Scientific Journal of Impact Factor (SJIF): 3.134
ISSN (Online): 2348-4470
ISSN (Print) : 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2, Issue 3, March -2015

SECURE DATA STORAGE ON MULTI-CLOUD USING DNA BASED CRYPTOGRAPHY

Prof. Mrs. D S Zingade[1], Sagar Dhuri[2], Prakash Naikade[3], Nilesh Gade[4], Abhijeet Teke[5]

[1,2,3,4,5] *Department of Computer Engineering, AISSMS IOIT*

Abstract — *Cloud computing has a great potential to enhance productivity and reduce costs , hence many users are preferring it over various traditional technologies, but at the same time it gives rise to many security risks and challenges such as theft of data , data leakage and denial of service. Hence using single cloud is becoming obnoxious and the concept of multi-cloud is gaining popularity. This paper demonstrates use of DNA based cryptography to ensure secure data storage on multi-cloud.*

Keywords-*DNA Sequence, DNA Base pairing rule, Cloud computing, Multi-cloud Architecture.*

I. INTRODUCTION

Cloud computing is gaining immense popularity but has many security risks and challenges associated with it. Due to service availability failure risks and possibilities of malicious attack, use of single cloud is becoming less likely favoured. Upcoming solutions include multi-cloud technology.

By dividing the data block into parts and distributing them among the various cloud service provider (CSP), better security aspects can be achieved. Each divided parts can be further protected by utilizing interesting features of DNA sequences.

Deoxyribonucleic acid [DNA] is a long polymer made from nucleotides. The double helix structure of DNA was first discovered by James Watson and Francis Crick. The DNA structures comprises of two helical chains. This helical structure contains four nucleotides namely Purine Adenine (A), Pyrimidine Thymine (T), Pyrimidine Cytosine (C), Purine Guanine (G).

Watson-Crick Base Pairing Rules:

To convert binary data into nucleotides as a DNA sequence, the base pairing rules are used. Synthesizing nucleotides naturally is done by constant rules:

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

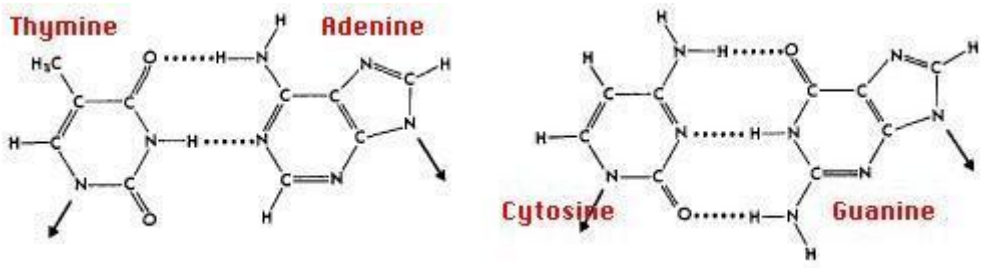

Fig. 1. Base Pairing Rules

DNA in Computing Area:

To increase the complexity and reduce risk of intrusion attack, universal rules can be modified. For Example, naturally C is synthesized to G while we can assume C to A or G or T or C itself.

II. SYSTEM MODEL.

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 readily 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 process the encryption algorithm at client side by using its own resources to enhance data security and reduce load on the server. This gives additional advantage to the model proposed.

The second tier, application provider is a server handling the incoming requests 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 that 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. PROPOSED METHODOLOGY

The system provides access to store and retrieve critical data over cloud for registered clients. In order to provide security, data hiding is achieved by strong mean namely DNA encryption followed by file segmentation algorithm.

A. Embedding Secret Data:

The embedding phase is divided into three successive sub-phases. The Figure 2 shows all the phases in brief.

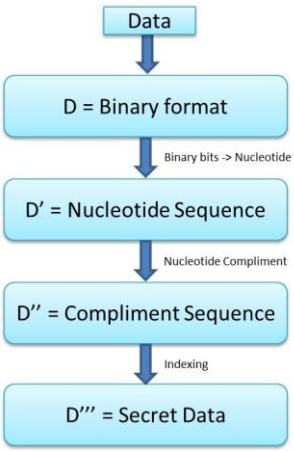

Fig. 2. Embedding Process

Let D be the binary format of data which client has to upload over the cloud space. The first phase is responsible for using DNA base pairing rule to produce D', which is nucleotide sequence. Base pairing rule is applied to convert binary bit pairs to representing nucleotides.

