

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 8, August 2016, pg.216 – 222

A Novel Approach for Efficient Data Slicing and Replication in Cloud

Shilpa Kodli¹, Priya Maski²

¹Assistant Prof, Department of MCA, Visvesvaraya Technological University, India

²VTU Student Department of MCA, Visvesvaraya Technological University, India

¹shilpakodli@gmail.com, ²priya13119119@gmail.com

Abstract— Cloud computing is one of the most efficient storage space environment for information on cloud. Cloudlet allows users to store the data distantly, where cloudlet is a local cloud which may consist of several broker and virtual machine. A broker may have any number of virtual machine connected to single broker. Many cloudlet are connected to actual cloud server. Conventionally, there were more problems regarding data storage, updating and retrieval from cloud server. We introduce the cloudlet mechanism to handle the data stored on local server. Each virtual machine stores only a single slice of information so that is secured and performance is high. Whenever we store a large file into cloud it takes more time for both upload and download. This is because large size of files takes a lot of time to transmit. Therefore server bandwidth is occupied for long period. To overcome the loop holes, here we propose a method called partitioning and reprint to specific virtual machine. It provides security by not allowing to recover the whole information. It offers fast and reliable service to users.

Keywords— cloud computing, file data, virtual machine and broker system, sky drive.

I. INTRODUCTION

Cloud computing is helpful in providing service, security along with scalable synchronization with near to the ground cost. Cloud handles enormous quantity of data very efficiently. Normally a file is being used with a server client model, clients generally upload the files in a server, server host the file when ever client wants to retrieve the file. Client has to give file address or url to the server and server returns the file as a binary file format. This has being normal file access technology which is being used over the years, however as the amount of data has being increased large volume of document it has become extremely difficult for this system to manage such kinds of files. This is due to two specific reasons- i) Server will be busy ii) As the number of clients increase the server effective bandwidth reduces which increase the latency of the file access. So a distributed file system has been proposed in past. Infrastructure as a service model uses virtualization technique to send data over cloud. In a cloud cluster there will be a several nodes, each one is known as a virtual machine. Each node will be having certain

processing capability, data accessing capacity. The part of any of the node which is available for any of the processing is known as a virtual machine. Cluster can be a determined as intercommunication of several virtual machine. The advantage is once a job arrives in cluster, cloud will be divided into several virtual machines for quick execution of work. When the task is completed it combines and send it back to user. This system is recognized as a broker or service broker. So a cluster is nothing but a combination of one or more brokers connected with many add up to of virtual machine.

II. RELATED WORK

As data stored on cloud should be accessible remotely to all users from any areas. Before storing data on cloud the concept used here are data imitation , fragmentation and T-Coloring to overcome security issues. The authors in [1] presented a methodology to ensure huge data storage service, confidentiality better security to system compared to previous system. Cloud helps data owner from unwanted attackers by protecting the information. The time period required for the uploading and downloading file over cloudlet is less, where users can be prevented from data loss. Instead of using measuring distance between two nodes it uses cloud service providers, which intern handles the request with multiple sub server. It provides slice of data to the users after automatic updating of nodes within local cloud. S.Sangeetha *et. al*. [2] proposed a technique Heuristic auditing strategic where user maintains a operational table to read record of node placement. Selection of particular cloud for storing the data with less cost, When users tries to update , upload and download the content of file it first checks the chunks with multiple of cloudlets so that the same copy is not repeated over the cloud. Gade Pooja V *et. al* . [3] propped data hosting scheme in hybrid multi cloud which reduces the time taken for data breakup and stores efficiently. This explains about the handling of data in cloud through automatic update. Data which is available on cloud is also accessible through request by client. If data is easily available within cloudlet means that users can easily access data stored in different nodes present within cloudlet with multiple copies of same data, by replication process which enhances recital level, scalability and backup information to handle large data over cloud. The authors determines in [4] reprinting process data can be easily recovered from different other nodes in case if any one of node is not available to users so as to increase the performance of system. Unscheduled task are sent back to the cloud which increases the efficiency of system this is known as reservation cluster. It uses pseudocode algorithm and solves the problem of resource usage and computation time of cloud. This is presented in [5] which is useful for finding the resource usage and execution time of cloud data. If all the task are scheduled then needs sent to reservation cluster will be executed.

