IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Efficient Data Backup Mechanism for Cloud Computing

Yashodha Sambrani[1], **Dr. Rajashekarappa**[2]

P.G Student, Dept of Information Science and Engg, SDM College of Engineering & Technology, Dharwad, India [1]

AP, Dept of Information Science and Engineering, SDM College of Engineering & Technology, Dharwad, India[2]

Abstract: Data generated in electronic format are voluminous in amount, as large generated data can be flexibly stored and accessed when needed by the users. Cloud computing has become the boon and emerging latest distributed computing technology which provides number of on demand services to the cloud users. This survey paper will be focusing more on the factors like security and disaster recovery aspects. The algorithm which will be proposed has two main objectives firstly providing highest security to the cloud users and secondly recovering of the data during natural destruction , paper also focus much on efficiency, time consumption to recover the data , data integrity, cost etc. Few of the recent techniques used for backup and security will be introduced giving a brief description on each of them.

Keywords: Remote server, Seed block, Data backup, Central Repository, Data Dynamics.

I. INTRODUCTION

Every users form home to professional workers would like to keep the backup of the critical data, even large organizations who are unable to store the heavily generated data in their own servers would like to give it to third party cloud service providers so that whenever the data is needed by them, or when the data is lost in their storage or may be due to natural disaster they can easily retrieve the data from the remote cloud. For the users to store the data on the cloud they need trust from the service providers that their data is unaltered and no other thirty party have access to the individuals data stored in the cloud. The data stored are not just limited for archiving as it involves data dynamics, data dynamics in cloud involves various operations like insertion, deletion, modification from time to time which must be reflected back to remote servers in the cloud within no time to achieve data integrity.

Cloud computing contains a network of servers that are running a low-cost consumer PC technology with the specialized connections that help in data processing across them [7]. This shared infrastructure contains pool of systems which are linked over a network and this concept of virtualization one of the key features of cloud computing will be used at maximum. As it involves sharing of computing of resources there are large number of users sharing the same storage space and other resources which the users are unaware of it, as the processing is made completely transparent to the users. Therefore, there is strong need of cloud security and privacy preservation techniques which does not allows our private data to be accessed by others either intentionally or unintentionally [12]. Even though accidently accessed and modified by the other users then it must be recoverable to its original state in a efficient manner. The data loss may also happen due to natural disasters which are unpredictable. Cloud storage will be as online storage where the information put away is as virtualized pool that is typically facilitated by third gatherings [8]. The facilitating organization works expansive information on huge server farm and as per the client prerequisites these data centre virtualized the assets and uncover them as the capacity pools that help clients to store their documents or critical information objects which can be accessed easily with low cost whenever needed. Data integrity plays an important role while recovering the lost data. To achieve all the requirements there is a need for highly efficient data recovery and backup technique.

II. LITERATURE REVIEW

1) HSDRT

This recovery involves the following features 1) It doesn't require utilization of costly rented lines when contrasted with customary frameworks and uses unused system resources(e.g. unused memory space of PC's , mobile phones and so forth). 2) It uses spatial scrambling and random dispatching technology to cipher important data files [2]. 3) As the number of users increases there will be more focus on security 4) As it involves stream cipher speed of encryption will be increased. The HS-DRT, that uses an effective ultra-widely distributed data transfer mechanism and a high-speed encryption technology. The system involves two sequences one is backup sequence and recovery sequence. The recovery sequence will be used when there is any data loss due to natural disaster when one of the components of HSDRT starts recovery.

There are some limitations involved in this approach due to which it cannot be considered as perfect technique for backup and recovery in cloud computing. Although this model can be used for movable clients such as laptops , smart phones etc. The data recovery cost is comparatively increased and also there is increased redundancy.

DOI 10.17148/IJARCCE.2016.5719

2) PCS
The proposed technique is based on parity cloud service, its performance is comparatively stable, simple and more convenient for data recovery, it covers the data with high probability .It does not require any user data to be uploaded to the cloud server for data recovery, it generates a virtual disk in user system for data backup, make parity groups across virtual disk, and store parity data of parity groups in cloud. It uses Exclusive-OR for getting information, the PCS [3] agent maintains a parity generation bitmap to indicate whether the parity block for each data block has been generated or not in the virtual disk. Each user will keep backup of their files to their own virtual disk for future data recovery process and whenever the user finds the original file is unavailable from the stored file system, the user will be requesting the PCS agent software for the recovery of the required file. The PCS agent software will be installed in each user system which creates virtual disk on the user storage device. The Virtual Disk Parity Group (VDPG), whose parity data are stored in the cloud storage, is generated and managed by the PCS server. PCS agent software consists of three key components Virtual Disk Interface (VDI), Recovery Manager (RM), and Storage Manager (SM). It addresses all issues like efficiency, reliability, convenience, especially it relieves users of their concern for privacy protection [14] as the user data are not stored on other servers. However this method has certain limitations as it is unable to handle implementation complexities.

