Internet of Things and Cloud Computing
2017; 5(5-1): 13-18
http://www.sciencepublishinggroup.com/j/iotcc
doi: 10.11648/j.iotcc.s.2017050501.13
ISSN: 2376-7715 (Print); ISSN: 2376-7731 (Online)

Efficient and Reliable Data Recovery Technique in Cloud Computing

Praveen S. Challagidad[1], Ambika S. Dalawai[1], Mahantesh N. Birje[2]

[1]Department of Computer Science and Engineering, Basaveshwar Engineering College, Bagalkot, India
[2]Center for Post Graduate Studies, Visvesvaraya Technological University (VTU), Belagavi, India

Email address:
praveensc07@gmail.com (P. S. Challagidad), asdambika@gmail.com (A. S. Dalawai), mnbirje@yahoo.com (M. N. Birje)

To cite this article:
Praveen S. Challagidad, Ambika S. Dalawai, Mahantesh N. Birje. Efficient and Reliable Data Recovery Technique in Cloud Computing. *Internet of Things and Cloud Computing*. Special Issue: Advances in Cloud and Internet of Things. Vol. 5, No. 5-1, 2017, pp. 13-18. doi: 10.11648/j.iotcc.s.2017050501.13

Received: July 4, 2017; **Accepted**: July 5, 2017; **Published**: August 25, 2017

Abstract: Cloud computing provides accessing of any kind of services dynamically over Internet on demand basis. One of the most significant service that is being provided is storage as a service. Cloud customer can store any amount of data into cloud storage results to huge amount of data at the datacenter. The data may get deleted by man-made disaster (either CSP or customer itself without their knowledge) or by natural disasters (either earth quakes or volcanoes) from the datacenters. Nowadays, data has been generated in large quantity that requires the data recovery services or techniques. Therefore there is a requirement for designing an efficient data recovery technique to recover the lost data. Many researchers have proposed different data recovery techniques but they lack in efficiency and reliability. In this paper, a multi-server system based on Enriched Genetic Algorithm to recover the lost data by using four cloud backup servers is discussed. To achieve reliability the proposed technique provides the flexibility for the user to collect information from any backup server when main cloud server loses its data and is unable to provide data to users.

Keywords: Cloud Computing, Data Recovery, Backup, Data Restore

1. Introduction

Now a days cloud computing is one of the beneficial technology. It overcomes the difficulties of additional technologies like cluster, grid and distributed computing. Cloud computing provides the thousands of server as a rent and executes the application on most powerful system available anywhere and anytime. It deals with data storage application, infrastructure using service oriented technology.

1.1. Cloud Service Models

Software as a Service (SaaS): SaaS is a collection of application and software; it allows the clients to subscribe the software instead of purchasing it. Software application is presented as service to the customer based on their demand. Twitter, Facebook, whats app provides Software as a service.

Platform as a Service (PaaS): This model provides platform as a service. This provides clients to develop his own application using the tools and programming languages. This service is hosted in cloud and accessed by clients using internet. Google App engine, Amazon AWS provides the platform as service.

Infrastructure as a Service (IaaS): This model provides the shared resource services. It provides the computing infrastructure like storage, virtual machine, network connection, bandwidth, IP address. IaaS is complete package for computing. Amazon, GoGrid provides the infrastructure as the service to the user [1].

1.2. Cloud Deployment Models

Public Cloud: A public cloud is available to any user with an internet facility, is less secure than the private cloud because it can be accesses by general public.

Private Cloud: Private cloud is available to a specific organization so that the user who belongs to that organization can have access the data. It is more secure than the public cloud because of its private nature.

Hybrid Cloud: The hybrid cloud is basically combination of no less than two clouds such as combination of private, community or public cloud.

Community Cloud: Community cloud allows the resources and system to be accessible by number of associated organization.

Data storage is one of the most significant services provided by cloud computing technology. But, recovering the lost data is one of the challenging issue in cloud computing paradigm. A brief overview of data recovery in cloud computing is discussed below.