Cloud computing is a shared service for interconnected servers, when compared to earlier storage technique used for dumping data for later use in hard drives this changes way of information store and running application saves cost. The breaking up of file given by user depends on certain criteria that no meaningful information is dumped in a single slice and each node contain a only single slice of information within cloudlet so as to increase security issues. M.Muthu *et. al*. [6] utilize fragment and shuffle algorithm is used and data forward algorithm to send reprint to server which generate key to all specific slice stored by using this key user can retrieve data from cloudlet. Infrastructure as a service is used in cloud computing, here virtualization technique is used to send data to cloud. If stored data is unavailable to user then this can be solved arranging resources based on provisioning of client. The problem occurs with over provisioning of stored resources are not utilized. P.Anand *et. al*. [7] analyze main theme is to minimize overall cost of provisioning data shared over network. To overcome with such issues an algorithm is used called optimal cloud resource provisioning algorithm it reduces cost with time.

III. EXISTING SYSTEM

A distributed file system is a file system where a particular directory is being distributed over several systems in a network. whenever a file needs to be stored, it is saved in one of the systems which is inside the directory. A file gets stored completely in a single system. The distributed file system by default also allows backup and recovery. Therefore when a file is uploaded in such a system a facsimile of the file is stored in one cloud server and some other copy that backup copy is automatically stored in the other interconnected server that is how dispersed file system works. So what distributed file system does is it increases the overall bandwidth by incorporating several number of servers and distributing the

overall file storing space among many number of servers. The data compromise may occur due to attacks by other users and nodes within the cloud. The employed security strategy must also take into account the optimization of the data retrieval time.

IV. PROPOSED SYSTEM

In order to overcome drawback of the present system that large files are stored in a single chunk into one of the node of the distributed system file system, we propose a novel approach by means of which large files are divided into smaller chunks and each chunk is stored in different machine. Also a copy of its backup is made which is stored in different machines of the system. In the proposed system we also create a snapshot of the network is defined by over all connectivity of the nodes. Therefore a system in proposed system will always have a copy of network and all the routes available from one node to the other node. The proposed system also incorporate a technique called routing technique by means of which distributing files to server calculates the probability of getting the file early. File chunks are distributed to the system which are well connected to each other such that file retrieval is faster.

Advantages are as follows

The main advantage of the proposed system is large files are distributed over many number of system, therefore retrieving file is easy.

Reliability for retrieving files because the multiple copies of the same file is being stored in multiple location.

Even if one system goes down, the file can be easily retrieved.

File integrity which means that when a file is being stored or modified, it is also modified in all the backup location. User does not have to manually change the backup file. As the main file gets changes, backup file also change.

Latency-the proposed system offers very quick file access, therefore from any place the file would be accessed if had an internet connectivity.

Limitation of Proposed System

The major limitation of proposed system is that files are not being encrypted before storing.

V. METHODOLOGY

The two algorithm used here are Algorithm for fragment placement and another Algorithm for fragment's replica. The file is fragmented based on the size of data and with some inputs and initialize and than the file is which is being fragmented is replicated to multiple nodes. While placing a copy of data in cloud there consist a ineligible nodes for storing any fragment by, possible option for replication of fragments. T-coloring is a method which provides information about the node that which node is connected with which node into the virtual machine. In cloud, file is divided into slices and a copy is stored to nodes as well as replica is saved in sky drive to enhance security to data by not allowing information to leak in case of attack.

