International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 3, Issue 2, March – April 2014 ISSN 2278-6856

DNA based Cryptography in Multi-Cloud: Security Strategy and Analysis

Richa H. Ranalkar[1], Prof. B.D. Phulpagar[2]

[1&2]Pune University, Department of Computer Engineering,
Modern College of Engineering, Shivajinagar, Pune – 05, India

Abstract: ***Cloud computing has a great capability to boost productivity and minimize costs, hence many companies are embracing it, but at the same time it constitutes many security 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 companies and new notion of Multi-Clouds usage is becoming perceptible to cope with these security issues. This paper demonstrates powerful security strategy of using DNA based cryptography which ensures secure data storage on Multi-Clouds.***

Keywords: Cloud Security, Cloud Service Provider (CSP), DNA Reference Sequence and Multi-Cloud.

1. INTRODUCTION

Cloud computing has a great capabilities, but at the same time it comprises many security challenges and risks. Due to possibilities of these risks such as service outage, theft of data, data leakage and the chances of malicious insider attack, using "single cloud" provider is becoming less favored. Materialized solution to this is use of multiple clouds i.e. "multi-clouds", "inter-clouds" or "cloud-of-clouds". In order to gain better data availability and security, user's critical data will be fragmented into parts and some interesting features of biological DNA sequences and data hiding principles will be applied to it. Finally, this DNA encrypted data pieces will be dispersed among the available Cloud Service Providers (CSP). Thus, this work proposes possible solution for security and privacy concerns of cloud.

2. LITERATURE REVIEW

Mohammad A. Alzain et al. [1] compares their own multi cloud database model with Amazon cloud. They concluded that data storage and retrieval can be done more efficiently using proposed model. Their analysis also addresses data intrusion, integrity and service availability issues.

Anil Kurmus et al. [2] 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. [3] 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.

According to S. Jaya Prakash et al. [4] service availability risk or loss of data can be reduced by using multiple CSPs with data replication technique. However, all CSPs should be totally unrelated with each other. This will overcome security risk of single contact point.

3. MULTI CLOUD

Most important problems that need to be addressed in cloud computing are data availability, data privacy, security and integrity [5], [6]. In order to achieve customer's privacy and security needs, CSPs are taking extra miles by implementing standard rules and regulations in line with powerful infrastructure. But still reports of data theft, data leakage, privacy breach, malicious insider attack and service outage are common in single cloud [7]. Solution to such trail of threats is storing critical data on multiple clouds from unrelated CSPs. Figure-1 shows captures example of Multi-Cloud architecture.

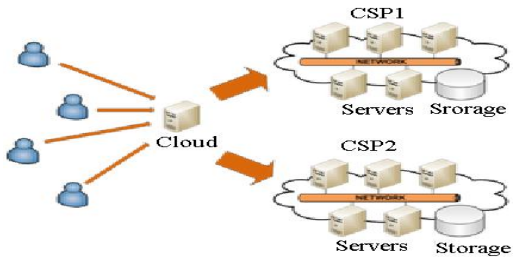

Figure-1: Multi-Cloud Architecture.

In order to achieve better security, user's critical data is broken into pieces and dispersed over multiple CSPs. Data piece allocated to server of CSPs should be small enough such that it should become impossible to retrieve any useful information from it. Consider p number of CSPs. Data will be divided into N parts. Let a_i is the data piece allocated for storage at CSP_i, alternatively $x_{i,j}$ is j^{th} data unit on i^{th} service provider.Then we have[8]:

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

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

To ensure no meaningful data retrieval, a_i should be far less than N

$$0 < a_i \ll N$$

After breaking down data into *N* pieces, next step is applying DNA based cryptography to each piece of data.

4. DNA BASED CRYPTOGRAPHY

One of the best known and popular techniques to secure data through the unprotected networks like Internet is Data hiding. Sureshraj and Bhaskaran proposed new data hiding technique based on DNA sequences [9]. Data will be DNA encrypted using two rules viz. complementary rule and binary coding rule. Then principle of data hiding will be used by embedding generated cipher-text in DNA reference sequence.

