using a simple algorithm.

3.4 Diffie-Hellman Key Exchange Algorithm

Diffie-Hellman is a way of generating a shared secret between two people in such a way that the secret can't be seen by observing the communication. That's an important distinction: You're not sharing information during the key exchange, you're creating a key together.

The basic idea works like this:
1. I come up with two prime numbers g and p and tell you what they are.
2. You then pick a secret number (a), but you don't tell anyone. Instead you compute ga mod p and send that result back to me. (We'll call that A since it came from a).
3. I do the same thing, but we'll call my secret number b and the computed number B. So I compute gb mod p and send you the result (called B)
4. Now, you take the number I sent you and do the exact same operation with it. So that's Ba mod p.
5. I do the same operation with the result you sent me, so: Ab mod p.
The magic here is that the answer I get at step 5 is the same number you got at step 4. Now it's not really magic, it's just math, and it comes down to a fancy property of modulo exponents. Specifically:

$(g^a \bmod p)^b \bmod p = g^{ab} \bmod p$
$(g^b \bmod p)^a \bmod p = g^{ba} \bmod p$

3.5 Architecture of Proposed Technique

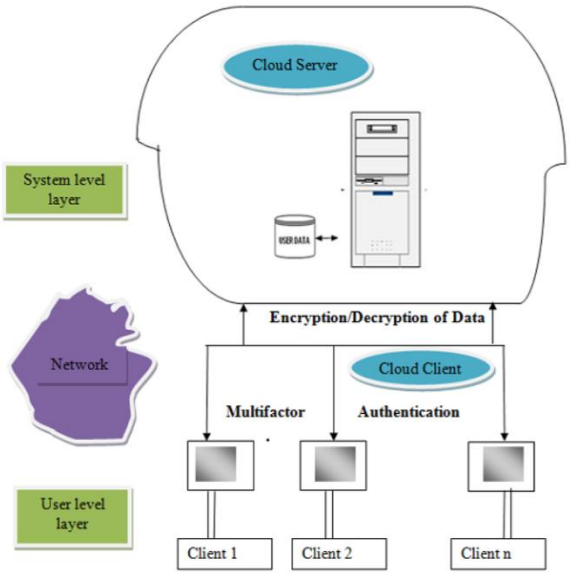

Figure 2: Proposed Architecture

Conclusion

a) In this thesis existing secure data transfer techniques have been analyzed and compared according to their features.
b) The design has been proposed while encryption/decryption working and DNA technique

has been used to provide the message integrity so that intruder can`t alters the operation in transit.

References

P. Mell (2011), The NIST Definition of Cloud Computing, National Institute of Standards and Technology, pp. 1-3.

L. Adelman (Nov. 11, 1994), Molecular Computation of Solutions to Combinatorial Problems, Science266:1021-1024.

Tushar Mandge (2013), A DNA Encryption Technique Based on Matrix Manipulation and Secure key Generation Scheme, ICICES Journal.

Guangzhao Cui Limin (2008) An encryption scheme using DNA technology. Bio-Inspired Computing: Theories and Applications, 2008. BICTA 2008. 3rd International Conference on Publication Date: Sept. 28 2008-Oct. ISBN: 978-1-4244-2724-6, page(s):37-42; Adelaide, SA.

Guangzhao Cui Limin (2008), An Encryption Scheme Using DNA Technology, IEEE, 978-1-4244-2724-6/08

Pramanik Sabari (2012), DNA cryptography, In Electrical & Computer Engineering (ICECE), 7th IEEE International Conference on, pp. 551-554.

Kazuo Tanaka (2005) Public-key system using DNA as a one-way function for distribution. Bios stems 81, 1, pp. 25-29.

Sherif T. Amin (2006) A DNA-based implementation of YAEA encryption algorithm, In Computational Intelligence, pp. 120-125.

A. Kurmus, M. Gupta (June 2011) A Comparison of Secure Multi-tenancy Architectures for File System Storage Clouds, ACM International Conference on Middleware, pp. 471- 490.

S. Lee, H. Park (2012)Cloud Computing Availability: Multi-clouds for Big Data Service, in Convergence and Hybrid Information Technology (6th International Conference, ICHIT 2012 Proceedings book), Volume 310, New York: Springer, pp 799-806.

A. Leier (2000), Cryptography with DNA binary strands, Bio Systems 57.

Bibhash Roy (2011), An improved Symmetric key cryptography with DNA Based strong cipher-ICDeCom, Feb' 24-25'2011, pp.1-5.

Bibhash Roy *et al* (2011), A DNA based Symmetric key Cryptography-ICSSA- 2011, 24-25 Jan'11.

Bibhash Roy (Dec 201), An Enhanced key Generation Scheme based cryptography with DNA Logic-IJICT-2010-11, Volume 1 No. 8,

Nucleotide base pairing of strands (2012), http://dedunn.edblogs.org