Secure Data Storage On Multi-cloud Using DNA Based Cryptography

Sagar Dhuri[1], Prakash Naikade[2], Nilesh Gade[3], Abhijeet Teke[4], Prof. Mrs. D S Zingade[5]
[1, 2,3,4,5] Department of Computer Engineering,
AISSMS Institute of Information Technology
[1]sag.ash773@gmail.com, [2]prakashknaikade@gmail.com, [3]nileshgade94@gmail.com, [4]abhiteke007@gmail.com

Abstract - Cloud computing has a great capability to boost productivity and minimize costs, hence many users are embracing it, but at the same time it constitutes many security risks and challenges. Due to possibilities of multiple risks such as service outage, theft of data, data leakage and the chances of malicious insider attack, using single cloud is becoming obnoxious by many users and new notion of Multi-Clouds usage is becoming perceptible to cope with these security issues. We aim to demonstrate powerful security strategy of using DNA based cryptography which ensures secure data storage on multi-cloud.

Keywords: DNA sequence, Cloud Computing, DNA base pairing rules, DNA binary coding.

I. INTRODUCTION

Cloud computing offers great potential, but at the same time it presents many security risks and challenges. Using single cloud provider is becoming less popular due to service availability failure risk and the possibility of malicious insiders in the single cloud. Solution that comes up recently is multi-clouds, or in other words, inter-clouds or cloud-of-clouds.

Better security and data availability can be achieved by breaking down the users critical data block into parts and dispersing them among the available Cloud Service Providers (CSP). Each divided part of data can be further protected by utilizing some interesting features of DNA sequences and data hiding.

This project aims to review DNA Encryption as a possible solution for security and privacy concerns of cloud. Deoxyribonucleic acid [DNA] is a long polymer made from repeating units called nucleotides. DNA was first identified and isolated by Friedrich Miescher and the double helix structure of DNA was first dis-covered by James Watson and Francis Crick, using experimental data collected by Rosalind Franklin and Maurice Wilkins. The structure of DNA of all species comprises two helical chains.

Watson-Crick Base Pairing Rules:

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).

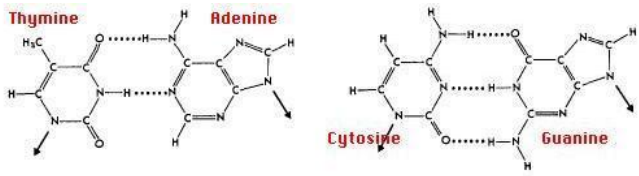


Fig. Watson-Crick Base Pairing Rule.

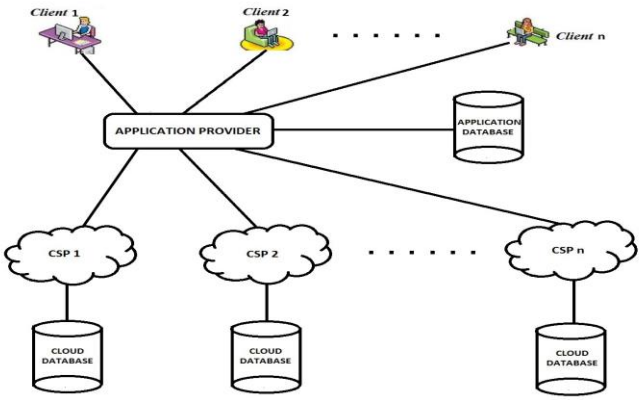
DNA in Computing Area:

In order to gain more complexity and to make it hard for intrusion by attacker, we will change these universal rules. For example, in biology A is synthesized to T while we can assume A to C, T or G, anything and so on, as we decide.

Multi-cloud:

Data availability, security, privacy, and integrity are the most critical issues to solve in cloud computing. Even though the cloud service providers have powerful infrastructure along with standard regulations to provide a better availability and ensure customers data privacy, there still exist many reports of service outage and privacy breach in last few years. Solution is using multiple clouds for storing critical data.