In real biological environment, DNA is double helix structure composed of pair of biopolymers, polynucleotide, known as Purine Adenine (A), Pyrimidine Cytosine (C), purine Guanine (G), and pyrimidine Thymine (T). Any composition of these four nucleotides will generate a DNA sequence. Below constant universal rules describes the basic synthesis of these nucleotides in real natural environment.

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 Figure-2.

4.2 Application in Computing Area

Base pairing rule: 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. Thus, possible number of such base pairing rules become 4×3×2×1 = 24. Hence chances of perfect guess by attacker are 1 /24.

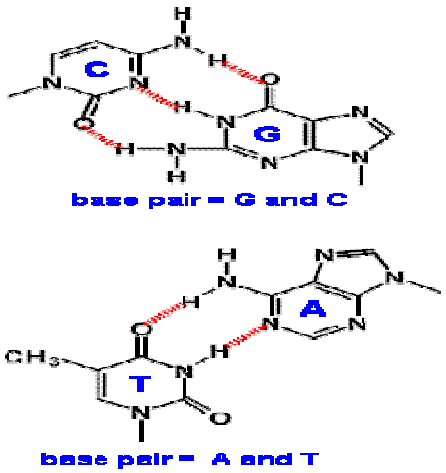

Figure-2: Natural Real World Nucleotides Synthesis.

Binary coding rule or Complementary pairing rule: Usage of this rule is for converting binary data to DNA sequence. Consider T = 00, A = 01, G = 10, and C = 11. To increase the complexity further, this will also change each time, so in next execution C = 00 or it can be G, A or T. Number of basic nucleotides are four, hence final numbers of possible rules are 4×3×2×1 = 24. Hence, chances of correct guess are 1/24.

Generate DNA reference sequence: One way is to directly select DNA reference sequence from European Bioinformatics Institute's (EBI) online database which consists of around 163 million unique DNA sequences. However in order to achieve more security, we will generate DNA sequence by randomly shift rotating four nucleotides. Thus at random system will generate the shift-key range from one to sixteen. Thus, system can generate 16^{16} combinations, as sixteen combinations of four nucleotides can be shifted, i.e. generating 185 million (more than EBI database i.e. more than163 million) unique DNA reference sequences [10], [11].

5. PROPOSED METHOD

Consider client within a company using cloud environment. The client needs to upload critical data on cloud by maintaining confidentiality. This strategy is divided into two phases as stated below.

- Data Embedding.
- Data Extracting.

5.1 Data Embedding

A Pseudo code steps are as follows:

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

Assume original data M = 110010011011 should be uploaded to the cloud. The following steps shows original data convert to Cipher-Text.

DNA reference sequence is:

1. $AA_1AT_2CC_3CG_4CT_5GA_6CA_7AC_8TT_9GT_{10}TC_{11}AG_{12}GG_{13}TA_{14}GC_{15}TG_{16}$
2. M = 110010011011.
3. Sub-Part$_1$ (T = 00, A = 01, G = 10, C = 11).
4. M' =CTGAGC
5. Sub-Part$_2$ ((AG), (CA), (GT), (TC)).
6. M''=ACTGTA
7. Sub-Part$_3$ (Picking Indexes); M'''=81614

Thus, embedding phase is completed; client sends 81614 to the cloud.

5.2 Data Extracting

Pseudo code steps are as per 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 in reverse way on M''.
5. Get M' = DNA Sequence.
6. Convert M' to binary using binary coding rule
7. Get M= original data

Assume secret data M = 81614 should be downloaded from the cloud. Below steps shows cipher-text conversion to Original data.
DNA reference sequence is:

1. $AA_1AT_2CC_3CG_4CT_5GA_6CA_7AC_8TT_9GT_{10}TC_{11}AG_{12}GG_{13}TA_{14}GC_{15}TG_{16}$
2. M'''= 81614
3. Sub-Part$_1$ (Picking Indexes from reference sequence); M''= ACTGTA
4. Sub- Part$_2$ ((AG), (CA), (GT), (TC)).
5. M' =CTGAGC.
6. Sub-Part$_3$ (T = 00, A = 01, G = 10, C = 11).
7. M = 110010011011.

Thus, data is extracted correctly.

