COMPREHENSIVE STUDY AND COMPARISON OF INFORMATION DISPERSAL TECHNIQUES FOR CLOUD COMPUTING

Ms. Priyanka
Department of Computer Science and Application
Chaudhary Devi Lal University
Sirsa, Haryana, India
priyankat137@gmail.com

Mr. Kapil Kumar Kaswan
Department of Computer Science and Application
Chaudhary Devi Lal University
Sirsa, Haryana, India
kapilkaswan@gmail.com

Abstract— Cloud systems refer to the collection of interconnected servers that are provisioned dynamically on demand, for execution of applications, to the customer like electricity grid. Cloud computing has gained great attention from industry but there are still many issues that are in their primitive stage which is hampering the growth of Cloud. One of the issues is security of data stored in the servers of datacenters of Cloud service providers. Many schemes have been developed till date for ensuring security of data in Distributed Systems. These schemes have been studied, analyzed and new method has been proposed which infix the parameters of security like recovery of data, confidentiality of data and integrity of data such that it ensure security of data stored in the servers of Cloud systems. The proposed scheme is based on two methods – Information Dispersal Algorithm and Fingerprinting. Information dispersal algorithm helps in maintaining confidentiality and integrity of data and Fingerprinting helps in recovery of data. Here Cloudsim simulation tool is used for simulation. CloudSim is an extensible simulation toolkit that enables modeling and simulation of Cloud computing systems and application provisioning environments.

Keywords: Information Dispersal Algorithm (IDA), Security issues

I. INTRODUCTION

Cloud Computing can be defined as "a type of parallel and distributed system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned, and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and the consumers". Some of the emerging Cloud-based application services include social networking, web hosting, content delivery, data processing and data storage. Cloud Computing is a distributed architecture that centralizes server resources on a scalable platform so as to provide on demand computing resources and services. Cloud service providers (CSP's) offer cloud platforms for their customers to use and create their web services, much like internet service providers offer costumers high speed broadband to access the internet. CSPs and ISPs (Internet Service Providers) both offer services.

II. ISSUES IN CLOUD COMPUTING

A. Privacy: There are various forms of privacy like "control of information about us". In Cloud Computing, data is processed and stored remotely so privacy is always an issue in the minds of customers. Since, Cloud services process users' data on machines that users do not own or operate, this introduces privacy issues

B. Security: While leading Cloud services providers employ data storage and transmission encryption, user authentication, and authorization (data access) practices, many people worry about the vulnerability of data to criminals like hackers, thieves, and disgruntled employees. Cloud providers are enormously sensitive to this issue and apply substantial resources to mitigating concern. Mostly, the uniqueness of the Cloud computing security is not recognized. Some researchers think that Cloud computing security is not much different from existing security practices and the security aspects can be well managed with the existing techniques such as digital signature, encryption, firewalls, and/or the isolation of virtual environments, etc .

C. Reliability: Some people also worry about whether a Cloud service provider is financially stable and whether their data storage system is trustworthy. Most Cloud providers attempt to mollify this concern by using redundant storage techniques, but it is still possible that a service can crash or go out of business, leaving users with limited or no access to their data. A diversification of providers can help alleviate this concern, albeit at a higher cost.

D. Data portability and conversion: Some people are concerned that if they wish to switch providers, they may have difficulty transferring data. Porting and converting data is highly dependent on the nature of the Cloud provider's data retrieval format, particular in cases where the format cannot be easily discovered. As service competition grows and open standards become established, the data portability issue will ease, and the conversion processes will become available supporting the more popular Cloud providers. Worst case will be when a Cloud subscriber will have to pay for some custom data conversion.

E. Intellectual property: A company invents something new and it uses Cloud services as part of the invention. Is the invention still patentable? Does the Cloud provider have any claim on the invention?

III. DATA STORAGE SECURITY SCHEMES

A.Shamir's algorithm*:* According to this scheme data is divided into n pieces and up to k pieces are required to get data. k-1 pieces will not reveal any information about data (secret). This scheme is based on polynomial interpolation: given k points (x^i, y^i) with distinct x such that for each x, there is one and only one polynomial q(x) of degree k-1 such that $q(x^i) = y^i$ for all i. Suppose data D is a number (ASCII value). To divide it into pieces D^i, a random polynomial $a^0 + a^1 x + \ldots\ldots\ldots a^{k-1} x^{k-1}$ of k-1 degree is selected in which $a^0 = D$.

$D_1 = q(1), \ldots\ldots\ldots D_i = q(i)$. If any of these k values are known, then other coefficients of polynomial are interpolated with the help of polynomial interpolation. On knowing the coefficients, data that is hidden is calculated with x = 0. Knowledge of just k-1 values does not reveal any data about secret data that is hidden.

B. Distributed fingerprints and secure information dispersal: This scheme uses the concept of Distributed fingerprints. Distributed fingerprints are used for calculating fingerprints of information in distributed systems. They are public fingerprints that are used for ensuring data integrity. They have the property that everyone in the system can compute them (using the same function and no secrets!) but no one can forge them i.e. no one can alter fingerprinted information without being noticed. These fingerprints require no secret keys and can be realized in a distributed environment with a majority of honest parties. They require the use of one-way hash function which is much simpler than general signature schemes like digital signature which usually use one-way hash functions as a sub component. They may replace some of the functions provided by signatures in distributed environments, but at a lower cost. IDA is helpful for ensuring security of data stored in servers of 'Distributed Systems' as long as active processors behave honestly; but, they require a security enhancement against dishonest processors i.e shares of file that are although present; but, are modified. The solution presented here is flexible enough to be adapted to most of the above scenarios even when malicious parties are expected. Distributed fingerprints require no secret keys at all, not for fingerprint generation neither for fingerprint verification. Everybody in the system can fingerprint any piece of information, but none can forge it.

