International Journal of Scientific Research and Engineering Development-– Volume3 Issue 1, Jan-Feb 2020

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Data Storage and Recovery in Cloud Environment Using Forensic Tools

Sajin R Nair[1],Vineeth M V[2],Smita C Thomas[3]

[1]P G scholar, Dept. of CSE, Mount Zion College of Engineering, Kadammanitta, Kerala, India

[2]Assistant Professor, Dept. of CSE, Mount Zion College of Engineering, Kadammanitta, Kerala, India

[3]Research Scholar VelsUniversity,Chennai,India

----------**********----------

Abstract:

Cloud computing is a computing paradigm and it provide the services such as upload and retrieve of different kinds of data. Accessing cloud storage service through internet and pay as your used subscription, methods and techniques to effectively store data and reduce storage security, susceptibility in the cloud storage. The customers might face with data lose in the cloud due to intentional deletion, unintentional deletion or device failure. In this paper deals with the most advanced data storage technique in cloud and also deals with the data recovery technique for deleted data. The forensic recovery is the next evolution of data recovery. The increasing technical and legislative complexity of cloud computing system the forensic tools helps IT specialists recover data that has been accidently deleted ,intentionally erased or damaged through corruption. The proposed system PhotoRec is the one of the data recovery software which was a free open source and it is safe to store the recovered data on the system. In cloud computing environment the PhotoRec ensures data recovery in various situations for the customers without losing their data in cloud. The goal of our proposed method is to completely recover the contents or files which are deleted in cloud and also the recovery tool compared with other existing recovery tools. In this section we ensure that the proposed system will be evaluated with other similar forensic data recovery tools and also presented that the simple method using PhotoRec application to recover the data after their deletion in cloud environment.

Keywords **—Cloud Computing, Data Storage, Cases For Losing Data In Cloud, Data Recovery, Cloud Forensic, Comparison Of Forensic Tools, Forensic Reconstruction.**

----------**********----------

I. INTRODUCTION

Recently, the processing and the storage of huge volumes of data have been enhanced enormously in these last decades due to the emergence of Cloud computing. This concept is determined as a computing paradigm for infrastructure for computing resources, where a large pool of systems are connected in private, public or hybrid networks. Cloud computing include on demand self-service, broad network access, resource pooling and rapid elasticity are the different characteristics. The cloud computing service models are Software as a Service (SaaS), Platform as a Service (PaaS) and the Infrastructure as a Service (IaaS) [1]. Consumer uses the provider's applications running on a cloud infrastructure in SaaS model. In PaaS, an operating system environment, hardware, and network platforms are provided, installs its own software and applications for customers. The hardware and network provides in the IaaS model.

Cloud services are typically provided through a private cloud, community cloud, public cloud or hybrid cloud [1]. Services are offered through the Internet in public cloud and are owned and controlled by a cloud service provider. In a private

cloud, services provided only for a single organization, and is managed by a third party. In a community cloud, infrastructure may be controlled by the organizations or by a cloud service provider (CSP), the service is shared by several organizations and made available only to those groups.

With the rapid development in the computer science from the last couple of years lead to latest trends like distributed processing, resource sharing, access controls and so on. This has even led to the development of newer technologies like Cloud Computing, Big Data analysis, and the Internet of Things (IoT). Cloud computing is one of the services which has been getting adopted at the faster rate because of the main of the advantages provided by it for various purposes.

Nowadays there are many cloud service providers, as it is easier for providers compared to previous times. Customers can use the services from anyone without knowing much about the service provider. Customer needs to choose an appropriate or correct service provider who can serve their needs and takes most of the security issues seriously and takes proper care in protecting their data and services at the competitive prices.

Data recovery is one of the most important steps in the forensic process as it takes place in the beginning of the forensic process and heavily affects the quality and effectiveness of all forensic phases in later stage, especially forensic analysis. If the customer has accidentally deleted their data, we wanted to check whether they will be able to reconstruct their data using any techniques. For that purpose, we have investigated the forensic tools which can recover the data from the devices even after deleting it from them. We proposed a simple framework adaption in the cloud to support our method and also compared the proposed method with other data recovery tools in the cloud. We have used one of the well-known tool for reconstructing the data using PhotoRec, we were able to reconstruct the data after deleting and recover the data. made errors, etc

II. LITERATURE SURVEY

Salman Akintoye, AntonieBagula, YacineDjemaiel, and NoureddineBourgin [1] view some methods and techniques to store data in cloud computing. The different techniques are, Storage Techniques for cloud, data management in cloud storage, data retrieval from cloud, authentication schemes for cloud storage, and data integrity and availability for cloud storage. This paper provides a survey of some proposed cloud storage methods and techniques, their advantages and drawbacks and makes stress on the current requirements for storage techniques in cloud computing.

