

File Recovery Technique using Mirroring concept in Cloud as Storage as a Service

Shilpi U. Vishwakarma

Computer Science and Engineering
P R Patil College of Engineering. & Technology
Amravati, Maharashtra, India
aditi12vish@gmail.com

Pravin D. Soni

Computer Science and Engineering
P R Patil College of Engineering. & Technology
Amravati, Maharashtra, India
pravindsoni@gmail.com

Abstract— Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network. The name comes from the use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud Computing entrusts remote services with a user's data, software and computation. It represents an emerging technology that permits the user to store big amount of data in cloud server and can use whenever required from anywhere. It provides various kinds of services to its users such as SaaS, IaaS, and PaaS. Storage-as-a-service is also a service delivered by cloud infrastructure in which big amount of data can be stored in cloud. As valuable and vital data of enterprises are stored at a remote location on cloud so we must be assured that user's data must be safe and be available at any time. In situations like hardware failure, data corruption or any accidental deletion our data may no longer remain available. To maintain the data security and we need a data backup technique for cloud platform to recover valuable and vital data efficiently. The purpose of recovery technique is to help user to retrieve information from any mirror server when main server lost his data and unable to provide data to the user.

Keywords: Cloud Computing, cloud mirroring, Recent Activity Table (RAT).

I. INTRODUCTION

Cloud computing is a pool of resources that are provided to the users through a network called INTERNET. It provides an ability to create, update and store user's files through any computer that has access to the web. We can say that Cloud Computing is nothing but the idea of accessing files, software and computing services via the Internet, instead from personal computer. Simply said, if user's software or files are "somewhere resides outside" instead of a computer's hard drive, we can using Cloud Computing services. The services provided by the cloud providers as SaaS, IaaS, PaaS, DRaaS, Storage as a service etc. Various companies like Amazon, Google, Microsoft, etc have their own cloud storage. [1]

Various Cloud Services are Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service(SaaS), Storage as a Service(STaaS), Security as a Service(SECaas), Data as a Service(DaaS), etc.

Mirroring

The general meaning of mirroring is the surface which reflects a clear image. To mirror a cloud, is to create a byte-for-byte copy of a cloud database at a different location. Actually mirroring is different than copying or backing up a cloud database in that, the mirror database is updated at the same time when the original database gets updated which is called as *synchronous* or as soon as possible after the original database is updated is called as *asynchronous*. There are three main purposes for mirroring [2]:-

1. To maintain another copy of a database for safe-keeping. The backup copy may be an on disk copy of an in-memory master database.
2. To offload reading of a database to another computer.
3. To be prepared to switch processing to another computer if the primary computer fails. This is often referred to as a Highly-Available (HA) database.

II. LITERATURE REVIEW

Cloud Computing provides sharing of computing resources which means huge number of users can share the same storage and other computing resources thus there is a strong need for a mechanism to prevent other users to access your important and useful data either intentionally or accidentally also. Several approaches have been taken by the various authors those are:-

Chi-won Song *et al*, 2011 have proposed a data recovery service framework for cloud infrastructure, The Parity Cloud Service (PCS) provides a "privacy-protected personal data recovery service". This technique work on the three steps those are Initialization Process, Parity Block Update & Data Block Recovery. In PCS user's data is not required to be uploaded on to the server for data recovery. [3] All the essential server-side resources that provide the recovery services are inside a reasonable bound. The advantages is, it provides a reliable data recovery at a low cost & the disadvantage is its implementation complexity is higher.

Yoichiro Ueno *et al*, 2010 have proposed a technique called HSDRT that uses an "ultra-widely distributed data transfer mechanism and a high-speed encryption technology". It has three component Data Center, Supervisory server, Client node specified by admin (PCs, Smart phones etc.). It has two sequences one is *backup sequence* that stores the data which is backed up and other is *Recovery sequence*, is used when there is disaster or any data loss occurs. The Supervisory Server starts the recovery sequence. [4] Even though this model can be used for movable clients such as laptops Smart phones etc. the data recovery cost is comparatively increased and also it increases the duplicity.

Linux box technique is proposed by Vijaykumar Javaraiah, 2011 that decreases the cost of the solution and protects data from disaster. Here migration process is very

easy i.e. transfer data from one cloud service provider to other. This technique is affordable to all consumers, Small and Medium Business (SMB), eliminates consumer's dependency on the ISP (Internet Service Provider) and its associated backup cost and integrates an application on Linux box that will accomplish backup of the cloud on local systems. The application will interface with cloud on a secured channel, check for updates and sync them with local storage. The data transmission will be secure and encrypted. After a valid login, the application secures the channel using IP Security and in-flight encryption techniques. The application then interacts with the application stack at the cloud service provider and does a onetime full backup. During subsequent check, it backs up only the incremental data to the local site. The limitation is that a consumer can backup not only the Data but Synchronize the whole Virtual Machine [5] which somehow consume more bandwidth.

