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 9, September 2016

A Survey on Coral - A Cloud-Backed Frugal File System

P. Rohini[1], K. Palraj[2]

PG Scholar, Computer Science and Engg, Sri Vidya College of Engineering & Technology, Virudhunagar, India [1]

Assistant Professor, Computer Science and Engineering, Sri Vidya College of Engineering & Technology, Virudhunagar, India [2]

Abstract: At present dramatically increase in the business and internet applications, storage is becoming a major issue in cloud computing. Storage costs are increasing day- by-day. Cloud backed is backing up data that involve distribution copy of the data over a community network to an off-site server. Uncomplicated access interfaces and elastic billing models, cloud storage has become a gorgeous solution to make simpler the storage organization for both enterprises and individual users. This paper presents a survey on the different cloud backed frugal file system. This enables effective storage management, increase the performance and reduce the cost in the cloud.

Keywords: cloud computing, cloud backup, billing models, frugal file system.

I. INTRODUCTION

Backup file, data archival and collaboration are the popular services in cloud companies [1], in general these services based on cloud storages like the Amazon S3, Drop box, Google Drive and Microsoft Sky Drive. These services are fashionable because of their everywhere accessibility, pay-as-you-go model, high capability, and ease of use. Such services can be generally grouped in two modules: (1) personal file synchronization services (e.g., Drop Box) - Personal file synchronization is based on back-end storage cloud model and the applications of client communicate with the local file system by monitoring interface [inotify -in Linux]. (2) cloud-backed file systems (e.g., S3FS [6]). Cloud-backed file system based on two architecture models: the First model is proxy based, second model is open-source solutions [S3FS [2] and S3QL [3]]. The two models are implemented at user – level. Proxy based model the proxy component placed in network infrastructure, performing as a file server to various clients. Functionality of Core files system is implemented by proxy, to calls the cloud and stores the files. The major limitation is bottleneck and single point of failure. Open source solution model the clients directly access the cloud, exclusive of proxy interaction as a result, there is no longer a single point of failure, but it's very harder to control the file sharing between the clients when miss the suitable rendezvous point for synchronization.

Cloud Backed File System:

Cloud backup [4] also identified by online backup, is an approach for backing up data that involves a replica of the data over a public network to an off-site system. Cloud Backed is models that provide data backed up remotely, maintained and managed. Users access the data through the network. Users normally compensate for their data storage on cloud as per-usage or monthly rate. The cloud Storage providers provide a platform as a service, is one of the infrastructure service on cloud storage to shorten storage management for enterprises and personality users.

Table I Cloud backup vs. Tape or disk

Services	Cloud backup	Tape or disk
Data	Non – critical data	Critical data.
Storage policy	Retention policy	Recovery time objective.
Recovery	Disk to disk- to tape strategies to disk to disk to - cloud.	Tape or some other portable storage media
Flexibility	No additional hardware is required	Additional hardware is required

Fig.1 Cloud back up service

DOI 10.17148/IJARCCE.2016.5997

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 9, September 2016

Implementing cloud data backup is able to help boost an organization data protection without raising the workload on information technology.

Online backup systems are classically built a client software application that run on a program determined by the purchase stage of service. Cloud backups contain the software and hardware component to keep an organization's data, include applications Exchange and SQL Server. Online backup is used by small and medium-sized businesses (SMBs) and larger enterprises to back up the data. For larger organization, cloud data backup as a complementary form of backup.

FRUGAL CLOUD FILE SYSTEM:

Different embodiment grant a techniques and tools of providing a frugal cloud file system[5] that proficiently uses the blocks of different types of storage devices with special properties for various purposes. The various types of storage strategy reduce the storage and bandwidth transparency. Favorably, the storage and bandwidth reduction in the clouds achieved by the frugal cloud file system, reduce the cost-effective of managing the file system at the same time, maintain high performance. Frugal file system is a structure that optimizes on the whole storage cost between various Cloud storage services and different type of price. The Frugal file system's storage services like a twin storage system, one is low latency (e.g., Amazon ElastiCache) and the other one is high latency (e.g., Amazon S3). In low latency the data transfer cost is low and cost of storage per byte is high (i.e. cache) but the high latency, cost of storage per byte is low and data transfer cost is high (i.e. disk).