The work in [2] elaborates on the idea of a several disk images representing the evolution of data stored on a computer as a result of user actions. The disk images were processed using a selection of recovery tools and the comparison results are presented. A holistic data comparison of the tools is conducted and the recovery of known marker files that were deliberately added to the disk image are assessed.

Ms.PriyankaSalunkhe, Mrs.SmitaBharne, Mrs.PujaPadiya [3] Forensic analysis is nothing but recovery of data from digital devices. Various tools and applications are built for digital forensic but they have certain limitations. Technical challenges implement small scale of data mining in which decision tree (DT) (add first time full form) can support for fast and efficient classification of data Finding victim system or recognize attack pattern is very difficult task to forensic investigator. Various forensic tool kits are available to detect a crime activity. But such kind of tools works within some limitations. These toolkits are expensive.

The work in [5] Forensic Investigator uses a special tool for the above purpose. Most familiar forensic tools, such as FTK and EnCase, are all commercial software. In that respect are several forensic tools are available, but they are for desktop application and are very costly. The investigator needs to convey the laptop and introduced tool with them on the crime sites. Today with the help of cloud technology we can access any information related to anyone from anywhere at any time, but

this arises a new threat to private and confidential information. And in that case the cloud based computer forensic tool will run better in this situation where examiners can access tool form cloud portal and perform investigations.

III. STORAGE TECHNIQUES IN CLOUD COMPUTING

In this section, we review two techniques to store data in cloud computing. They are Dynamic Data Deduplication in Cloud Storage and Deduplicated Service with CDMI Standard [1].

A. Dynamic Data Deduplication In Cloud Storage

The dynamic deduplication scheme for cloud storage, improve storage efficiency and maintaining redundancy for fault tolerance. Data deduplication could be a technique accustomed cut back cupboard space and network information measure. In existing deduplication systems duplicated information chunks establish, store only one replica of the data in storage and logical pointers are created for other copies instead of storing redundant data. The existing deduplication schemes could forestall the system fault tolerance since it's going to be that many files talk over with an equivalent information chunk which can be unavailable because of failure.

The dynamic deduplication theme was planned to balance between storage potency and fault tolerance needs and address limitation of static deduplication theme that can't deal with dynamical user behavior.

For instance, data usage in cloud changes overtime; some data chunks may be read frequently in a period of time, but may not be used in another period. Dynamic deduplication theme has the potential to adapt to numerous access patterns and dynamical user behavior in cloud storages. The planned system relies on client-side deduplication mistreatment whole file hashing. Hashing process is performed at the client, and connects to any one of de duplicators according to their loads at that time then identifies the duplication by comparing with the existing hash values in Metadata Server. The system is composed of the following components: load balancer that requests from clients sending to any one of de duplicators according to their loads at that time; Deduplicators which establish the performed duplication; Cloud Storage, a Metadata Server to store metadata and File Servers to store actual files and their copies; and Redundancy Manager to identify the initial number of copies, and monitor the dynamical level of Quality of Service (QoS). The system model was stimulated using HDFS, one Name node as Metadata server, and five Data nodes as Fileservers. Three events were simulated: upload, update, and delete.

The transfer event is once the file is 1st uploaded to the system. If files exist already within the system, and are uploaded once more, the number of copies of the files will be recalculated according to the highest level of QoS.

B. Deduplicated Service With Cdmi Standard

The data deduplication private cloud storage system with Cloud Data Management Interface (CDMI) standard based on the fundamental DFS. A data deduplication scheme is implemented in the system to reduce cost and increase the storage efficiency. Gluster is chosen as the basic for DFS to implement proposed private cloud storage system. The planned non-public cloud storage system consists of 5 parts as shown in Figure 1:Client that communicates with Controller in Front-end node and exchange information; Front-end node contains Apache server that redirects the requests that square measure sent from CDMI request sender to reinforce load balance; adapter node receives CDMI request and stores files via Gluster consumer.

Storage nodes contains GlusterFS server, which might produce differing kinds of volume for various functions. The system provides the following three main functionalities: upload file, download file, and delete file. During the file transfer method, hash value of the file is calculated by Hash generator in Client and sent to the Controller where the hash value is compared with all file metadata stored in Database. Controller will notify Client that file doesn't exist, if the file is not duplicated. Then, the file's information are going to be sent to Controller to insert it into info and CDMI sender in consumer can send transfer file request. After receiving the

request, Load balancer can route it to CDMI Server by the apache load balancer computer hardware algorithms.

Finally, adapter node can store the file to Storage node and can inform consumer with uploading finished response message. But, if the file is duplicated, Controller will notify the Client that the file exits, then Client will send this file's metadata to Controller to insert it into database which will produce associate empty information file with an equivalent file name Storage node. For transfer practicality, Client initiate request and send a file_ path to the Controller.