Giuseppe Pirr'ò *et al*, 2010 proposed ERGOT technique which is fully based on the semantic analysis which does not focus on time and implementation complexity. It provides the support for Service Discovery in cloud computing. [6] Building components of ERGOT is, Distributed Hash Table protocol, Semantic Overlay Network and a measure of semantic similarity among service description. This technique provides the retrieval of data in an efficient way that is totally based on the semantic similarity between service descriptions and service requests. ERGOT is based on semantic-driven query answering approach in DHT-based systems by building a SON over a DHT but it does not go well with semantic similarity search models. The downside of this model is an increased time complexity.

Lili Sun *et al*, 2011 proposed a technique which is a gradual increase in cost with the increase in data i.e. The Cold and Hot back-up strategy that attains backup and recovery process and triggered on the basis of failure detection. In CBSRS (Cold Backup Service Replacement Strategy) recovery process, it is triggered when a service failure is detected and it will not be triggered when there is no failure i.e. when the service is available. The HBSRS (Hot Backup Service Replacement Strategy), is a "transcendental recovery strategy for service composition", used for dynamic network. [7] During the implementation of process, the backup services remains in the activated state and the first returned results of services will be used to ensure the successful implementation of service composition.

Eleni Palkopoulouy *et al*, 2011 proposed this technique that mainly focuses on the significant reduction of cost and router failure scenario i.e. (SBBR). It involves logical connectivity of IP that will be remain unaffected even after a router letdown. Importantly it provides the network management system via multilayer signaling. [8] Additionally this model shows how service imposed more outage requirements that directly effect on the setting of the SBBR architecture (e.g. imposing a less number of network-wide shared router resources locations).The problem with model is that it is unable to include optimization concept with cost reduction.

B. Cully *et al*, 2008 proposed REN technique that provides the lowest cost point of view a model "Rent out the Rented Resources". REN focuses on reducing the cloud service's monetary cost. It proposed a model for cross cloud federation which consists of three phases that are Discovery, Matchmaking and Authentication. This model is completely based on the conception of cloud vendors where the cloud customers rent the resources from different venture(s) and after virtualization, rents it to the clients as cloud services. [9]

Ms. Kruti Sharma *et al*, 2013 proposed a technique SBA in which there is a remote Backup server which is a copy of original cloud server, physically placed at a remote location. This method is based on the concept of Exclusive-OR (XOR) operation of digital computing. The whole mechanism consists of three main parts as the Main Cloud Server, Clients of the Cloud and the Remote Server. This algorithm uses a random number and a unique client id associated with each client. Every time when a new Client is get registered with the cloud its unique client id is get XOR with a random number. The result of this XOR operation is called as a Seed Block which will be used only for that specific client. When a client stores their Data on the Cloud it is saved in Cloud and simultaneously it is XORed with its Seed Block and the resultant Data is stored in the remote server. If any accidental data loss occurs in the main Cloud then in such cases the original data is recovered by XORing the Data with the Seed Block of that particular client to obtain Data i.e. the original Data file. [10] This technique is fully capable of recovering the data files. The downside is, inefficient because there is wastage of storage space.

III. PROPOSED METHODOLOGY

Now for big companies if ever data is loss then they can manage their data and resources but for small companies that will be the bigger issue to manage because data loss is a real risk that's why we planned to make use of storage as a service and developed an "exact copy of Main Database which we called as MIRROR Database". So in case if any of the company or user lost their data from the main database then we can retrieve their data from mirror database. In proposed technique user can upload any kind of data like files, images, audio, video etc. on cloud whether it is in plain text or in encoded format.

Mirroring algorithm:-

Mirroring scheduling algorithm will check the mirror copy of the user data. Mirroring starts when the CPU utilization goes below the threshold value (we assume the CPU threshold value is 60%), and also we will do the mirroring according to time (we assume time threshold is midnight (2 a.m.)). By using the concept of CSP (Cloud Service Provider) we will maintain the log through which we will continuously(say after 5 minutes) check the row mirror counter, after analyzing the log, CSP can dynamically change the threshold values. The advantages of using this technique is though we are using hard disk for file recovery, eventually the cost for recovery will be reduced along with this the proposed technique is applicable to any kind of cloud. This technique will focus on entire mirroring of cloud as we are using the asynchronous mirroring the overhead of the RAT.

