RESEARCH ARTICLE OPEN ACCESS

Review on Multi-Cloud DNA Encryption Model for Cloud Security

Richa H. Ranalkar *, Prof. B.D. Phulpagar**
*(M.E. Student, Department of Computer Engineering, Modern College of Engineeting, Pune – 05)
** (Professor, Department of Computer Engineering, Modern College of Engineeting, Pune – 05)

ABSTRACT
Cloud computing is one of the fastest growing areas in the IT industry, with huge potential of cost-saving capabilities along with increased flexibility and scalability for system resources. Although there are great benefits associated with cloud computing, it also presented new security threats/challenges. Due to risks of service availability, failure and the possibility of malicious insiders in the single cloud, single cloud is becoming less popular and new concept of using Multi-Clouds is becoming evident to solve these security issues. This paper reviews DNA Encryption to secure data while storing it on Multi-Clouds.

Keywords **-** Cloud computing, Cloud security, Cloud service provider (CSP), DNA Encryption, DNA sequence and Multi-Cloud

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 user's 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 paper aims to review DNA Encryption as a possible solution for security and privacy concerns of cloud.

II. LITERATURE REVIEW

Mohammad A. Alzain *et al.* [1] proposed a multi cloud database model. The proposed work encompasses comparison of this model with Amazon cloud. They have demonstrated how this proposed model is better for data storage and retrieval. The model is analysed for addressing data integrity, data intrusion, and service availability.

Anil Kurmus *et al.* [2] have proposed two multi-tenancy architectures one at hypervisor level and other at operating-system kernel level. Architectures are compared as a solution to security problems such as malicious customer, confidentiality, data integrity and unauthorized data access.

Sangdo Lee *et al.* [3] proposed a rain cloud system model. Different cloud interface providers are managed by library. His work aimed to demonstrate data storage for a big rain cloud system.

According to S. Jaya Prakash *et al.* [4] a multi cloud system where data is replicated into different cloud providers is a solution to reduce the service availability risk or loss of data. To overcome security risk of single point of contact, multiple clouds from different unrelated cloud providers should be used.

III. 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 customer's 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. An example of multi- cloud architecture is as shown in Fig. 1.

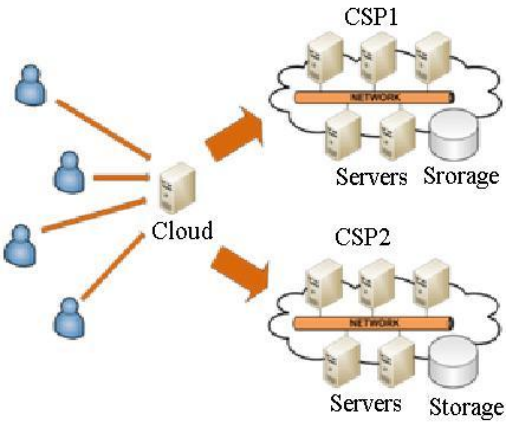

Fig. 1: Multi-Cloud Architecture.

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 storage are 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 [5].

Thus minimum number of pieces that must be chosen for data retrieval is *k*, for which at least *q* service providers are required. a_i is the data pieces allocated to be stored at CSP_i, Thus, we have:

$$\sum_{i=1}^{p} a_i = N$$

$$\sum_{j=1}^{p} x_{i,j} \geq q$$

$$\sum_{j=1}^{p} . \sum_{i=1}^{a_j} x_{i,j} \geq k$$

Where, $N \geq k$ and $p \geq q$ $x_{i,j}$ is j^{th} data unit on i^{th} service provider.

Now, to assure no meaningful information retrieval by single CSP, less than *k* number of data pieces should be allotted to each CSP:

$$0 < a_i < k$$

Each of such data pieces should be converted to binary in order to apply DNA encryption and data hiding using DNA sequence.

IV. DNA ENCRYPTION MODEL

Data hiding is one of the most popular ways to protect data through the unsecured networks like Internet.