II. VARIOUS FILE SYSTEM FOR CLOUD BACKUP

Distributed File System (DFS)

Distributed file system (DFS) is the file systems for cloud. DFS allows the clients to access data. Data files are separated by parts as chunks that stored on various remote systems which offer the parallel execution. Data are stored in files in the format of hierarchical tree structure. Directories are denoted by nodes. DFS facilitate any type of enterprises (such as large, medium, small) that allows storing the data and accessing the data on remotely. DFS allow two type of file system as GFS and HDFS. Both file systems are handled the batch processing. Hadoop distributed file system (HDFS) designed for access the terabyte's data or peta bytes data. HDFS is master slave architecture. It consists of Name node and Data node mechanisms. Name node manages the storage of Metadata and Data Node manages the node storage. HDFS file systems the files are divided into blocks. Each block contains various data nodes and every node is replicated for availability. This is the block level replication. Name node manages the operation of name space and map the block to data node. HDFS is characterized by method of Data re balancing. Google file system (GFS) is distributed file system that provide few faults –tolerant, high data storage and performance many clients access the data simultaneously. GFS allow the map reduce concept for access the multiple machines. GFS is Single master server and multiple client architecture. The files are separated into chunks that stored in server chunk. Coordinating the resource storage and Metadata files managements is responsible for master server. In the GFS file processing done by two phases sending phase and writing phase.

BLUE SKY

Blue Sky is a file system of network based cloud storage. Cloud backup provide a persistent data storage, provided by Amazon s3 or windows Azure. Blue sky allows the consistency and large storage capacity and reduce use of hardware sever. Client access the cloud storage server with help of proxy running on –site. Cloud optimization is achieved by log-structured design and secure cloud log cleaner. It uses multiple protocols as NFS and CIFS for various providers. Blue sky file system is maintained by object data structures format and log structured format[8] for cloud organization. Blue sky provides version ed store data for backups in the file system. Data and Metadata is represented by blue sky object; it's presented in log structured file system, format is data blocks, inode , inode maps and check points. For storage purpose log segments objects are aggregated. Blue sky standard file system semantics is POSIX with atomic hard links and renames. Blue sky provides a transparent proxy based structural design, with the purpose of store the data per mentally on cloud storage providers to clients in an enterprise. File data are store in data blocks. Files are separated by constant size blocks. Blocks size is 32 kb. Inodes contain the details of basic Metadata like ownership, access control timestamps, directory inode reduce the path traversals. Inode map list provide the location of inodes because inode data are not stored in permanent location. Origin of File structure snapshot is determined by checkpoint objects that locate the present inode map objects. The main use of checkpoint is managing the file system integrity. For every file structure, Blue Sky maintains a part of log for every writer to file structure. Classically there are two: the proxy managing the file system and cleaner - garbage collect data. Every writer stores the log segments to separate index that provide independent updates to file system, all log contains of a number of log segments, and log segment group the objects mutually into the fixed size block. Blue Sky needed a file system cleaner to garbage collect.

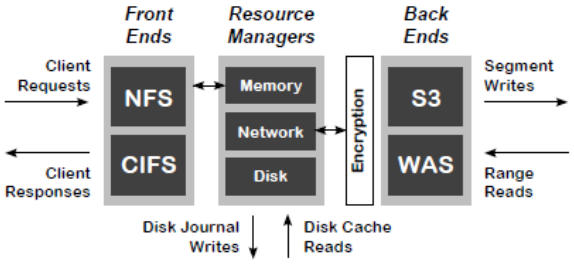

Fig.2 Back up file system - blue sky

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 9, September 2016

Different traditional disk-based systems contain the Flexible nature of cloud, the cleaner is not essential. Blue Sky cleaner's run on proxy. Example of blue sky cleaner provider is Amazon EC2 and S3. The log structured design allows snapshots for backup purpose. Snapshot tool record the list of checkpoints to protect the log segments from deletion.