The Controller can search info for the real_ path of this file and answer consumer.

Then, client's CDMI sender will send CDMI download file request and the Load balancer redirects the request to CDMI server by the apache load balancer scheduler algorithms. Finally, adapter nodes get the requested file from GlusterFS and transmit it back to consumer.

The delete practicality is classed into 2 varieties in line with the connection among file_ path, real_ path and whether the file is shareable. If file_ path is capable real_ path and also the file is shareable, the file is moved to the other place rather than delete directly. Controller question info by file_ path to seek out an appropriate path for moving files. After removing the files, the Database will be updated and Client will receive the delete response message. But if file_ path isn't capable real_ path, or file path is equal to real path and the file is not shareable, the file can be deleted directly. After receiving a delete message that contains file_ path, Controller can delete file in line with the file_ path directly and consumer can receive the delete response message.

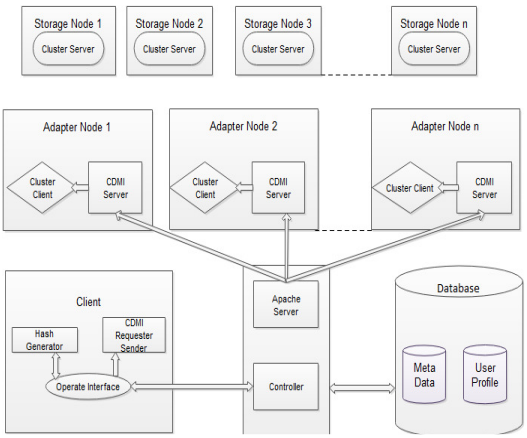

Fig. 1 A file- deduplicated private cloud storage service with CDMI standard

IV. CASES FOR LOST DATA IN CLOUD

Deleted or lost files in any system is a common issue which can occur due to many reasons such as:

A. Intended Delete

Sometimes when we use the devices, we think some of the information is not required. But after some time when we start to look for some other information, we recall that particular files which were deleted are also important for this particular task which we are looking in to at the moment. In such scenarios, it gets completed. If we have not added new content or files to those memory locations (which is unknown to the customer), there is it still a probability that content still present on the memory locations. If we use appropriate tools to recover those data might solve huge problems.

B. Unintended Delete

It is possible customers to delete some of the files unintentionally. After the deletion customer might realize that was a mistake. In such scenarios, using the appropriate recovery tools will help the customer to recover the data without hustle.

C. Hardware/Device Failure

Failure in memory devices or hardware in the cloud might happen, but customers will not know about these failures. But cloud service providers might have a backup for the customer data. In case of service providers doesn't have the backup, they can also use forensic tools to recover the data using memory reconstruction.

V. METHOD OF RECOVERY FOR DATA IN CLOUD

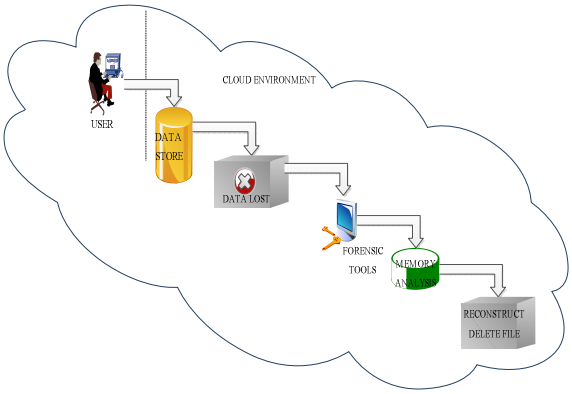

Fig. 1.Method of Data Recovery in Cloud Environment

A. PhotoRec

It is one of the data recovery software available to reconstruct lost or deleted files or contents from various types of memory devices [2]. It works on different types of formats irrespective file system. PhotoRec tries to read the data from the memory blocks from unallocated and allocated memory blocks. If the file system is corrupted, still it can recover the data from reading the memory blocks. Even if the memory devices are damaged, still PhotoRec will be able to recover the data from those devices. PhotoRec is a free and open source application which supports various operating systems such as Windows, Linux, Mac OS X, and so on. Which is mainly distributed under GNU General Public License (GPLV v2+).PhotoRec works on the principle of recovering the lost data partitions on different kinds of files systems. PhotoRec reads the devices or drives to recover the lost data. To recover the data from a spoiled device or accidentally deleted, it is more appropriate to start the recovery using the PhotoRec before the lost files are overwritten with the any of the changes made after their deletion. It is safe to store the recovered data on a different device to make most out of the recovery.

B. Install

It is very easy to install the PhotoRec application. It is available within the package of testdisk. For Linux machines in the cloud, customers can install the required packages with install command based on their operating system.