Sureshraj and Bhaskaran proposed new data hiding technique based on DNA sequences [6]. First DNA encryption is applied by using two rules viz. base pairing rules and complementary rules. Then generated cipher-text is embedded in DNA reference sequence, thus data hiding is used.

In real environment (biology) DNA is a pair of biopolymers, Polynucleotide, forming the double helix. Nucleotides synthesis is done in real natural environment using below constant rules:

4.1 Watson-Crick Base Pairing Rules

- Purine Adenine (A) always pairs with the pyrimidine Thymine (T).
- Pyrimidine Cytosine (C) always pairs with the purine Guanine (G). For that See Fig. 2.

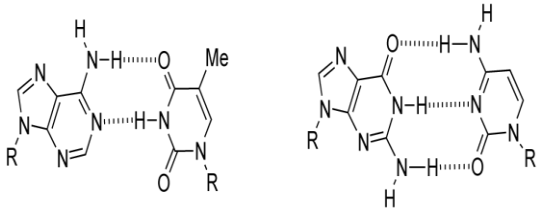

Fig. 2: Synthesizing Nucleotides in Natural Environment.

4.2 Application in Computing Area

Base pairing rule - These natural rules will be changed in order to increase complexity and make it difficult to intrude by attacker. For example, in biology G is synthesized to C while we can assume G to T, A or C, anything and so on, as we selected. Thus, all the base pairing rules are 4×3×2×1 = 24. The possibility of correct guess by attacker is 1 /24.

The letters are A, C, G, and T known as nucleotides, any composition from them will make a DNA sequence.

Binary coding rule or Complementary pairing rule - This rule is used to convert binary data to DNA sequences. Consider G = 00, A = 01, T = 10, and C = 11. In order to further increase the complexity, again this will change every time, So Next time G = 00 or it can be T, A or C. As these are four basic nucleotides, hence four possibilities of complementary rule for every DNA sequences. The final numbers of possible rules are 4×3×2×1 = 24. Hence, likelihood of correct guess is 1/24.

Generate DNA reference sequence - One way is to directly pick up reference sequence from European Bioinformatics Institute (EBI). This is online database that consists of around 163 million unique DNA sequences. However more secure method is generating DNA sequence based on the user-id and shift-key. Each user will be assigned with four character user-id e.g. "TAGC" and at random system will generate the shift-key range from one to sixteen.

Thus, system can generate 16^{16} combinations, as sixteen combinations can be shifted based on the key, i.e. generating more than 163 million unique DNA reference sequences [6], [7].

V. PROPOSED METHOD

There exist clients (client-1 and client-2) of a same company which are using cloud environment. The client-1 wants to upload critical data on cloud such that data confidentiality should be maintained. First, client-1 must apply the method of DNA encryption and data hiding on its data.
This is divided into two phases.

- Embedding data.
- Extracting the original data.

5.1 Embedding Data – Client-1

Pseudo code steps are as follows:

1. M is original data piece in binary.
2. Apply binary coding rule.
3. M' = DNA sequence (Converted to DNA nucleotides from binary).
4. Apply base pairing rule.
5. M" = new form of M'.
6. Find index of Nucleotides in DNA reference sequence.
7. M"' = Cipher text.

Assume original data M = 111001001011 should be uploaded to the cloud. The following steps shows original data convert to Cipher-Text.
DNA reference sequence is:

1. $AT_1CG_2AA_3TT_4CG_5CG_6CT_7GA_8GT_9AC_{10}CA_{11}AT_{12}TC_{13}GC_{14}GC_{15}TG_{16}AG_{17}TC_{18}AA_{19}CC_{20}$.
2. M = 111001001011.
3. Sub-Part$_1$ (A = 00, T = 01, C = 10, G = 11).
4. M' = GCTACG.
5. Sub-Part$_2$ ((AC), (CG), (GT), (TA)).
6. M" = TGACGT.
7. Sub-Part$_3$ (Picking Indexes); M"'=16109.

Thus, embedding phase is completed; sender sends 16109 to the cloud.