II. SYSTEM MODEL



ARCHITECTURE DIAGRAM

The proposed conceptual architecture is three–tier architecture. The first tier is user, second is application provider (server), and third is cloud service provider (CSP).

The client i.e. user interface level is ready developed interface which is capable to register, upload files, retrieve saved files, delete previously saved file and even update its own information as and when required. The system proposes to do encryption model at client side by using its own resources to enhance data security and reduce load on the server. This gives double advantage to the system model proposed.

The second tier, application provider is a server handling the incoming request and outgoing replies from and to the clients. It plays vital role of an interface between clients and cloud service provider. It is even responsible for file segmentation.

The cloud service provider gives storage space so the clients can store their critical data. As client does not have direct access to this storage space, so it's difficult to crack this system. The security levels are discussed in next sections.

III. PREVIOUS WORK

One of the famous ways to protect data through the Internet is data hiding. Because of the increasing number of Internet users, utilizing data hiding or Steganographic techniques is inevitable. Eliminating the role of the intruder and authorizing the clients are eventual goals of these techniques. Therefore, the role of data hiding has become more eminent nowadays. Before employing biological properties of DNA sequences [1], the common way of embedding a secret data into the host images was the traditional way of data hiding. It unfortunately leads to some liabilities. The most important ones was the detection of the distortions of the image when the host image changed to some degrees. That was the best spot to start the wholly detection of the secret data through the image. By advent of biological aspects of DNA sequences to the computing areas, new data hiding methods have been proposed by researchers, based on DNA sequences. The key portion of their work [1] is, utilizing biological characteristics of DNA sequences.

Always, the rules are done naturally because the opportunities to synthesize hydrogen bonds [2] between A and T (two bonds), and also between C and G. These concepts are named Watson-Crick base pairing rules [2].

Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort.

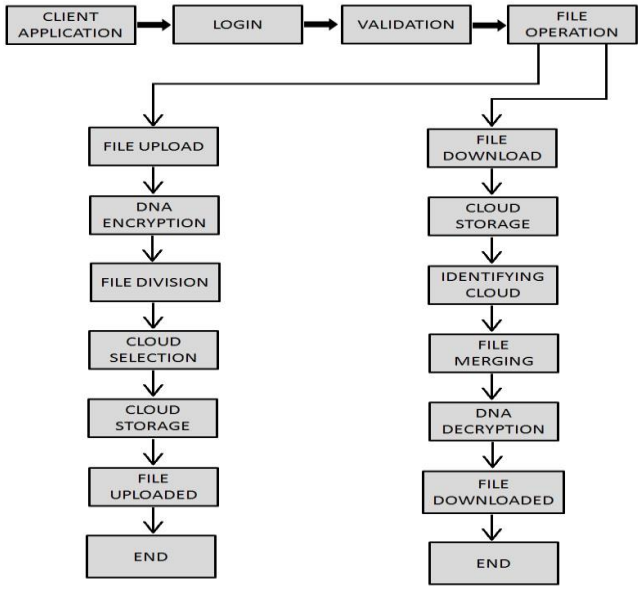
The SaaS (software as a service) cloud model can be defined as a cloud deployment model in which centrally built software is used for users to access the cloud. Cloud computing offers a data storage system which enables the users to be less dependent on the client system and provides an architecture to upload the data in a cloud that can be shared by multiple users and also provide security through authentication of the user. This architecture [3] introduces a secondary cloud controlled by a single administrator which provides the data backup for primary cloud after undergoing specific segmentation and encryption algorithms [4] to ensure security and integrity of data.

Better security can be obtained by dispersing the user data over multiple cloud service providers (CSP) in such a way that, none of the CSP can successfully retrieve meaningful information from the data pieces allocated at their servers. Also, because of redundancy in data distribution, user is assured of data availability. If a service provider goes bankrupt or suffers service outage, the user still can access his critical data from other CSPs. Thus advantages of using Multi-cloud [5] storage is data availability, avoid vendor lock-in, business continuity and disaster recovery.