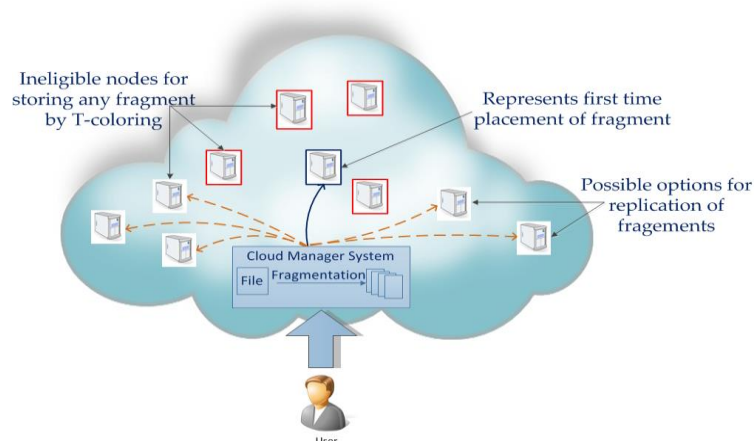


Figure1- Data Slicing and Replication Methodology

Virtual Machine- Users outsource the data to cloud, the data is then fragmented (splitting of data) and a replica (a copy) of data is sent to different nodes of virtual machine. No virtual machine gets more than a single slice of data, Virtual machine is also known as a guest which is used for creating another environment within the host environment. Virtual machines are used to perform a specific task that is different from the host environment. Virtual machines are divided into two types 1) system virtual machine and 2) process virtual machine.

System virtual machine that supports the sharing of host environment's physical resources between virtual machines.

Process virtual machine provides platform independent programming environment.

Virtual machine gives the file to the service Broker, where broker will fragment files and send the replica to the multiple nodes. As well as a copy is stored in backup. Broker handles the data, then send to the server for storing data through the multiple chunks in cloud network.