SCFS: A Shared Cloud-backed File System

In spite of their increasing Technology, current cloud-backed storage systems still have various restrictions related to reliability, durability assurances and inefficient file sharing. SCFS, a cloud-backed file system overcome these challenges and provides strong consistency and POSIX semantics for cloud backed services. It uses a plug gable back plane allows the various cloud storage or a cloud-of-clouds. The main goal of SCFS is assurance of security like integrity, confidentiality, availability and also supports consistency-on-close semantics [28], SCFS is not proposed to be a big-data file system, because file data is downloaded from and uploaded to one or more clouds. SCFS contain the backend cloud storage, co-ordination service and SCFS agent. File data are maintained by backend cloud storage. Metadata management and synchronization supported by co-ordination service. SCFS functionality and client file system mounted by SCFS agent. SCFS provide the strong cloud consistency based on two approaches as maintaining the Meta with limited capacity data and save the data itself. SCFS support two prototype co ordination services: zookeeper [29] and Depspace [13]. The two services are integrated with SCFS wrappers. The co –ordination services simulated the fault tolerance. Zookeeper need 2f + 1 replica for tolerate f crashes by using Paxos-like protocol. Depspace need 3f + 1 or 2f + 1 replica for tolerate f arbitrary faults by using BFT-SMaRt replication engine. SCFS cloud storage services are Amazon S3, Windows Azure Blob, Google Cloud Storage, Racks pace cloud files and all services using cloud of cloud back end. Cloud back end use the extended version of DepSky, support a new operation as read the all versions of hashes data are stored in Depsky's Metadata internal objects and stored in cloud. Based on the consistency and sharing requirements of stored data the SCFS operation divided into 3 modes. 1) Blocking mode 2) Non –blocking mode 3) Non- sharing mode. The main uses of SCFS are backup, disaster recovery, and file sharing control and without require dependence on any single cloud provider.

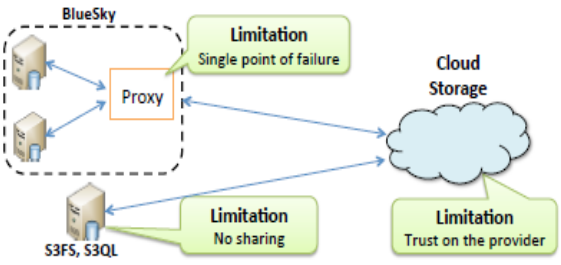

Fig.3 Back up file system - SCFS

DEPSKY: Dependable and Secure Storage in a Cloud-of-Clouds

DEPSKY, cloud backup improve the availability, integrity and confidentiality of information store in the cloud with help of encoding, replication and encryption of the data on varied clouds that make a cloud-of-clouds. DEPSKY is a reliable and protected storage system that gives the profit of cloud computing by using an arrangement of diverse commercial clouds to cloud-of-clouds. DEPSKY also give the virtual storage, it is accessed by users while invoking the operations. DEPSKY also provide the four limitations such as Loss and corruption of data, Loss of privacy, Vendor lock-in, Loss of availability. DEPSKY System use data and system models. It contains two main algorithms as DEPSKY – A and DEPSKY – CA and also contains the set of auxiliary protocols. Two algorithms are implemented by software library in the clients. Data model contain the three abstraction levels. First level, the conceptual data unit has unique name, version number – support the object updates, data verification – a cryptographic data hash. Second level, Conceptual unit is implemented by generic data unit, has two types of files: signed Metadata file and storage file. Third level, data unit are implemented. Data unit support the operation of storage objects like creation of Metadata file, destruction of data unit, write operation and read operation. System Model uses the asynchronous distributed system; it's composed by writers, readers and cloud storage providers. Quorum protocols can provide as the backbone of storage systems. Quorum protocols contain the individual storage nodes instead of servers. Many protocols involve several steps to access the shared memory, it makes unrealistic for geographically isolated distributed systems such as DEPSKY. The DEPSKY protocols need two communication round-trips to read or write the metadata and data files. Byzantine fault-tolerant (BFT) storage is implemented by several protocols. But its require server for execute code and functions; it's not available on cloud storage. This the key difference between DEPSKY protocols and BFT protocols. DEPSKY – A is the protocol of DEPSKY, it improves the availability and integrity of storage cloud by replication using quorum techniques. DEPSKY -A include read and write algorithm.