The second phase applies the DNA complimentary rule. Because of this phase the complexity of hidden data is increased. The product of this phase is D'' produced from D'.

The role of third phase is to extract the index of the nucleotide couple from the DNA reference string. All the DNA nucleotide couples in D'' are replaced by their respective indexes to form secret data D'''.

DNA Reference Sequence:

$CA_1AT_2TG_3CC_4GG_5CG_6CT_7AG_8GT_9GA_{10}AG_{11}TC_{12}CA_{13}\,GC_{14}AA_{15}TA_{16}TT_{17}AC_{18}TG_{19}CT_{20}$

Data = N
D = 01001110
Sub-phase1 $_{(A=\,00,\ C=\,01,\ G=\,10,\ T=\,11)}$: D´= CATG
Sub-phashe2 $_{((AT)\ (CA)\ (GC)\ (TG))}$: D´´= ATGC
Sub-phase3 $_{(Indexes)}$: D´´´= 0214

Secret data (D''') is sent to the server for further processing of file segmentation.

B. Extracting Original data:

While retrieving the original data from the secret data, the reverse algorithm is applied. Figure 3 shows the extraction of original data in brief.

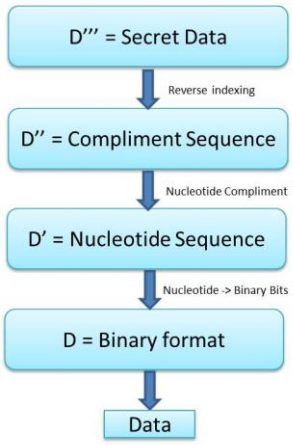

Fig. 3. Extracting Process

Consider first phase which manipulates secret data (D''') which is in the form of decimal numbers. These numbers are nothing but indexes of nucleotide couples in the DNA reference string. Replacing the indexes with their respective couple produces D''.

D'' is data in which every nucleotide represents the compliment of original nucleotide. Hence applying complimentary rule D' is obtained in second phase.

Converting D' to D is the responsibility of third phase. D' is nucleotide representation of binary pair. Hence every nucleotide is replaced by its respective binary values.

DNA Reference Sequence:
$CA_1AT_2TG_3CC_4GG_5CG_6CT_7AG_8GT_9GA_{10}AG_{11}TC_{12}CA_{13}\,GC_{14}AA_{15}TA_{16}TT_{17}AC_{18}TG_{19}CT_{20}$

Sub-phase1 $_{(Indexes)}$: D´´´= 0214
Sub-phashe2 $_{((AT)\,(CA)\,(GC)\,(TG))}$: D´´= ATGC
Sub-phase1 $_{(A=\,00,\,C=\,01,\,G=\,10,\,T=\,11)}$: D´= CATG
D = 01001110
Data = N

The original data is extracted from the secret data applying these phases successively.

C. File Segmentation

To increase complexity in data hiding the file is not stored as a whole on a single cloud. Multi-cloud technology comes into existence. The file is divided into different parts which can be stored over different cloud spaces to give immense security to client's critical data.

Consider D''' = {0214}, here {02} and {14} can be stored on different cloud spaces. The data can even be further divided and stored.

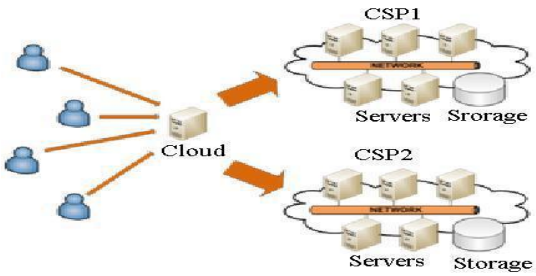

Fig. 4. Multi-cloud architecture

To increase data availability different divided parts can be combined and stored such that even if at all any cloud has failed still the data can be retrieved from the remaining clouds.

IV. CONCLUSION

The major drawbacks of single cloud architecture are the security challenges faced. A solution is provided by applying DNA encryption for data hiding and dividing this critical data to be stored on multiple clouds. This technique

has ability to provide more secured storage and the methodology discussed will help to build a strong security architecture in cloud computing. It will definitely improve the data availability resulting in customer satisfaction.

V. 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.MuraliBhaskaran, "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, BharanidharanShanmugam, "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/