C. Working

Many of the file systems such as ext2/ext3/ext4, FAT, NTFS store the files in blocks, these blocks will be stored in some fixed number of sectors once the file has been created. In many of the present operating systems (OS), OS will try to keep the data in continuous sectors to speed up the system performance and minimize fragmentation. This helps in the system performance while performing the several read and write operations. When a user tries to delete a file from the system, metadata related to the files will be deleted but the actual data will be still present on the drive until that memory location has been written with a new file in physical memory. As these files are still present on the drive, it is still possible to recover the deleted files. To recover these files, PhotoRec will first find the block size. This information usually available in the boot record for windows or super block for Linux machines. If the file system is corrupted, PhotoRec reads the data by sectors and find first 10 files. Based on these files it will calculate the block size from their location on the devices. Once the block size is calculated, it will start fetching the data block by block. Then these blocks will be compared with database available in the PhotoRec to recover the correct format of the files.

VI. IMPLEMENTATION OF SYSTEM FOR DATA RECOVERY

Our implementation for recovery does not include any structural changes to the existing cloud architecture [3]. The goal of our method is to completely recover the contents or files which are deleted in the Cloud. When a file or content is deleted, and the customer wants to recover the data, customer needs to call the PhotoRec application with the device to recover the data. When a file is deleted in most of the operating systems, the memory location will be moved to available memory instead of completely deleting the file. So, when the customer calls the PhotoRec application, they will be asked to choose the device from which

the files need to be recovered and provide the location where those recovered files need to be shared. Based on the type of files selected, PhotoRec will read Free or complete memory to reconstruct the files. The application will just read the memory, it will not make any changes to those memory locations unless we store on the same device.

A. Sequence of steps

Once the customer has deleted, they can follow the following sequence of steps to recover the deleted files from memory traces using any Forensic Tool as shown in the Fig.3:

1. If the customer needs to install testdisk to have PhotoRec on the machine. If it exists, this step can be skipped.
2. Customer needs to start the Forensic Tool (PhotoRec) and select all types of data which needs to be selected and give the path to store the recovered files.
3. Forensic Tool (PhotoRec) will start memory analysis based on free and unallocated memory available in the machine or system.
4. Then Forensic Tool will start to recover the databased on memory analysis for specific selected types of data. It will show us what type of files were recovered and how many of them were recovered on the terminal output.
5. Once the recovery is completed, the customer can quit the PhotoRec application.

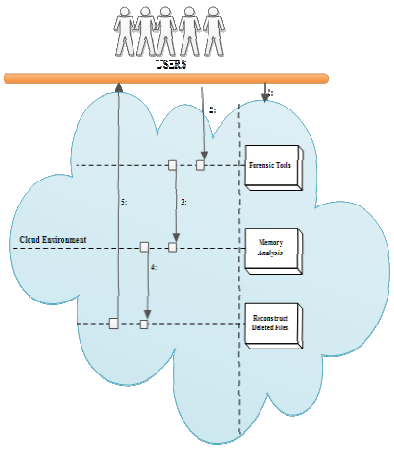

Fig. 3 Sequence diagram for data recovery

VII. CONCLUSIONS

In this paper, we have to implement a simple method, PhotoRec application to recover the data after their deletion in cloud environment. We have discussed the usability, comparison with other forensic data recovery tools and evaluate with various factors. PhotoRec has good performance on the case of recovering deleted files. However the logically damaged cases, unexpected deletion of data it can be recovered at least half an hour to recover the data. The method has achieved almost 98.49 % of recovery from deleted data. Various factors such as device usage, performance, applicability time of deletion, and time of recovery we can concluded that the PhotoRec is more applicable for recovering the deleted data from the cloud environment. In our future work, we would like to check more advance techniques that support to this method.

REFERENCES

[1] Salman Akintoye, AntonieBagula, YacineDjemaiel, and NoureddineBourgin “A Survey On Storage Technique In Cloud Computing” International Journal Of Computer Applications (0975-8887) April 2017.

[2] Joe Buchanan_Wollaston, Tim strer and William Gilsson “Comparison of the Data Recovery Function of Forensic Tools”.(331-347)Feb 2017.

[3] Ms. PriyankaSalunkhe, Mrs.SmitaBharne, Mrs.PujaPadiya “Data Analysis of File Forensic Investigation”International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)(372-376)-2016.

[4] Elisa Bertino and Ravi Sandhu”Digital Forensic Science Issues, Methods, and Challenges”Synthesis Lectures On Information Security, Privacy, & Trust 2016.

[5] Monali P. Mohite and S. B. Ardhapurkar “Design and Implementation of a Cloud Based Computer Forensic Tool”Fifth International Conference on Communication Systems and Network Technologies (1005-1009)2015.

[6] TheodorosSpyridopoulos and VasiliosKatos “Chapter 17Data Recovery Strategies for Cloud Environments”(377-379) 2015.