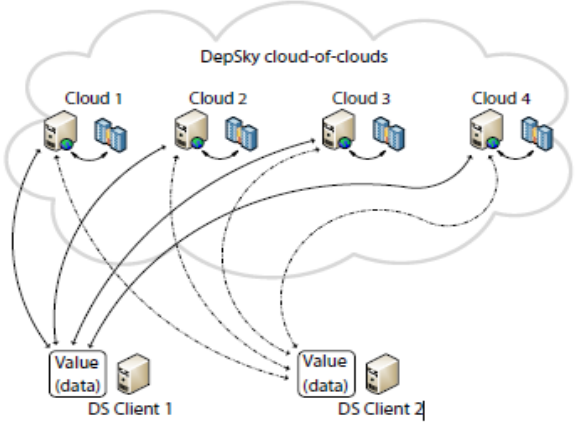

Fig.4 Back up file system – Depsky

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 9, September 2016

The DEPSKY-A protocol has two major restrictions. First, one is data unit size and costs, the data stored in single cloud. Second one is data storage is clear text, so it does not provide confidentiality guarantee. DEPSKY – CA protocol overcome these problems and also has the additional cryptographic function and coding functions. DepSky-CA write algorithm's encryption technique generates the key sharing.

CHARON

CHARON is one of the cloud backed file system that able to store and share the large amount of data between various cloud providers and cloud storage system in secure, reliable manner. The two main feature of CHARON is server less design and efficient management of file system. CHAON support three types of data locations as cloud of clouds, public cloud storage and private cloud storage. Cloud of clouds provides mufti cloud availability, confidentiality. Single storage cloud is low cost compared to cloud of clouds but it requires confidence provider. Private cloud storage based on adopted method and solution, also provides the dependability level. CHARON data are separated by file data and Metadata. Metadata are stored in cloud of clouds. CHARON use data centric. Byzantine-resilient leasing algorithm which ignores the concurrency conflicts. CHARON divides the files into constant size blocks. Files are stored in various data location based on the requirements. POSIX interface is provided by CHARON that allow the user interact with any file system.

CHARON cloud storage providers are Amazon S3, Windows Azure Storage, Backspace Cloud Files, and Google Cloud Storage. CHARON consists of two design concepts: first design is writes on absorbs file and the second design is remove write – write conflicts and mechanism of ruling out optimistic. CHARON design implementation has main three challenges a: 1) Ability to deal multiple cloud storage locations, 2) Proper file system management and 3) concurrent access to the file system. CHARON use modular based approach for non fault tolerant, that build service of cloud.

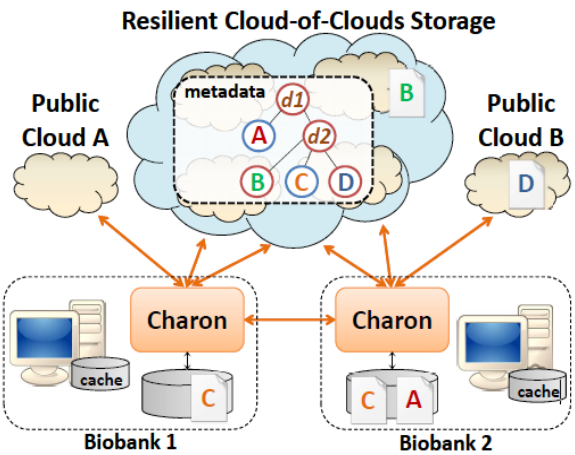

Fig.5 Back up file system – CHARON

Drop box:

Drop box is a cloud based services. It is used for data storage, data access and data management. Client can storage and access the data through desktop, mobile os. The Drop box back up is done automatically when the drop box client drop the file in to selected folder. Drop box server and client programs are written in python language. Drop box use version control for revise and re post the files. Drop box use revision history which is used for recover the data easily. Version history is combined with delta encoding technique. Drop box uses AES -256 bit encryption method and SSL for synchronization.LAN sync technology is provided by Drop box that allows the file download in securely. Drop box business model is Freemium. The user can access the cloud storage via free account and paid account. Control server and data storage server is major component of Drop box architecture.