1.3. Data Recovery in Cloud Computing

Data stored at the datacenter is increasing day by day it leads into huge amount of data storage in cloud and results into issues such as data loss, data breach etc. There is a need of an efficient technique if the data get destroyed or deleted by mistake to recover the data from any backup server. In business continuity if the system crashed or any type of natural or human made disaster occurred then there is chance of data loss and it may also cause the financial loss. By using some of the data recovery techniques the original data can be recovered. But, the existing recovery techniques are not efficient and reliable hence, to recover the lost original data a technique is needed to meet efficiency and reliability.

2. Related Work

This section presents summary of some of the data backup and recovery techniques in cloud computing.

In paper [2], author has proposed Hierarchical attribute based user classification algorithm to prevent the access of information from les privileged user. To do so author proposed a delegation approach. Further to provide the physical control of data to data owner author has divided the data into three categories such as Privacy Not Required (PNR), Privacy Required with Trusted Provider (PRTP) and Privacy Required with Non-Trusted Provider (PRNTP).

In paper [3], author has proposed the DR-Cloud model which is fault tolerant multi cloud storage, it makes use of DR XOR codes which provides data redundancy and uses minimum repair traffic during data transmission. DR-Cloud acts as interface between user application and multi cloud server.

The paper [4] presents the novel technique to recover the data. It solves all existing problems with data recovery by automatically compressing and decompressing the data before the backup of the data. Dual backup system was used. The dual system provides the high reliability and better bandwidth utilization of data storage.

In paper [5], author has proposed the Advanced Encryption Standard (AES) and Seed Block algorithm (SBA) method to perform the smart remote data backup in cloud computing environment. The proposed technique uses the AES and seed block algorithm. If the data gets deleted by mistake then we can get it from the remote server. This method takes less time to recovery the data and solves the time related issues. Thus the method provides an efficient security mechanism for the data stored in the cloud environment.

The paper [7] presents the cloud mirroring technique. It uses the mirroring algorithm. The method provides the high availability, integrity of the data, recovery of the data and minimizes the data loss. This method can be applied to any kind of the cloud. Cost to recover the data is also less.

In paper [8], author proposed the data backup and recovery technique. This technique provides the data protection from the service failure and also decreases the cost of solution. By using this technique the process of migration becomes simple and also removes the cloud vendor dependency. They proposed an effective data backup technique to recover the data from the server in case of data loss. For every business it is essential to back up the data to avoid the data loss.

The paper [9] presents a method which includes business service procedure (BSP) and disaster recovery procedure (DRP) with an assistance of cloud environment in order to avoid disaster recovery problems. The work employs priority based technique towards data recovery. The proposed approach ensures that it can provide security to entire organization datasets, which may contain log, account files. It also ensures that it can minimize the time required to get better organization data within small amount of time.

In paper [10], review has been done on distribution of data in cloud environment by construction of privacy preserving techniques and RBD. In order to perform smart RBD the system employs encryption and compression methods. The paper aims to preserve user privacy. System verified that it can overcome time related issues and are also solved by encryption and compress techniques.

In paper [11], the author implements the PRS algorithm for distributed disaster recovery system. The system On DDR provides the data security through 1+1+N distributed architecture in case of multi-node damage. To improve the system performance it uses the RS erasure coding and it also helps to reduce the storage resource consumption caused by data redundancy.

Paper [12] discussed the tools to study the disaster management. Now-a-days cloud computing technology is increasing day by day; the huge amount of data is stored on cloud. There is a chance of data loss and disaster. There is necessary to study the tools to manage the disaster in cloud environment. This paper aims to analyzing the various types of disaster and recovery techniques.

The paper [13] presents the Secure Erasure Coding (SEC) technique. This technique helps to retrieve the data from remote server in the absence of network connection and helps to recovery the data even if the data deleted from the server or cloud get destroyed. This method does not uses any kind of encryption techniques but also provides the security. It uses the less time to recovery the data.

From literature survey we found different techniques to recover the data. Each technique has its own advantages and disadvantages. During the data recovery process there present some issues. These issues are discussed below.

Issues Identified

Data Storage: All the enterprises stores there large amount of data in the cloud. For providing the security to data the computing is distributed but storage is centralized. Therefore single point failure and data loss are critical challenge to store the data in cloud.

Data security: User stores their huge data in the cloud. The stored data may be confidential and sensitive data. Providing the security to these data is important [6].

