REDUSING THE STORAGE SIZE USING COMPRESSION ALGORITHM AND CLOUD BACKUP USING EXOR

[1]S. N. Uke, [2]Mohini Shinde, [3]Shweta Joshi, [4]Priya Sharma, [5]Rushikesh Rathod
[1]Assistant Professor ,Smt. Kashibai Navale College of Engineering,
Vadgoan, Pune, India
Email:[1]snuke@sinhgad.edu, [2]mohini.shinde2401@gmail.com, [3]shwetapari@gmail.com,
[4]priya19feb@gmail.com, [5]rishirthd@gmail.com

Abstract – Cloud computing is a computing terminology based on utility and consumption of computing resources. It involves software networks and remote servers that allow centralized data storage and access to computer resources or services. In cloud computing we need to maintain data efficiently, hence there is a need of data recovery services. To achieve this we use seed block algorithm (SBA). The objective of proposed algorithm being, first to help the users to collect information from any remote location in the absence of network connectivity, second is to store data in compressed form and saving storage space, and third to recover the files in case of the file deletion or if the cloud gets destroyed due to any reason. The time related issues are also being solved by proposed SBA. Such that it will take minimum time for the recovery process. This algorithm also focuses on the security concept for the back-up files stored at remote server. we are also implementing compression algorithm to overcome the problem of storage size.

Keywords—Seed Block Algorithm, Remote Data Backup server (key words)

I. INTRODUCTION

The data files or information regarding clients which is stored in computer or laptop is lost due to hardware problem like if the system gets physically crashed or data gets corrupted, we are left with no other source to recover it. It is a very tedious job to manage various client records since work is done manually and there are a lot of chances made that create errors in maintaining the user account. Also there is a large data storage problem in the centralized system. Hence at times the data is lost from main server and there is no other backup facility to restore this data. This application provides a feasible solution that collects data and sends it to a centralized storage location smartly and we can access the data remotely.

Cloud computing is defined as a model for enabling convenient, on-demand network access to a shared pool of configurable computing service that can be provisioned rapidly and released with minimal management effort or services provider. In today's world, Cloud Computing itself is a gigantic technology which is surpassing all the previous technologies of computing in this competitive and challenging IT world. Its advantages overcome the disadvantage of various early computing techniques, increasing its demand day by day. Cloud storage provides online storage where data stored in form of virtualized pool that is usually hosted by third parties. The hosting company operates a large data on a large data centre. According to the requirements of the customer this data centre virtualizes the resources and exposes them as the storage pools that helps user to store files or data objects. As a number of users share the storage and other resources, it is possible that other customers might be able to access your data. Either the human error, faulty equipment's, network connectivity, a bug or any criminal intent may put our cloud storage at risk. Also changes in the cloud are also made very frequently; we can term it as data dynamics. The data dynamics is supported by various operations such as insertion, deletion and block modification. Since services are not

limited for archiving and taking backup of data, remote data integrity is also needed. The data integrity always focuses on the validity and fidelity of the complete state of the server that takes care of the heavily generated data which remains unchanged during storing at main cloud remote server and transmission. Integrity plays an important role in back-up and recovery services. However, various successful techniques are lagging behind some critical issues like cost, implementation complexity, security and time consumption. To cater these issues, we propose a smart remote data backup algorithm.

II. RELATED WORK

The recent back-up and recovery techniques that have been developed in cloud computing domain are HSDRT [2], (PCS) [3] Parity Cloud Service, (ERGOT) [4] Efficient Routing Grounded on Taxonomy, Linux Box [5], Cold/Hot backup strategy etc. Survey shows that none of these techniques are able to provide us with best performances under all uncontrolled circumstances such as cost, security, implementation complexity, redundancy and recovery in short span of time.

There are many techniques like HSDRT with many advantages like low cost privacy it has some disadvantage like implementation complexity was very high. Another technique is ERGOT which performs exact match retrieval but problem with this system is time complexity, implementation complexity. Linux box is also one of the technique to get the data backup but it requires Required higher bandwidth, Privacy is not provided, it takes Complete server backup at a time. in Cold/Hot Back-up Strategy there is problem when Cost increases as data increases gradually so to overcome all such problems we have introduced another technique using Seed Block Algorithm(SBA) for data recovery.