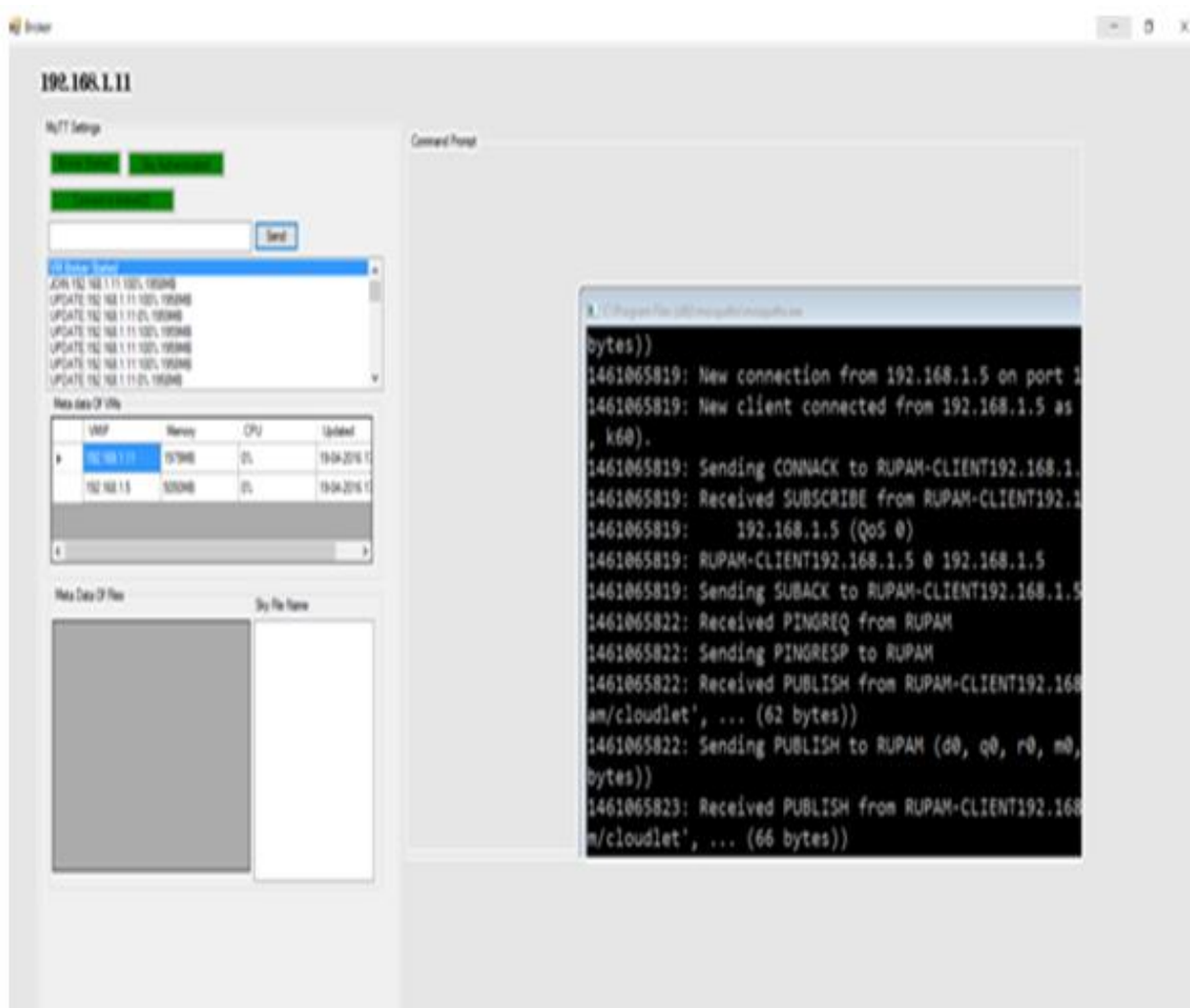


Figure 2: Data is split to two different virtual machines.

To broker any number of nodes or virtual machine can be connected, which shows details of chunks of data stored with in connected virtual machine.

Broker- Broker manages overall task of division and reprint of data. All virtual machines are connected to a broker, when user uploads file, broker checks for how many virtual machines are active then splits text to connected nodes and also backup is stored in sky drive.

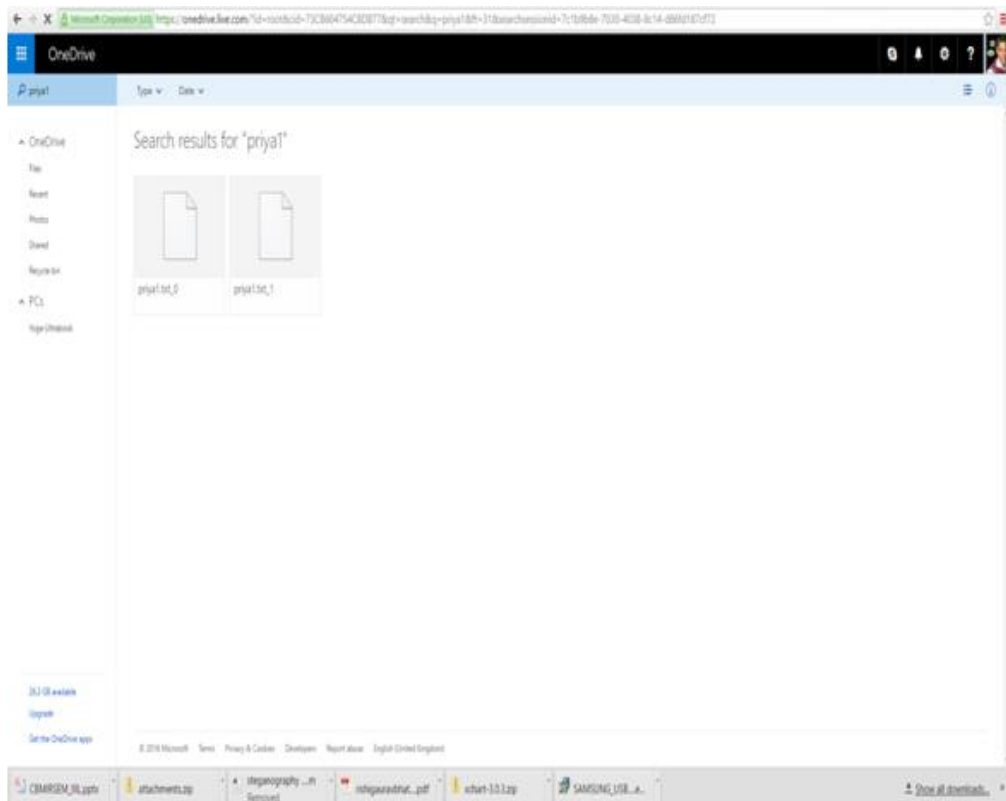


Figure 3 : Uploaded file is splitted into two chunks and stored in sky drive.