Lack of redundancy: If the cloud gets destroyed due to any reason then secondary site gets activated in order to provide the data to user when primary storage fails to provide the data.

Dependency: Customer doesn't have control on their system and data. Backup service is provided to overcome this drawback.

3. Proposed Data Recovery Technique in Cloud Computing

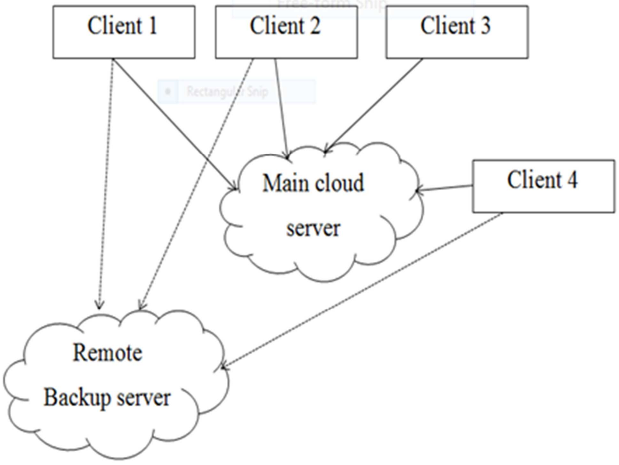

Figure 1. *Architecture of remote server.*

The proposed architecture imbibes three modules such as 1. Remote Backup Server 2. Main Cloud Server 3. number of Clients/users. Remote Backup server maintains the replicated copies of main server and is called as remote repository. The main server is called as central repository it stores all the user data. The user uploads the file to main cloud server; the main cloud server stores all the data in backup server. If user wants to retrieve the file from cloud then file is searched in main cloud server firstly, if the data is not present in main server then it is checks in backup server to retrieve lost data.

The loss of data or data crash happens due to natural disasters or human made disasters in main server. To recover the lost data a recovery technique is essential. Recovery of data can be achieved through the use of proposed algorithm efficiently. To provide the reliability two or four backup data cloud storages could be used. The Figure 2 shows the system architecture imbibing four backup cloud storages.

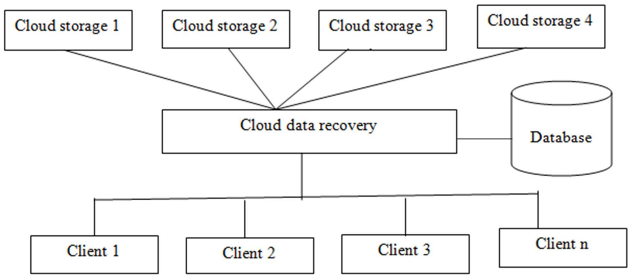

Figure 2. *System architecture.*

The system architecture illustrates four backup servers. The replicated copies of data are maintained in more than one server to recover data. When data loss occurs at one location it can be retrieved from other backup server using Enriched Genetic Algorithm (EGA). The EGA operation can briefly explained as imbibing following steps:

1. Initialization - Cloud Initialization.
2. Evaluation - Total file in cloud i.e. calculating size and cloud service provider status.
3. Selection - Selection of user.
4. Crossover - Comparing files for the user in cloud server.
5. Mutation - Deleted file is getting restored.
6. Termination - Completion of Restore.

3.1. Procedure of EGA

1. User uploads the file F to the N cloud servers.
2. From the file F generates hash code H1 and stored in data base.
3. Calculate the size of file.
4. User has to select the file to be downloaded.
5. If the file is deleted then it is retrieved from the backup server
6. Select the file to be downloaded and generate the hash code H2.
7. If both the hash codes are same then we retrieved the original file.

3.2. File Uploading Algorithm

1. From the client system, user uploads the file F to the cloud server.
2. Let N be the number of copies needed.
3. From file F generate hash code H1 and store it in database for the integrity check purpose.
4. I=1
5. Select I^{th} server and its cost per KB.
6. TOT Cost=Cost * Size of the file in KB.
7. New Balance=Available Balance – TOT_Cost
8. If New Balance < 0 then go to step 12.
9. Upload the file to server and update the balance.
10. 10 I=I+1
11. If I<N go to step.
12. Stop.