C. A tree based recursive Information hiding scheme:In this scheme, additional information is added in the shares of the secret. This additional information is a message and the message is retrieved along with file (secret) on reconstructing the file (secret). Reconstruction of correct message shows the data is same as it was submitted which insures integrity of data (secret file). Message is known to user in advance. Work has been done in a finite field Zp, where p is a prime and it is public knowledge.

It has been assumed that a secret S is represented as string of numbers $S = s^1 s^2 \ldots s^r$, where each $s^i \in Zp$ and $|S| = r = n^h$, where |S| denotes the length of secret S for some integer h. For example, if we assume that the secret is a text message composed of ASCII characters, then it can be represented as a string of numbers less than p = 257 [efficient dispersal of information]. Furthermore, it has been assumed that there is another string denoted by $M = m^1 m^2 \ldots m^t$, $m^i \in Z^p$, where $|M| = t = n^h - 1/(n-1)$, that is to be hidden within the shares of the original secret S.

IV. INFORMATION DISPERSAL ALGORITHM AND FINGERPRINTING

In practice, an IDA is implemented as follows: The original file F is firstly divided into m segments S1, S2,.. Sm, each of size L=m. Then, the m segments are encoded into n unrecognizable pieces F1, F2, □ □ □ , Fn using a non-systematic m-of-n erasure code. The reliability of an IDA is clear : no more than n-m lost pieces of the n pices F1,F2….Fn will not results in data loss. However, the confidentiality of an IDA is not straightforward and deserves a sytamatic study.

Information Dispersal Algorithm (IDA) divides the file into shares such that no share reveals any information about file/secret and these shares are dispersed among servers and later when the file is to be retrieved back, then the shares are accumulated for reconstructing file out of them. Up to k shares are required out of n (total) number of shares to get complete secret file back.There are many algorithms that come under the category of Information Dispersal Algorithm. For implementing IDA, additional information is added within the shares of original secret. For implementation of IDA, additional information to be added is a message. The implementation of IDA in the proposed scheme helps in ensuring confidentiality and availability of data. Second scheme that has been used in the proposed algorithm is Fingerprinting. Fingerprinting helps in ensuring integrity of data.

V. SIMULATION

To enhance the security of data stored on Cloud servers, the proposed algorithm is to be implemented on Cloud platform. CloudSim simulation tool is considered best for this implementation; because, CloudSim supports system and behavior modeling of Cloud system components like virtual machines, datacenters, resource provisioning policies, servers, datacenter broker etc. CloudSim is an extensible simulation toolkit that enables modeling and simulation of Cloud computing systems and application provisioning environments. It implements generic application provisioning techniques that can be extended with ease and limited efforts. Currently, it supports modeling and simulation of Cloud computing environments consisting of both single and inter-networked Clouds (federation of Clouds). Moreover, it exposes custom interfaces for implementing policies and provisioning techniques for allocation of VMs under inter-networked Cloud computing scenarios . CloudSim simulation tool is like Java. In this tool, all entities are classes and the functions that these entities can perform are enlisted as methods. After extending an entity class, methods are called to perform the application.In CloudSim, other than entities' classes, there is core simulation framework. In this framework, different classes are available that work behind the scenes during execution of entities like SimEntity, CloudSimTags, SimEvent etc.

VI. COMPARISON OF THE PROPOSED ALGORITHM WITH FEW DISCUSSED ALGORITHMS

The comparison has been performed between the proposed algorithm and the existing algorithms in order to understand that whether the proposed algorithm is better suited for data security in Cloud. The comparison has been done based on the following parameters:

a) Confidentiality of data stored in servers

b) Recovery of data from modified pieces,

c) Integrity of data

d) Recovery of data from missing pieces

Parameters / Algorithms	Recovery of data from modified pieces	Recovery of data from missing pieces	Integrity of data	Confidentiality of data
Shamir's Algorithm [52]	✗	✔	✗	✔
Distributed Fingerprints and Secure Information Dispersal [56]	✔	✔	✔	✔
A Tree Based Recursive Information Hiding Scheme [57]	✗	✔	✔	✔
PROPOSED ALGORITHM	✔	✔	✔	✔

CONCLUSION

After studying various algorithms and understanding their advantages and disadvantages with respect to security requirements of data stored in servers of Cloud, an algorithm has been proposed which achieves the objectives of this research. The algorithm has been very well implemented in Cloud using CloudSim simulation tool.The proposed algorithm achieves the listed objectives of this research and can be very well implemented in Cloud. This algorithm is best suited for those Cloud service providers where main focus is to give secured data storage services and space is not a problem.

REFERENCES

[1] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, I. Brandic, "Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th Utility," Elsevier Science Future Generation Computer Systems, pp. 599-616, Jun. 2009,
[2] "Case Studies", http://aws.amazon.com/solutions/case-studies/.
[3] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, I. Brandic, "Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th Utility," Elsevier Science Future Generation Computer Systems, pp. 599-616, Jun. 2009,
[4] "Google App Engine," http://en.wikipedia.org/wiki/Google_App_Engine.
[5] "About the Nebula Platform," http://nebula.nasa.gov/about/.
[6] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. D. Rose, R. Buyya, "CloudSim: A toolkit for modeling and simulation of Cloud computing environments and evaluation of resource provisioning algorithms," Software: Practice and Experience, vol. 41, no. 1, pp. 23–50, 2011.
[7] Mark C. Chu-Carroll, "Cloud computing," http://scienceblogs.com/goodmath/2009/05/Cloud_computing.php, May 2009.