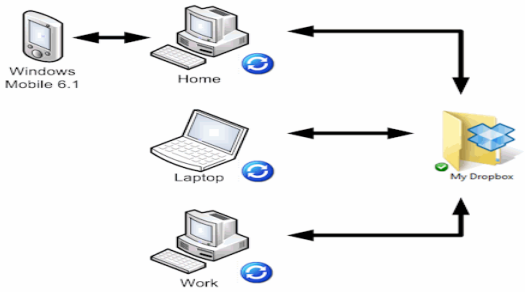

Fig.6 Back up file system – Drop box

III.DIFFERENENT ALGORITHM FOR CLOUD BACK SYSTEM

Seed block algorithm: The user information is collected from several remote locations without network connection. The files are recovered by seed block algorithm. The seed block algorithm mainly concentrates on security. It is used for back up file storage on remote system and excluding any encryption techniques. Seed Block Algorithm (SBA) provides the efficient and simple Back-up and recovery process. SBA computation based on XOR (exclusive – OR) operation. Example: The two data files samplefile1 and samplefile2 are stored in cloud storage. SBA performs the computation operation XOR on these two files and produces the result as result file (samplefile1 + samplefile2). When the samplefile1 is missing in the cloud storage, to recover the samplefile1 by XOR (i.e.) samplefile1 = (result file samplefile2).

HSDRT: HSDRT is a modern file backup method. It is suitable for laptop, smart phones etc. HSDRT generally use distributed data transfer method with high rate encryption technique. HSDRT system segregates into two sections: backup and recovery. The cost of implementation is high.

(ERGOT): Relatively Efficient Rounding Grounded on Taxonomy (ERGOT) depends on semantic analysis but the focus of time complexity is failed. Semantic based

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 9, September 2016

technique supports cloud computing Discovery service. ERGOT is provides a data retrieval technique not a backup technique. ERGOT contains three mechanisms: Distributed Hash Table (DHT) protocol, Semantic Overlay Network (SON), semantic measure.

Linux Box: Linux Box is the uncomplicated methods for back-up and recovery through minimum cost. Still, security is very low level. Migration is possible between the one cloud service providers to another cloud service provider. The data transmission method use encryption technique. The main drawback of Linux box is entire virtual machine is sync to waste bandwidth.

Cold and Hot Backup Service: Cold and Hot Backup Service is based on triggering technique. If the service failures is detected then the backup service is triggered else not triggered the backup services. The two approaches are CSBRS and HSBRS. In Hot Backup Service substitutes the HBSRS approach through the execution of backup services in dynamic state. The recovery time is reduced by HSBRS.

SBBR: Shared Backup Router resources (SBBR) mainly focus on failure of router and cost reduction. It also provides the mufti layer signalling for network management system. SBBR uses IP logical connectivity for failure of router. The main drawback is inconsistency between the physical and logical configuration, which affects the performance of system.

Advantage and disadvantage of cloud backup:

Table III Pros and cons of cloud back up

Advantage	Disadvantage
Consistency and efficiency: Cloud provider use the technology as server virtualization, data deduplication, compression, and disk based backup, encryption, application specific protection, storage virtualization and also provide the monitoring and managing data 24/7, migration etc.	**Nonexistent – service level agreements:** Band width availability, data transfer, system accessibility, back up data guarantee is depends on agreement of service level.
Scalability and Cost: Organization provides a boundless scalability without upfront of principal payments. Pay – as – you go method allow the management of operational cost and grow of capacity.	**Size limitation:** It is major impact on cloud backup. Based on availability of bandwidth the organization fixes the threshold for data capacity.
Recovery time: Cloud storage file recovery is faster and effectiveness. In cloud storage the recovered data are located, streamed through the WAN. It minimizes the time and eliminating the infrastructure of local tape.	**Full recovery and seeding data:** Based on the capacity of data, the complete backup /complete recovery of cloud data site confirm the time consuming and data production.
Accessibility: Data are easily accessible from anywhere and anytime.	**Discontinuation of the service:** Exit strategy is important testing specific features as extraction of data, notification cancellation, and termination.

IV. CONCLUSION

This paper presented a survey on the different frugal file system for cloud backed services. The frugal cloud based file system improves the performance and cost for end users. Frugal cloud back up is the combination of intellectual data backup & recovery and simple unified solution that safe the organization data. It provides the organization's management services, disaster recovery plan, energy efficiency and cost reduction.