After user uploads files, text is divided into parts by broker and stored usually both connected virtual machine and sky drive cloudlet as slices where no information can be leaked to others.

VI.RESULTS

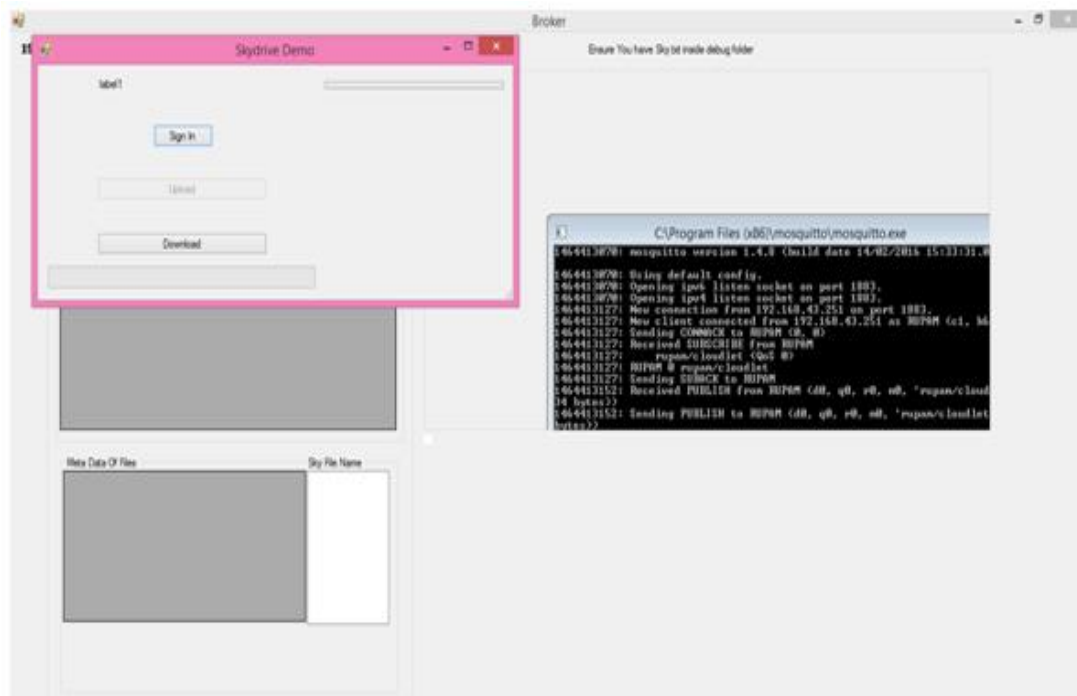


Figure 4 : skydrive signin

Skydrive is the actual cloud used in this project , where many local clouds or cloudlets are connected to globe cloud server to access data stored.