IV. SYSTEM IMPLEMENTATION



Figure 1: Login Module

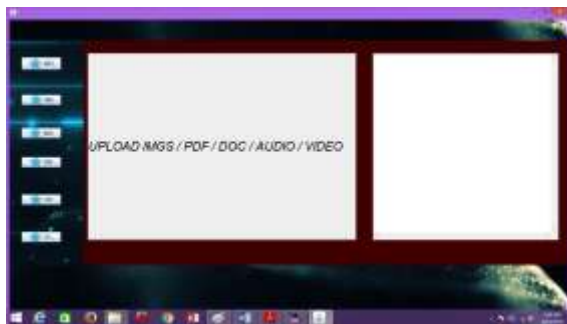


Figure 2: choose to upload a file

Select desired file to be uploaded on the server for that various options will be there.



Figure 3: Uploading Module

In Uploading module user has a facility to upload any type of file such as audio, video, images etc. once user upload his file we will encode that file by using Base64 encoding technique, to provide security to the users data.



Figure 4: Downloading Module

Now at the time of downloading the user has two options to download the file that is “with encoded format” and second

one is “Plain format”. So if user selected the file to be in encoded form then the desired file will not be open and if user choose file to be download as a plain form then the file will be open on his/her local disk, if file is audio then another option is provided, play song or not, if user choose yes then song will be play.

V. RESULT ANALYSIS

If files stored in cloud can be corrupted then it is impossible to recover the original file and unfortunately the respective data will be lost, but we are providing a technique called “Cloud Mirroring” that will ensure that whatever data stored by the user on the cloud the data(files) can be retrieve in any kind of failure also.

Uploading and Downloading: To upload or download a file on the server it will completely depend on the size of that particular file along with speed of internet. Uploading large files, or even many files, can take quite some time if user using a basic cable connection then upload speed will be a shadow of download speed so even uploading a few gigabytes of data could take a few hours, or even days if you are uploading tens or hundreds of gigabytes.

Mirroring: To mirror a data we provide a timer which will continuously mirroring (say after 5 min) the data whichever present on the RAT table so mirroring process will depend on the Recent Activity Table because after every 5 min we will check whether the RAT is full or not and accordingly we will start mirroring process, this process will not take much time it will take few seconds if the file size is less but for bigger file it will take quite some time.

Recovery: Here in this technique, suppose a file takes downloading time is say 2min, but if the desired file is lost from main database then we have to give that file to the user from mirror server to it will take 2.5 min. it will take few more seconds rather than original server because all the server present on the remote location.

VI. CONCLUSION

The data stored by the user is always valuable for him but no one can assure whether his data cannot be corrupted or lost so recovery plays a vital role in such scenarios. We have proposed file recovery technique by the concept of cloud mirroring by which we provide the user’s desired data.

REFERENCES

- [1] Shilpi Vishwakarma, Pravin D. Soni *A Survey on various backup and recovery technique in cloud*, International Journal on Recent and Innovation Trends in Computing and Communication Volume: 3 Issue: 12 (December 2015)
- [2] Shilpi Vishwakarma, Praveen D. Soni *Cloud Mirroring: A Technique of Data Recovery*, International Journal of Current Engineering and Technology, Vol.5, No.2 (April 2015).
- [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] Yoichiro Ueno, Noriharu Miyaho, Shuichi Suzuki, Muzai Gakuendai, Inzai-shi, Chiba, Kazuo Ichihara, 2010, *Performance Evaluation of a Disaster Recovery System and Practical*

-
- Network System Applications*, Fifth International Conference on Systems and Networks Communications, pp 256-259.
- [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. M. Young, the Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
- [6] 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
- [7] Lili Sun, Jianwei An, Yang Yang, Ming Zeng, 2011, *Recovery Strategies for Service Composition in Dynamic Network*, International Conference on Cloud and Service Computing.
- [8] Eleni Palkopoulou, Dominic A. Schupke, Thomas Bauscherty, 2011, *Recovery Time Analysis for the Shared Backup Router Resources (SBRR) Architecture*, IEEE ICC.
- [9] B. Cully, G. Lefebvre, D. Meyer, M. Feeley, N. Hutchinson, and A. Warfield, *rmus: high availability via asynchronous virtual machine replication*, in Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation, 2008, pp. 161–174.
- [10] Ms. 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.
- [11] Somesh P. Badhel, *A Review on Data Back-up Techniques for Cloud Computing*, International Journal of Computer Science and Mobile Computing, Vol.3 Issue.12, December- 2014, IJCSMC
- [12] S Sankareswari, S Hemanth *Attribute Based Encryption with Privacy Preserving using Asymmetric Key in Cloud Computing (IJCSIT)* International Journal of Computer Science and Information Technologies, may 2014