PCS has proved to be more convenient for data recovery based on parity recovery services. It is simple, more reliable and easy to use but is unable to control implementation complexities. On the other hand HSDRT has come out an efficacious technique for movable clients like laptops. But implementation of the recovery cost is high, also fails to control data duplication.

ERGOT is unable to focus on time and implementation complexity. However, Linux Box model provides data backup and recovery at a comparatively low cost but its protection level is very low.

The table below shows the comparison of these techniques,

III. NEED FOR BACK-UP IN CLOUD COMPUTING

Cloud computing provides on demand resources to the consumer/user. It requires management of resources among every client/user accounts. Such management includes various aspects of proper utilization of the resources. The resources can be any hardware or software, software like any application development kit, application programming interface or a type of data file. There are various choices among various implementations for back up of the data that maintain its security among various users. Cloud computing must be able to provide reliability such that users can upload their sensitive and important data which is kept secure. Also, implementation of this system over a large scale leads to an enormous amount of data being stored on the server. To deal with this efficiently, the data needs to be stored in compressed for so as to save storage space. The cost-effective approach is the main concern while implementing any cloud.

During the study of cloud computing, we found various advantages of cloud computing. It is found that the cloud is capable enough to store huge amount of data of various different clients with complete security. This enables the Internet Service Provider (ISP) to provide a huge storage in a cloud to the user. The users are allowed to upload their private and important data to the main cloud. At the same time we found a critical issue regarding storage i.e. if any of the client's data file is missing or disappears for some reason or the cloud gets destroyed either due to any natural calamity (like flood, earthquake etc.), then for back-up and recovery consumer/client has to depend on service provider which means the data has to be stored in the server.

To overcome this problem, it requires an efficient technique for data backup and recovery so that the client is able to contact the backup server where private data is stored with high reliability whenever a main cloud fails to provide the user's data. These techniques must be provided at low cost as well as for

implementation of the recovery problem's solution, it should be easy to recover the data after any disaster. Due to heavy storage of the clients, the need of back-up and recovery techniques in cloud computing arises.

A. Problem Statement

In cloud computing, to maintain the data efficiently, there is a necessity of data recovery services. To cater this, we propose a smart remote data backup algorithm, Seed Block Algorithm. Using Seed block algorithm we recover the files in case of the file deletion or if the cloud gets destroyed due to any reason. The time related issues are also being solved by proposed Seed Block Algorithm such that it will take minimum time for the recovery process. Proposed algorithm also focuses on the security concept for the back-up files stored at remote server, without using any of the existing encryption techniques.

The objective of proposed algorithm to help the users to collect information from any remote location, to store data in the cloud in compressed form to save space and to recover the files in case of the file deletion or if the cloud gets destroyed due to any reason. Usually Backup server of main cloud is the copy of main cloud. When this Backup Server is at remote location (i.e. far away from the main server) and has the complete state of the main cloud, this remote location server is termed as Remote Data Backup Server. The main cloud is termed as the central repository and remote backup cloud is termed as remote repository. To tackle the challenges like low implementation complexity, low cost, security and time related issues we propose Seed Block Algorithm.

IV. REMOTE DATA BACK-UP SERVERS

Remote Data Backup server is a server which stores the main cloud's entire data as a whole and located at remote place (far away from cloud). There are an anormous number of users using the storage space. Hence, we compress the data when it is being stored in the backup server. This helps us use the least amount of storage space. If the central repository loses its data, it uses the information from the remote repository. The purpose is to help clients to collect information from remote repository if the main cloud is unable to provide the data to the clients. As shown in Fig 1, if clients found that data is not available on central repository, then clients are allowed to access the files from remote repository (i.e. indirectly).