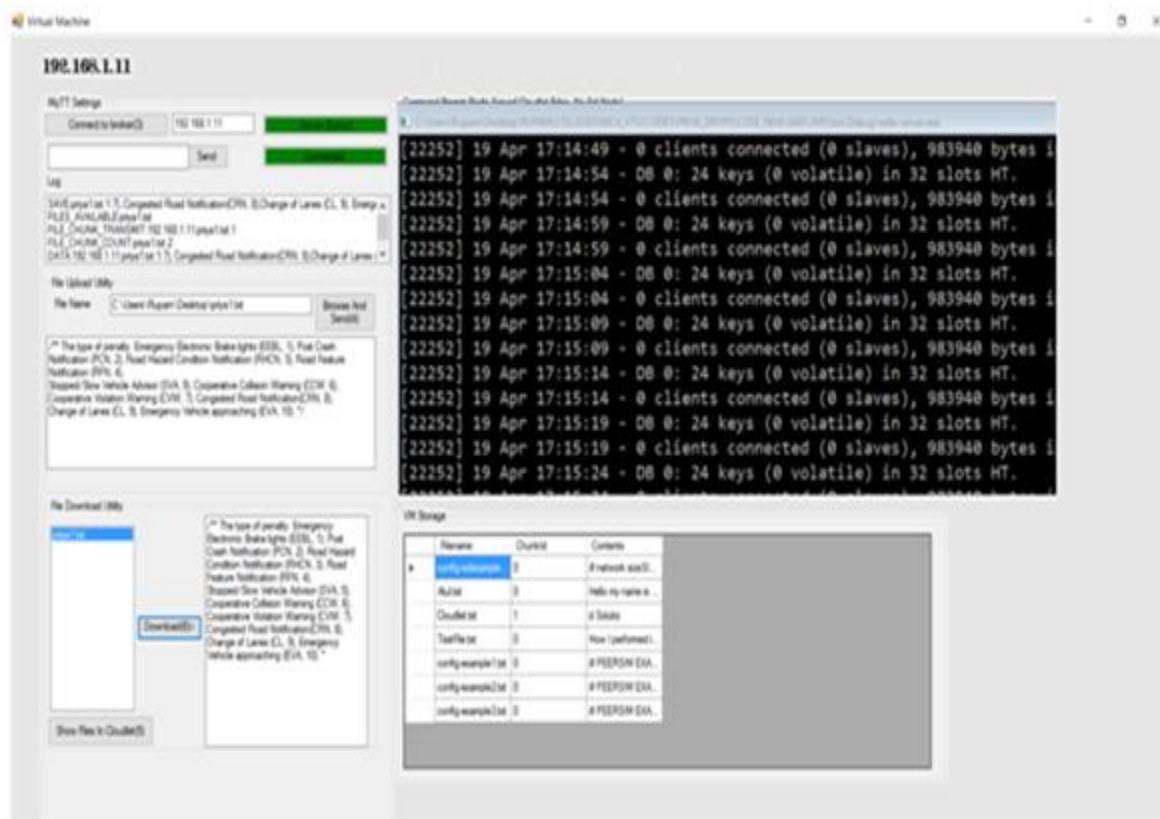


Figure 5: Download successful.

VII. CONCLUSION

Currently issues of storing large data leads to less secure as it is stored in single node. We proposed methodology called slicing and replication of data within cloudlet, where user upload and download file can be performed quickly as compared to the earlier. Files can be successfully stored to cloud. The performance of the file with in cloudlet requires very less time for user. The methodology of slicing and replication of data can be more secured using Advanced Encryption Standard and cipher text encryption which includes encryption of a file after uploading and decrypt file while downloading , key should be generated to the authenticated users. Based on the replication process used scheduling and unscheduling algorithm can be used.

REFERENCES

- [1] Khodade Nittin Balasaheb, Najan Vaibhav Kailas and Kalyankar Amrut Gopalrao, "Dynamic data possession in cloud computing system using fragmentation and replication," Imperial Journal of Interdisciplinary Research (IJIR) Vol-2, 2016.
- [2] S.Sangeetha, G.Vidhya, "An enhancement of secure data object in cloud environment," International Journal of Emerging Technology and Advanced engineering, Volume 6, Issue 1, January 2016.
- [3] Janjalkar Priti S, Gade Pooja V and Mate Shweta A, " Multi cloud hosting for performance optimization and security," International Journal for Scientific Research & Development Vol. 3, Issue 08, 2015.

- [4] Manisha Kakal, Sona Malhotra, “*Replication for improving availability & balancing load in cloud data centres,*” International Journal of advanced research in computer science and software engineering, Volume 5 ,Issue 4, 2015.
- [5] G.Malathy, Rm.Somasundaram,”*Performance enhancement in cloud computing using reservation cluster,*” European Journal of Scientific Resarch ISSN 1450 Vol.86 No 3 September, 2012,pp.394-401.
- [6] M.Muthu,R.Sachin Kumar Jaiswal,M.C.Sarthkumar,A. Sudhanram,”*Security enrichment of data in web cloud using DROPS,*” International Journal on Applications in Engineering and Technology Volume 2: Issue 1 : January 2016.
- [7] P.Anand,R.Bharath,C.Ganapathy,S.Sam Victor ,M.Vigneshwaram,”*Secure Dynamic fragment and replica allocation in cloud storage,*”International Journal Of Informatic & Technological Engineering Volume(3) Issue(3) March 2016.