3) ERGOT
Efficient Routing Grounded on Taxonomy (ERGOT) features the semantic analysis and fails to focus on time constraints and implementation complexity. It is a Semantic-based System which helps for Service Discovery in cloud computing. It is a unique technique for data retrieval [4]. It was observed that, this technique is not a back-up mechanism but it provides an efficient retrieval of data that is completely based on the semantic similarity between service descriptions and service requests [9],[10],[11]. ERGOT is built upon 3 components viz. 1) A DHT (Distributed Hash Table) protocol 2) A SON (Semantic Overlay Network) [15] 3)A measure of semantic similarity among service description. DHT based search is efficient but its limitations to retrieve exact terms for instance service name, hence it does not suit well for semantic similarity.

4) Linux Box
Linux Box model is having very simple concept of data back-up and recovery with very low cost [5]. Process of migration from one cloud service provider to other seems to be very easy. It is economical for all consumers and Small and Medium Business. This solution removes consumer's dependency on the internet service provider and its associated backup cost. It incorporates an application on Linux box that will perform backup of the cloud onto local drives. Care is taken to Keep data protected by using encryption techniques during transmission over the network. The described method is simple to implement and can be easily affordable by small and medium business . Even though it can be implemented with low cost, it requires higher bandwidth as it performs backup of the entire virtual machine .

5) Cold and Hot back-up strategy
Here the author proposes two approaches for recovery purpose for service composition which is one of the major concern in dynamic network [6],[15]. The two techniques provide improvement to the existing Backup Service Replacement Strategy(BSRS) particularly designed for managing failures in dynamic network. However, certain limitations in this strategy fails to provide good service. In the Cold Backup Strategy, the service triggers whenever there is sign of unavailability of services. The hot backup strategy provides greater service replacement when compared with Cold Replacement strategy. It is suitable when the chances of failure is high, as it restores the service before there is an interruption by keeping multiple service backups to maintain high availability.

III. CLOUD COMPUTING

The term 'Cloud Computing' has countless definitions and interpretations. Cloud computing allows end users to run software applications and access data from anywhere, anytime and from any computer. This is one of the most important elements of cloud computing and why it became so popular today which is surpassing all the previous technologies [13]. Cloud computing can be portrayed as a synonym for distributed computing over a network, where simultaneous computing from multiple users takes place. It refers to a cluster of hardware machines referred as a server connected through a communication network such as the internet, an intranet, a local area network or wide area network. Individual users who have permission to access the server can use the server's processing power for their individual computing needs like to run an application, store data etc. Therefore, instead of using a personal computer every-time individuals can access the data from anywhere in the world. Users need not have complete knowledge of cloud, expertise in, or control over the technology infrastructure in the "cloud" that supports them.
Cloud computing exhibits the following key characteristics:

Agility: It is the capability of rapidly and cost efficiently adapting to changes. Users can easily access more resources and then release them rapidly back to the pool. As virtual machines require more resources or when their hosts encounter difficulty, they may be moved automatically and instantly to other servers.

Virtualization: It allows to create a virtual version of a device or resource such as server, storage, network. It allows servers and storage devices to be shared and the rate of utilization is increased where applications can be easily migrated from one physical server to another. using virtualization technology creates multiple virtual machines

DOI 10.17148/IJARCCE.2016.5719

on a single physical machine can significantly reduce the hardware and power costs.

Cost: Using the cloud can reduce total cost of ownership of infrastructure significantly. Some clients report savings of fifty to seventy-five percent. Because each client is unique, the potential savings achieved by leveraging cloud technologies or services will vary.

Utilization and efficiency: It improves the utilization rate of systems that are often used only 10-20% and hence the resources can be used efficiently.

Reliability: It is improved if multiple redundant sites are used, which makes well-designed cloud computing suitable for business continuity and disaster recovery.