5.2 Extracting Data – Client-2

Pseudo code steps are as given below:

1. M"' = Cipher text.
2. Find Index of Nucleotides in DNA reference Sequence.
3. M" = Previous Form of M'.
4. Apply base pairing Rules on M' (In reverse way).
5. Get M' = DNA Sequence.
6. Convert M' to binary using binary coding rule.
7. Get M= original data

Client-2 takes the secret data in form of some numbers. Assume secret data M = 16109 should be downloaded from the cloud. Below steps shows cipher-text convert to Original data
DNA reference requence is:

1. $AT_1CG_2AA_3TT_4CG_5CG_6CT_7GA_8GT_9AC_{10}CA_{11}AT_{12}TC_{13}GC_{14}GC_{15}TG_{16}AG_{17}TC_{18}AA_{19}CC_{20}$.
2. M"' = 16109.
3. Sub-Part$_1$ (Picking Indexes); M"= TGACGT.
4. Sub-Part$_2$ ((AC) (CG) (GT) (TA)).
5. M' = GCTACG.
6. Sub-Part$_3$ (A = 00, T = 01, C = 10, G = 11).
7. M = 111001001011.

So, the receiver extracted the original data, accurately.

VI. SECURITY MEASURES

In terms of security, each intruder must have correct knowledge of following information. Without this basic knowledge, possibility of decrypting original data is scientifically near to zero.

DNA reference sequence: As stated previously system generates more than 163 million DNA reference sequences.

Base pairing rule: As stated in section IV-B, likelihood of correct guess of this rule is 1/24.

Binary coding rule: As stated in section IV-B, The possibility of correct guess by any attacker is 1 /24.

Hence, the final probability of correct and successful guess by attacker is [6]

$$\frac{1}{163 \times 10^6} \times \frac{1}{24} \times \frac{1}{24}$$

VII. SUMMARY OF LITERARURE REVIEW

Sr. No.	Authors	Work
1	M. Alzain et al.	Compares Multi cloud database model Vs Amazon cloud.
2	A. Kurmus et al.	Proposed Operating system based multi-tenancy architecture, OS resource segregation.
3	S. Lee et al.	Proposed Rain cloud system model, Rain cloud library-Interface, Demonstrated Data storage.
4	J. Prakash et al.	Suggests replication of data reduces Service

		availability risk or loss of data, Multiple unrelated cloud architecture should be used.
5	D. Sureshraj et al.	Proposed new data hiding technique using DNA sequence. DNA Encryption is used to protect data while storing on Multi-cloud.

VIII. 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 DNA encryption using data hiding, 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.

REFERENCES

[1] M. A. Alzain, B. Soh and E. Pardede, "MCDB: Using Multi-Clouds to Ensure Security in Cloud Computing", *IEEE transactions on Dependable, Autonomic and Secure Computing (DASC)*, pp. 784 – 791, Dec– 2011.

[2] A. Kurmus, M. Gupta, R. Pletka, C. Cachin, and R. Haas, "A Comparison of Secure Multi-tenancy Architectures for File System Storage Clouds", *ACM International Conference on Middleware*. pp. 471- 490, June - 2011.

[3] S. Lee, H. Park, and Y. Shin, "Cloud Computing Availability: Multi-clouds for Big Data Service", *6th International Springer Conference*, pp. 799 - 806, August – 2012.

[4] S. Prakash, Dr. K. Subramanyam, and S. Prasad, "Multi Clouds Model for Service Availability and Security", *IJCST Vol. 2, Issue -1*, March – 2012.

[5] Y. Singh, F. Kandah, and W. Zhang, "A Secured Cost-effective Multi-Cloud Storage in Cloud Computing", *IEEE Workshop on Computer Communications and Cloud Computing*, pp. 619 – 624, April – 2011.

[6] D. Sureshraj, and V. Bhaskaran, "Automatic DNA Sequence Generation for Secured Cost-effective Multi-Cloud Storage", *IEEE Conference on Mobile Application Modelling and Cloud Computing*, pp. 1 – 6, December – 2012.

[7] European Bioinformatics Institute, http://www.ebi.ac.uk/.