REFERENCES

[1] Future of cloud computing - 2nd annual survey results.http://goo.gl/fyrZFD, 2012.
[2] S3FS - FUSE-based file system backed by Amazon S3.
[3] http://code.google.com/p/s3fs/.
[4] S3QL - a full-featured file system for online data storage.
[5] http://code.google.com/p/s3ql/.
[6] [http://searchcloudstorage.techtarget.com/definition/cloud-storage.
[7] Krishna P.N. Puttaswamy, Thyaga Nandagopal and Murli kodialam "Frugal storage for cloud file system," in proceeding EuroSys'12 of the 7th ACM European conference on computer Systems,2015 pages 71-84.
[8] https://en.wikipedia.org/wiki/Distributed_file_system_for_cloud.
[9] Michael Vrable, Stefan Savage, and Geoffrey M. Voelker "Blue Sky: A Cloud-Backed File System for the Enterprise, "in Proceedings of the 10th USENIX conference on File and Storage Technologies, Feb 2012.
[10] M. Rosenblum and J. K. Ousterhout."The Design and Implementation of a Log-Structured File System," ACM Transactions on Computer Systems (TOCS), 1992.
[11] Alysson Bessani, Ricardo Mendes, Tiago Oliveira, Nuno Neves, Miguel Correia, Marcelo Pasin, and Paulo Verissimo "SCFS: A Shared Cloud-backed File System, " in Proceedings of the USENIX ATC on USENIX Annual Technical Conference 19&20-June -2014.
[12] P. Mahajan, S. Setty, S. Lee, A. Clement, L. Alvisi, M. Dahlin, and M. Walfish "Depot: Cloud Storage with Minimal Trust," In Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation (OSDI), Oct. 2010.

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 9, September 2016

[13] J. Howard "Scale and performance in a distributed file system" ACM Trans. Computer Systems, 1988.
[14] P. Hunt, M. Konar, F. Junqueira, and B. Reed. "Zookeeper: Wait-free coordination for internet-scale services," In USENIX ATC, 2010.
[15] A. Bessani, E. P. Alchieri, M. Correia, and J. S. Fraga "DepSpace:A Byzantine fault-tolerant coordination service," in EuroSys, 2008.
[16] StorSimple. StorSimple. http://www.storsimple.com/.
[17] TwinStrata. TwinStrata. http://www.twinstrata.com/.
[18] Alysson Bessani, Miguel Correia, Bruno Quaresma, Fernando Andr´e, and Paulo Sousa "DEPSKY: Dependable and Secure Storage in a Cloud-of-clouds," EuroSys 11-April- 2011.
[19] Ricardo Mendes, Tiago Oliveira, Vinicius Cogo, Alysson Bessani " The CHARON file system,"
[20] Idilio Drago, Marco Mellia, Maurizio M. Munafò, Anna Sperotto and Aiko Pras "Inside Drop box: Understanding Personal Cloud Storage Services," in Proceeding of IMC -12 of ACM conference on internet measurement conference,2012 PP.481-494.
[21] http://www.techrepublic.com/blog/five-apps/keep-your-data-safe-with-one-of-these-five-cloud-backup-tools/
[22] http://www.cloudwards.net/spideroak-or-wuala-which-is-more-secure/
[23] Kailas Pophale, Priyanka Patil, Rahul Shelake, Swapnil Sapkal **"**Seed Block Algorithm: Remote Smart Data- Backup Technique for Cloud Computing," International Journal of Advanced Research in Computer and Communication Engineering**,** Vol. 4, Issue 3, March 2015.
[24] Lili Sun, Jianwei An, Yang, and Ming Zeng "Recovery Strategies for Service Composition in Dynamic Network," International Conference on Cloud and Service Computing,2011
[25] Giuseppe Pirr´o, Paolo Trunfio, Domenico Talia, Paolo Missier and Carole Goble "ERGOT: A Semantic-based System for Service Discovery in Distributed Infrastructures," 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010.
[26] Xi Zhou, Junshuai Shi, YingxiaoXu, Yinsheng Li and Weiwei Sun "A backup restoration algorithm of service composition in MANETs," Communication Technology ICCT 11th IEEE International Conference, 2008, pp. 588-591.
[27] Ms...KrutiSharma, and Prof K.R. Singh "Online data Backup and Disaster Recovery techniques in cloud computing: A review", JEIT, Vol.2, Issue 5, 2012.
[28] Eleni Palkopoulouy, Dominic A. Schupke, Thomas Bauscherty "Recovery Time Analysis for the Shared Backup Router Resources (SBRR) Architecture", IEEE ICC, 2011.
[29] http://searchcloudstorage.techtarget.com/definition/cloud-storage.