3.3. Recovering Algorithm

1. User has to select the file to be downloaded.

2. From the transaction table get the numbers of cloud storage containing the file and N server configure uration details.
3. I=1
4. Select Ith server status, if status is Deactivate then go to step 9.
5. Download file from Ith server
6. Generate hash code H2 from the file.
7. Fetch the hash code from Data base.
8. If H1=H2 the go to step 12.
9. I=I+1
10. IF I<=N go to step 12.
11. Display "File recovered Successful": STOP.
12. Display "File Integrity check is successful": STOP.
13. Display "File Integrity check is not successful": STOP.

3.4. Modules

Multi server system

Four different cloud servers are used for storing the data. Using all available cloud servers' reliability can be achieved in cloud computing paradigm. The data is stored in all four cloud server. The multi-server system increases the data availability in cloud environment.

Data hosting

Replication and erasure coding techniques are used to store the data. Data storage is based on size of file and frequency. Storage Mode Switching (SMS) will decide the replication process for storing the client data.

File Uploading with Hashing

In File uploading module, user has to select the file to upload to cloud by selecting the number of copies of replication required to store. While uploading charm application will read the file size in kb. Then it will select the best cloud storage server based on the storage availability, pricing cost, predictor, size etc. For integrity Verification process it will generate the Hash Key (HK1) using MD5 algorithm and it will keep it in the user DB. Finally based on the Replication details the File will be stored in the Cloud Storage Server.

File recovery

When user request the file from main cloud server he has to select the file from the Data Recovery Application then the data Recovery Server will select the corresponding cloud sever details from the DB and also it will check for the cloud availability for recovering the file, if cloud server is not available then it will be recovered from another cloud server. While recovering it will generate the HASH Key (HK2), then it will check for the HK1 &HK2 for the integrity check. Finally the file is recovered from the backup server.

4. Result Analysis

The proposed model is simulated using java language in windows 10 PC. The experiment is conducted by taking different types of files and its sizes as shown in table 1.

Table 1. *Different types of file and its size.*

Type	File Size	File size in remote server	File size of recovered file
.txt	250KB	250KB	250KB
.pdf	580KB	580KB	580KB
.jpg	30KB	30KB	30KB
.png	40KB	40KB	40KB

The Figure 3 shows the processing time taken by the main cloud server and the backup server.

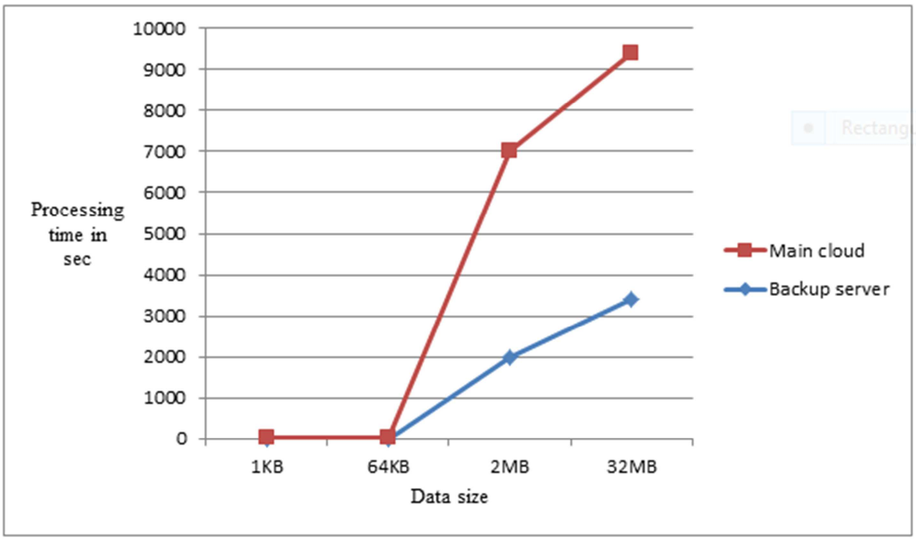

Figure 3. *Processing time of data in main cloud and backup server.*

The file uploading process is done either 2 blocks or 4 blocks. In 2 blocks the file is stored in two cloud servers. In four blocks the file is stored in four cloud server.