Given p number of cloud service providers, User will divide his data into N data pieces where at-least k data pieces out of N data pieces are necessary to recover any meaningful information of the data; as data redundancy is used. This (k, N) is the first threshold. Second threshold is of (q, p); which implies, at least q out of p number of CSPs must be a part of retrieval process for successful data retrieval [6].

IV. PROPOSED METHODOLOGY

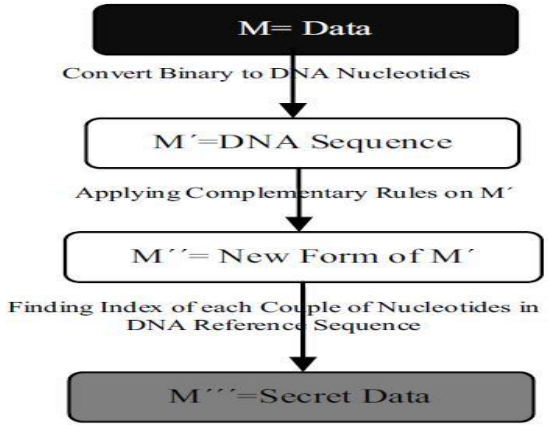
In this system if there are some no of clients register, then they can store their data on multiple clouds as segments. The following figure displays flow of data.



BLOCK DIAGRAM OF PROPOSED SYSTEM

Embedding Secret Data:

In order to explain embedding phase, separating the phases into some successive and vivid sub-phases, is the best way of proposing current method. In below, sub-phases have been shown, respectively.



Obviously, there is an original data M which the client decides to upload via a network to cloud computing environments. So, there are three sub-phases to provide the final form of M which is M´´´ and upload it to cloud. The first sub-phase is, converting by DNA base pairing rules. The product is M´. M´ contains nucleotides sequences. By applying DNA base pairing rules, the data can convert from binary to DNA sequence. Not only DNA base pairing helps to encrypt the data from binary to DNA sequence but also it is applied to decrypt the secret data to original one, truly.

The next (second) sub-phase is, applying complementary rules. Increasing the complexity is the real and exact purpose of this step. By applying the complementary rules, the new form of the M´ which is M´´ emerges. Now, M´´ is appeared. As mentioned before, both of clients have a DNA reference sequence from a large number of possibilities base on EBI [7] or NCBI [8] database. It means that, they have selected the same DNA reference sequence, exactly. The exact role of the third sub-phase is, extracting the index of each couple nucleotides in DNA reference sequence, numerically. When all the indexes have been extracted, M´´´ has been made, properly. M´´´ is precisely the secret data with some changes through the embedding phase. Now, sender can send the data (M´´´) to cloud. Clarification of the current phase is continued by demonstrating an example, step by step. In this example, assume original data M=100111000011 should be uploaded to the cloud.

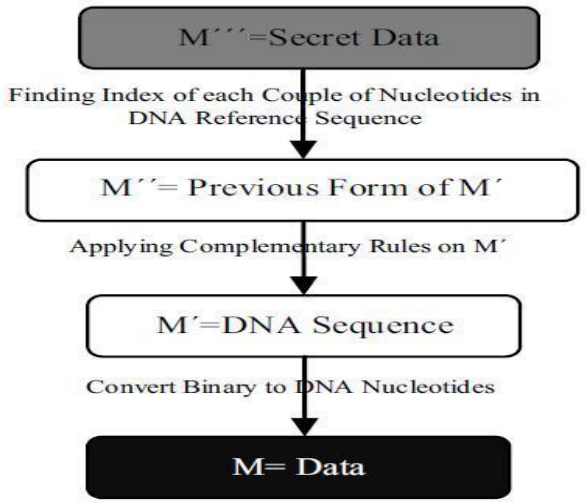
DNA Reference Sequence:
$AT_1CG_2AA_3TT_4CG_5CG_6CT_7GA_8GT_9CA_{10}CA_{11}AT_{12}TC_{13}GC_{14}$
$GC_{15}TG_{16}AG_{17}TG_{18}AA_{19}CC_{20}$
M=100111000011
Sub-phase1 $_{(A=\ 00,\ T=\ 01,\ C=\ 10,\ G=\ 11)}$: M´= CTGAAG
Sub-phashe2 $_{((AC)\ (CG)\ (GT)\ (TA))}$: M´´=GATCCT
Sub-phase3 $_{(Indexes)}$: 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 for extracting the original data by using three consecutive phases.