6. SECURITY MEASURES

Suggested strategy is very secure. In order to penetrate and decrypt the original data, attacker must have below knowledge. 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 4.2, chances of correct guess of this rule is 1/24.

Binary coding rule: As stated in section 4.2, likelihood of correct guess by any intruder is 1 /24.

Hence, the final probability of correct and successful guess by attacker is [9].

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

7. IMPLEMENTATION PLATFORM

System is developed using Microsoft Visual Studio 2010, MS SQL 2008 tools on windows 8 operating system. Programming language used is ASP C#.net. Minimum hardware requirement is Processor Pentium –IV of Speed 1.1 GHz, RAM of 2 GB and Hard Disk of 80 GB.

Internet connectivity is required for implementing Multi-Cloud, as free storage services from different CSPs (like Mozy, Asus web storage, Cloud Me, Amazon cloud, Mega etc.) established in the market are used.

8. RESULT ANALYSIS

8.1 System is first developed using concept of single threading serial execution, and then same is extended to Multi-threading implementation. In Multi-threading implementation algorithm is programmed for parallel execution. Hence for *N*-CSPs total *N* threads runs in parallel as background process. Both implementations are analyzed for performance measurement in terms of data size and execution time. It is observed that execution time is reduced 10 times in Multi-threaded implementation. Figure-3 shows the performance measurement graph of both implementations. X-axis depicts data size and Y-axis depicts time of execution.

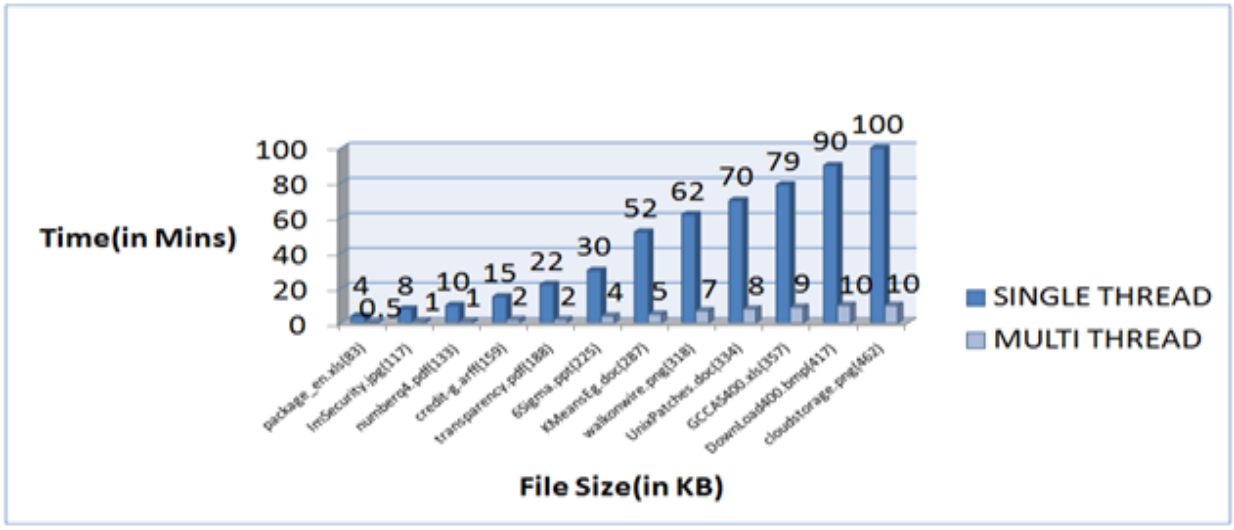

Figure-3: Performance Statistics: Single Vs Multi-Threaded Implementation.

8.2 In Multi-threaded implementation, initially data is divided in 4 pieces and stored on 4 CSPs, and then work is extended to 8 and 12 CSPs. Accordingly system performance is analyzed in terms of data size and execution time. Figure-4 shows the performance measurement graph of three implementations. X-axis depicts data size and Y-axis depicts time of execution. It is observed in final reading that execution time is reduced 4 times as the numbers of CSPs are increased to 12.