The Figure 4 shows the total time taken to upload the file in the two servers and the four servers.

The file uploading process it takes more time to upload the file in four cloud servers than to the two cloud servers. The uploaded file may be text, pdf or images.

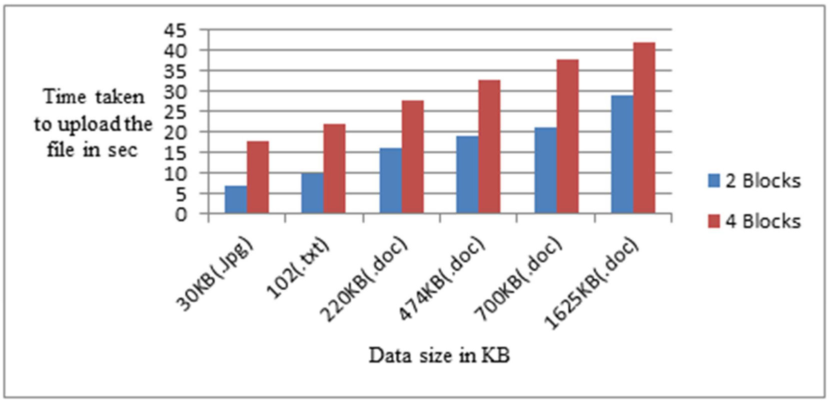

Figure 4. *Time taken to upload the data in sec.*

The user uploads the file to main cloud server if the file is deleted or not found in the main server then requested file is retrieved from the backup server.

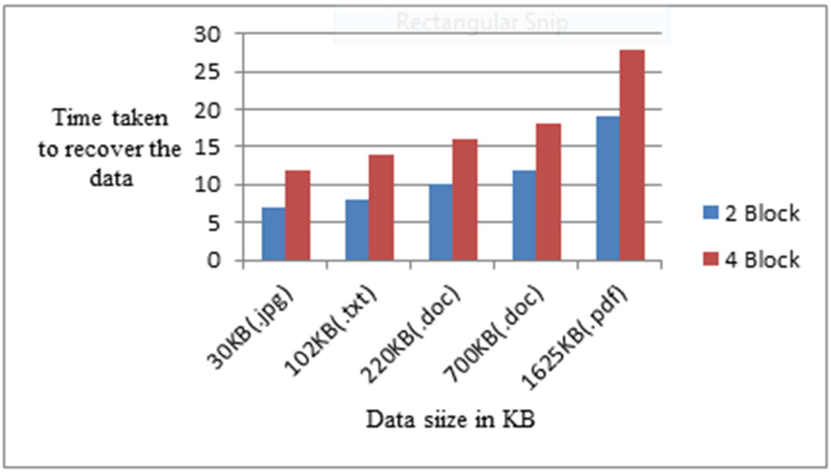

Figure 5. *Time taken to recover the data in sec.*

The Figure 5 shows time taken to recover the data in seconds. When the data is deleted from the main cloud server then the data is recovered from the backup server.

In two block method if the file is deleted or not found in the main cloud server then it is retrieved from the backup server.

In four block method: The graph shows when data is deleted or not found in the main cloud server and data is only present in one backup server then data is retrieved from another backup server.