IV. REMOTE DATA BACKUP SERVER

A Remote Data Backup server [1] is the copy of the main cloud which is kept at a remote location far away from the main cloud and having the complete state of the main cloud, this remote location server is termed as Remote Data Backup Server. The main cloud is termed as central repository and remote backup cloud is termed as remote repository. The main purpose of keeping the entire data as whole at remote location is it help users to access the information from the remote cloud if the main cloud lost its data accidentally or due to unpredicted natural disaster. And also if the network connectivity is not available with the main cloud then they can collect information from remote cloud.
The Remote backup services must satisfy the following viewpoints:

1) Data security
Giving full assurance to the customer's information is the most extreme need for the remote server. Furthermore, either deliberately or inadvertently, it ought to be not ready to access by outsider or some other client's.

2) Data Integrity
Data integrity assumes a vital part in backup and recovery service, it is worried with complete state and the entire structure of the server. It confirms that information such that it stays unaffected amid transmission and gathering. It is the measure of the legitimacy and devotion of the information present in the server.

3) Data Confidentiality
Client’s data files should always be kept confidential such that even though number of users simultaneously access the cloud, the data files that are personal to only particular client must be able to hide from other clients on the cloud during accessing of file.

4) Trustworthiness
The remote cloud must have the Trustworthiness trademark. Since the client/customer stores their private information, therefore the users must be assured that their data will be completely secure and inaccessible to other users.

5) Cost efficiency
The recovery cost should be lesser. Lesser the cost of recovery, better will be the percentage of number of users.

V. DESIGN OF THE PROPOSED SYSTEM

The algorithm mainly aims on providing backup and recovery process. The Fig 1 shows the architecture of the proposed system consisting of main cloud , remote cloud and the clients. First for every client whomever registers with the main cloud, a random number and an unique client id will be generated for them, so that the client gets a unique identification with which they can further store or modify their data. Second, when the client id is being registered in the main cloud , the client id and random number will get EXORed with each other to generate seed block for the particular client which will be stored at the remote server .

When the client creates the file in cloud first time, it is stored at the main cloud. The main file of client is being EXORed with the Seed Block of the particular client when it is stored in main server. And that EXORed file is stored at the remote server in the form of file’ (pronounced as File dash). Due to any natural disaster if either file in main cloud is destroyed or file is being accidentally deleted , then the client will get the original file by EXORing file’ with the seed block of the corresponding client to produce the original file and return the resulted file which is the original file back to the requested client. Consider the algorithm given below.

Algorithm:

```
Initialization: Main Cloud: Mc; Remote Server Rs;
Clients of Main Cloud: Ci; Files: a1 and a’1;
Seed block: Si; Random Number: r;
Client’s ID: Client_Idi
Input: a1 created by Ci; r is generated at Mc;
Output: Recovered file a1 after deletion at Mc;
Given: Authenticated clients could allow uploading,
downloading and do modification on its own the
files only.
Step 1: Generate a random number.
Int r = rand ( );
Step 2: Create a Seed Block Si for each Ci and Store
Si at Rs.
Si = r XOR Client_Idi (Repeat step 2 for all clients)
Step 3: If Ci //Admin creates/modifies a a1 and stores at
        Mc, then
a’1 create as a’1 = a1 XOR Si
Step 4: Store a’ at Rs;
Step 5: If server crashes a1 deleted from Mc, then, we
        do
EXOR to retrieve the original a1 as; a1 = a’1 XOR Si
Step 6: Return a1 to Ci.
Step 7: END.
```

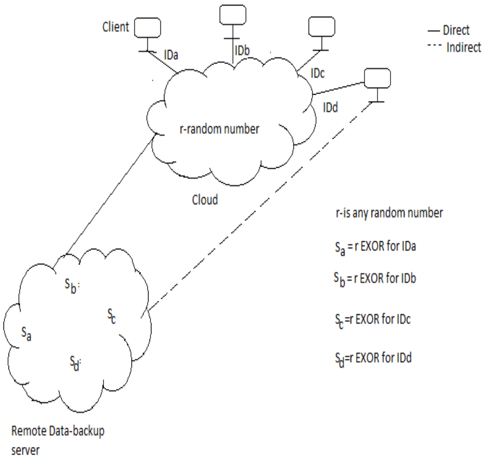

Fig 1. Architecture of Proposed System

VI. CONCLUSION

The minimal system requirement for main cloud server and remote cloud server is shown in the Table I. By observing the table we come to know that memory requirement is more in the remote cloud when compared to the main cloud because , the remote cloud need to store additional information such as seed block for every individual client. The experimentation results show the size of the original file which is stored at the main cloud is exactly similar to size of the file stored at remote cloud. To make sure it works the same with different file types and file with different sizes, the experimentation is carried out with different file types as shown in the Table II. The proposed SBA is very much robust in maintaining the size of recovery file same as that the original data file. From this we conclude that proposed SBA recover the data file without any data loss.