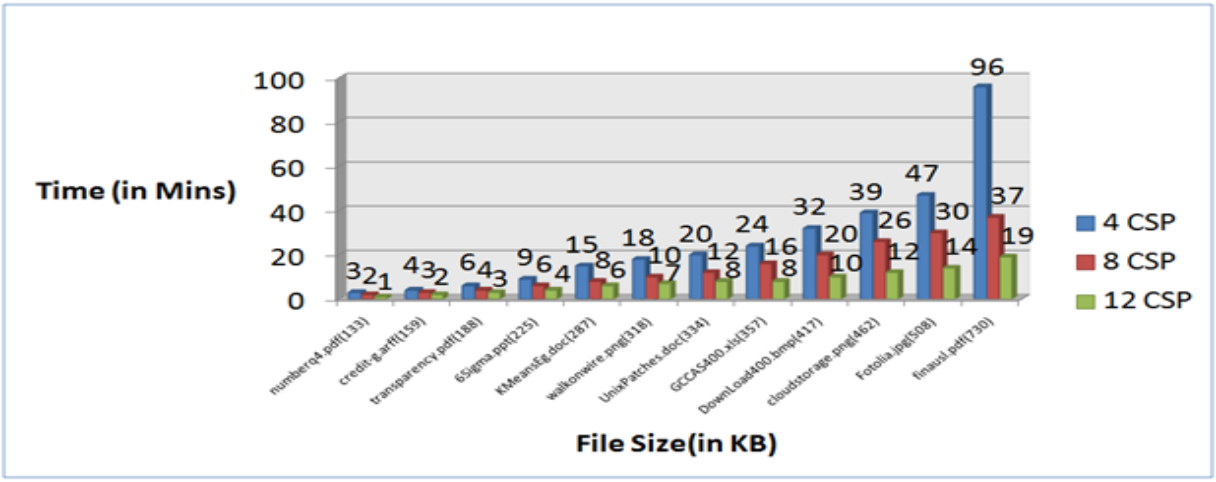

Figure-4: Performance Statistics 4 Vs 8 Vs 12 CSPs.

9. CONCLUSION

Cloud computing offers real benefits to companies seeking a competitive edge in today's economy, but obstruction to cloud implementation is security challenges and risks. By splitting user's critical data into parts and then applying powerful DNA based cryptography to each part of data and ultimately uploading it on multiple clouds; this strategy has shown its ability of providing a cloud user with a more secured storage. Future scope for this work is implementation of virus scanner and malware detector and embedding it within application. This will prevent uploading of any malicious data. However, the implemented concepts are definitely beneficial in building strong cloud security architecture. This will certainly meet with customer's expectations and will attract more investors for industrial as well as future research farms.

ACKNOWLEDGMENT

I am thankful to my project guide Prof. B. D. Phulpagar for his guidance and valuable comments. Without his support and co-operation this work would not have been possible.

REFERENCES

[1] M. Alzain, B. Soh and E. Pardede, "MCDB: Using Multi-Clouds to Ensure Security in Cloud Computing", IEEE conference on Dependable, Autonomic and Secure Computing, December– 2011, pp. 784 – 791.

[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, June – 2011, pp. 471- 490.

[3] S. Lee, H. Park, and Y. Shin, "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, 2012, pp 799-806.

[4] S. Prakash, K. Subramanyam, and S. Prasad (2013, February), Multi Clouds Model for Service Availability and Security, IJCSET, vol. 4, issue 2, pp. 158-161.

[5] N. Gruschka, M. Jensen, "Attack surfaces: A taxonomy for attacks on cloud services", IEEE 3rd International Conference on Cloud Computing, July 2010, pp. 276 – 279.

[6] W. Liu, "Research on Cloud Computing Security Problems and Strategy", IEEEE conference on Consumer Electronics, Communications and Networks, April-2012, pp. 1216 – 1219.

[7] F. Sabahi, "Cloud computing Security Threats and Responses", IEEE conference on Communication Software and Networks, May- 2011, pp. 245 – 249.

[8] 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, April – 2011, pp. 619 – 624.

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

[10] A. Leier, C. Richter, W. Banzhaf, H. Rauhe (2000, June), Cryptography with DNA binary strands, Elsevier BioSystems Vol. 57 , issue 1, pp. 13-22.

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