Extracting Original data:

Now, client2 takes the secret data in form of some numbers. For the purpose of extracting the original data from DNA reference sequence, phase two with its sub- phases will extract the original data, correctly.

So, the first sub-phase manipulates the M´´´. Because of the nature of the secret data which is some sorts of numbers (exact positions (indexes) of the original data on DNA reference sequence), extracting the data starts by finding the indexes on DNA reference sequence one by one according to the numbers which sender has sent in form of the current secret data. M´´ is the exact product of the first sub-phase. Consequently, the second sub-phase applies complementary rules on M´´ in order to extracting M´, correctly.



The importance of the M´ is the form of it. M´ is the last form of data, based on DNA nucleotides. Converting the M´ to the M is the third sub-phase. Transforming from DNA nucleotides to the binary is the responsibility of the last sub- phase. Now, the client2 has truly extracted the original data M. Those steps are demonstrated through the example in below:

DNA Reference Sequence:
$AT_1CG_2AA_3TT_4CG_5CG_6CT_7GA_8GT_9CA_{10}CA_{11}AT_{12}TC_{13}GC_{14}$
$GC_{15}TG_{16}AG_{17}TG_{18}AA_{19}CC_{20}$
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. In the next section, security and liabilities of the algorithm will inspect, briefly.

V. CONCLUSION

Cloud computing is restructuring how IT resources and services to be used and managed, but major problem in cloud implementation is security challenges. By dividing user's data and applying data hiding using DNA encryption, and then storing it on multiple clouds; this model has shown its ability of providing a cloud customer with a more secured storage. The proposed concepts discussed here will help to build strong security architecture in cloud computing. This will also improve customer satisfaction and will attract more investors for industrial as well as future research farms.

VI. FUTURE SCOPES

Many applications are moving to the cloud, so, it is possible to think of new applications that would use the storage cloud as a back-end storage layer. The system software can be extended in the future to include Java Platform Enterprise Edition technologies like JSP, Servlets along with other advanced functionalities such as storing and sharing the data. Storing and sharing the data on cloud is much easy and secure with this proposed system. Data availability issue will be solved completely in future as we are addressing.

REFERENCES

[1] Deepak Kumar, Shailendra Singh, "Secret Data Writing Using DNA Sequences", 978-1-4577-0240-2/11/$26.00 IEEE, 2011.
[2] D.Sureshraj, Dr.V.Murali Bhaskaran, "Automatic DNA Sequence Generation for Secured Effective Multi-Cloud Storage", SR Journal of Computer Engineering (IOSR-JCE)e-ISSN: 2278-0661, p-ISSN: 2278-8727Volume 15, Issue 2, Nov- Dec2013.
[3] Mohammad Reza Abbasy, Bharanidharan Shanmugam, "Enabling Data Hiding for Resource Sharing in Cloud", IEEE World Congress on Services, 2011.
[4] Richa H. Ranalkar , Prof. B.D. Phulpagar, "Review on Multi-Cloud DNA Encryption Model for Cloud Security", ,Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 3, Issue 6, Nov-Dec 2013.
[5] Zicheng Wang, Xiaohang Zhao, Hong Wang and Guangzhao Cui, "Information Hiding Based on DNA Steganography", 978-1-4673-5000-6/13/$31.00 IEEE, 2013.
[6] K. Menaka, "Message Encryption Using DNA Sequences", World Congress on Computing and Communication Technologies, 2014.
[7] National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/
[8] European Bioinformatics Institute, http://www.ebi.ac.uk/