Requirements	Central Repository	Remote Repository
CPU	Core2 Quad Q6600 2.40GHz	Core2 Quad Q6600 2.40GHz
Memory	8GB(DDR2-800)	12GB(DD2-800)
OS	Any Window/Linux platform	Any Window/Linux platform
HDD	SATA 250GB or more(7200rpm)	SATA 500GB or more(7200rpm)

Table-I System Requirements

Type	Size Of Original File in main Server	Size Of Back-up File in Remote Server	Size Of Recovered File after recovery Process
Text (.txt/.doc/.docx/.xl/.pdf)	434KB	434KB	434KB
	2.5MB	2.5MB	2.5MB
Image (.jpeg/.gif/.png/.bitmap)	80KB	80KB	80KB
	4MB	4MB	4MB

Table-II Performance Analysis

Proposed SBA(Seed Block Algorithm) is simple to implement when compared with other techniques discussed in literature review. The Algorithm can still be enhanced with additional security algorithms for providing greater security.

REFERENCES

[1] Kruti Sharma, Prof. Kavita R Singh, 2013 “Seed Block Algorithm: A Remote Smart Data Back-up Technique for Cloud Computing” International Conference on Communication Systems and Network Technologies IEEE.

[2] Yoichiro Ueno, Noriharu Miyaho, Shuichi Suzuki,Muzai Gakuendai,Inzai-shi, Chiba,Kazuo Ichihara, 2010, “Performance Evaluation of a Disaster Recovery System and Practical Network SystemApplications,” Fifth International Conference on Systems and Networks Communications,pp 256-259.

[3] Chi-won Song, Sungmin Park, Dong-wook Kim, Sooyong Kang,2011, “Parity Cloud Service: A Privacy-Protected Personal Data Recovery Service,” International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11.

[4] Giuseppe Pirr´o, Paolo Trunfio , Domenico Talia, Paolo Missier and Carole Goble, 2010, “ERGOT: A Semantic-based System for Service Discovery in Distributed Infrastructures,” 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing.

[5] Vijaykumar Javaraiah Brocade Advanced Networks and Telecommunication Systems (ANTS), 2011, “Backup for Cloud and Disaster Recovery for Consumers and SMBs,” IEEE 5th International Conference, 2011.

[6] Lili Sun, Jianwei An, Yang Yang, Ming Zeng, 2011, “Recovery Strategies for Service Composition in Dynamic Network,” International Conference on Cloud and Service Computing.

[7] S. Zhang, X. Chen, and X. Huo, 2010, “Cloud Computing Research and Development Trend,” IEEE Second International Conference on Future Networks, pp. 93-97.

[8] M. Armbrust et al, “Above the clouds: A berkeley viewofcloudcomputing,”http://www.eecs.berkeley.edu/Pubs/TechRpts/2009//EEC S-2009-28.pdf.

[9] O. Sahin, C. Gerede, D. Agrawal, A. El Abbadi, O. Ibarra, J. Su. SPiDeR: P2P-Based Web Service Discovery. ICSOC, 2005.

[10] L. Vu, M. Hauswirth, K. Aberer. Towards P2P-based Semantic Web Service Discovery with QoS Support. BPS, 2005.

[11] F.B Kashani, C. Chen, C. Shahabi. WSPDS:Web Services Peer-to-Peer Discovery Service. ICOMP, 2004.

[12] Wayne A. Jansen, 2011, “Cloud Hooks: Security and Privacy Issues in Cloud Computing, 44th Hawaii International Conference on System Sciences Hawaii.

[13] Timothy wood, Emmanuel Cecchet, K.K.Ramakrishnan, Prashant Shenoy, Jacobus van der Merwe, and Arun Venkatramani, 2011, “Disaster as a Cloud service: Economic Benefits & Deployment Challenges”.

[14] Dark Reading, “Security is chief obstacle to cloud computing adoption, study says, ”http://www.darkreading.com, 2009.

[15] A. Crespo, H. Garcia-Molina. Semantic Overlay Networks for P2P Systems. AP2PC, 2004.