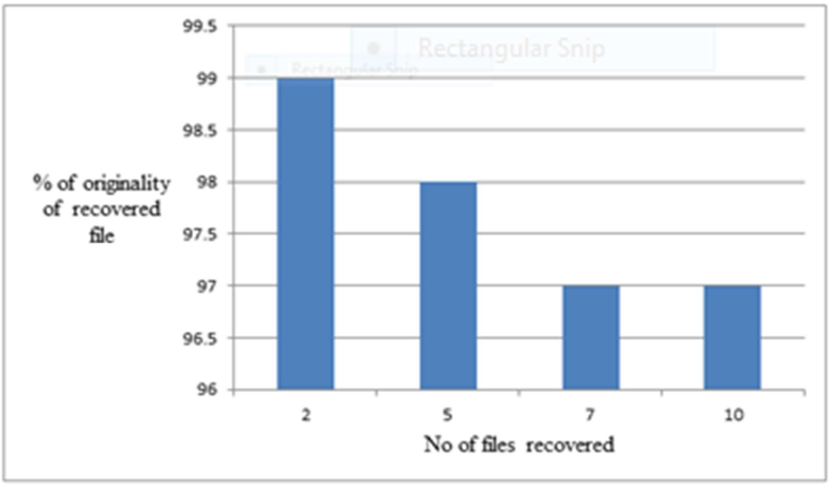

Figure 5. *Percentage of the originality of recovered file.*

The user uploads the file to main cloud server. If the file is deleted by mistake then that file is recovered from the backup server. The Figure 5 shows the % of originality of the recovered file.

5. Conclusion

Now a day's large amount of data is stored in the cloud and becoming very important to all the organization. A thorough literature survey is presented in this paper. The efficient recovery technique for recovering the deleted data is discussed in detail. Result and analysis section shows that the proposed algorithm is efficient and reliable. The four backup servers concept is used to recover the deleted data. Proposed method provides the flexibility for the user to recover their data from any server among four backup servers.

References

[1] Mahantesh N. Birje, Praveen S. Challagidad, "Cloud computing review: concepts, technology, challenges and security", *International Journal of Cloud Computing*, InderScience Publishers, vol. 6, issue 1, 2017.

[2] P. S. Challagidad, M. N. Birje, "Hierarchical Attribute-based Access Control with Delegation Approach in Cloud", Proceedings of the 11th INDIACom; INDIACom-2017; IEEE Conference ID: 40353 *2017 4th International Conference on "Computing for Sustainable Global Development*", 01st - 03rd March, 2017.

[3] Greeshma Radhakrishnan, Chenni Kumaran, "DR – Cloud: Multi- Cloud Based Disaster Recovery Service", *International Journal of Innovative Research in Science, Engineering and Technology*, Vol. 5, Issue 3, March 2016.

[4] Megha Rani Raigonda, Tahseen Fatima, "A Cloud Based Automatic Recovery and Backup System with Video Compression", *International Journal of Engineering and Computer Science,* ISSN: 2319-7242, Vol. 5, Issue 09, and September 2016.

[5] Tanay Kulkarni, Sumit Memane, "Intelligent Cloud Security Back-Up System", *International Journal of Technical Research and Applications,* Vol. 3, Issue 2, Mar-Apr 2015.

[6] M. N. Birje, P. S. Challagidad, M. T. Tapale, R. H. Goudar, "Security Issues and Countermeasures in Cloud Computing", *International Journal of Applied Engineering Research*, ISSN 0973- 4562, Vol. 10, No. 86, 2015.

[7] Shilpi U. Vishwakarma and Praveen D. Soni, "Cloud Mirroring: A Technique of Data Recovery", *International Journal of Current Engineering and Technology*, Vol. 5, No. 2, March 2015.

[8] PS. Vijayabaskaran, "Efficient Backing up Data for Migrating Cloud to Cloud", *International Journal of Computer Science and Information Technologies*, Vol. 6, 2015.

[9] Atesh Kumar, Saurabh Mishra, "Priority with Adoptive Data Migration in Case of Disaster using Cloud Computing use style", *International Conference on Communication, Information & Computing Technology,* 2015.

[10] Ruchira. H. Titare, Prof. Pravin Kulurkar, "Remote Data Back-up and Privacy Preserving Data Distribution in the Cloud: A Review*", International Journal of Computer Science and Mobile Applications,* Vol. 2, Issue. 11, November 2014.

[11] Jian Wan, Huijia Xuan, "Research and Implementation of Distributed Disaster Recovery System Based on PRS Algorithm", *International Journal of Database Theory and Application*, Vol. 7, No. 3, 2014.

[12] Chintureena Thingom, "A Study on Tools for Cloud Disaster Management", *International Journal of Interdisciplinary and Multidisciplinary Studies,* 2014.

[13] Kolipaka Kiran, Janapati Venkata Krishna, "Smart Data Back-up Technique for Cloud Computing using Secure Erasure Coding", *International Journal of Computer Trends and Technology*, vol. 16, number 3 – Oct 2014.