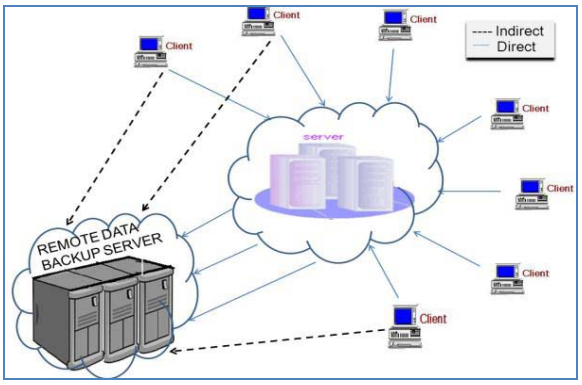


Fig. 1. Remote data Backup Server and its Architecture

V. SBA - ALGORITHM IMPLEMENTATION

The proposed SBA algorithm is as follows:
Algorithm 1:
The proposed SBA algorithm is as follows:
Algorithm 1: Initialization: Main Cloud: Cm: Remote Server: Rs;
Clients of Main Cloud: Cn; Files: f1 and f1';
Seed block: Sb; Random Number:
Rn ; Client ID:Cilent-Idi
Input: f1 created by Cn; Rn is generated at Cm;
Output: Recovered file f1 after deletion at Cm
Given: Authenticated clients could allow uploading, downloading and do mod-ification on its own the files only.
Step 1: Generate a random number.
Int Rn=rand ();
Step 2: Create a seed Block Sb for each Cn and Store
Sb at Rs
Sb = Rn EXOR Cilent-Idi (Repeat step 2 for all clients)
Step 3: If Cn/Admin creates/modifies f1 and stores at
Cm, then create as f1'
f1'= f1 EXOR Sb
Step 4: Store a at Rs
Step 5: If server crashes f1 deleted from Cm
Then we do EXOR to retrieve the original f1 as:
f1= f1' EXOR Sb
Step 6: Return f1 to Cn
Step 7: END

Since a backup system contains at least one copy of all data worth saving, the data storage requirements can be significant. Organizing this storage space and managing the backup process can be a complicated undertaking. Hence we compress the data that is being stored in the backup server saving up most of the space. A data repository model can be used to provide structure to the storage. Nowadays, there are many different types of data storage devices that are useful for making backups. The Remote backup services should cover the following issues:

- Privacy and ownership.
- Relocation of servers to the cloud
- Data security
- Reliability
- Storage space management
- Cost effectiveness
- Appropriate Timing

Privacy and ownership:
Different clients access the cloud with their different login or after any authentication process. They are freely allowed to upload their private and essential data on the cloud. Hence, the privacy and ownership of data should be maintained; Owner of the data should only be able to access his private data and perform read, write or any other operation. Remote Server must maintain this Privacy and ownership.

Relocation of server:
For data recovery there must be relocation of server to the cloud. The Relocation of server means to transfer main server's data to another server; however the new of location is unknown to the client. The clients get the data in same way as before without any intimation of relocation of main server, such that it provides the location transparency of relocated server to the clients and other third party while data is been shifted to remote server.

Data security:
The client's data is stored at central repository with complete protection. Such a security should be followed in its remote repository as well. In remote repository, the data should be fully protected such that no access and harm can be made to the remote cloud's data either intentionally or unintentionally by third party or any other client.

Reliability:
The remote cloud must possess the reliability characteristics. Because in cloud computing the main cloud stores the complete data and each client is dependent on the main cloud for each and every little amount of data; therefore the cloud and remote backup cloud must play a trustworthy role. That means, both the server must be able to provide the data to the client immediately whenever they required either from main cloud or remote server.

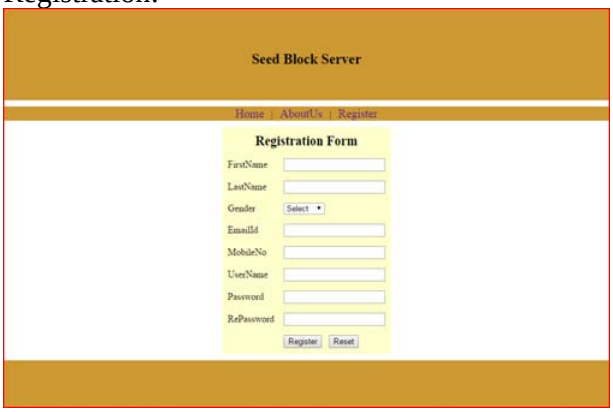
Storage space management:
The data a client/user uploads gets stored on the cloud. The number of users using the cloud is large, so is the amount of data being stored in the cloud. To use the resources effectively, we compress this data at the storage. This saves up a lot of space for other data to be stored.

Cost effectiveness:
The cost for implementation of remote server and its recovery & back-up technique also play an important role while creating the structure for main cloud and its correspondent remote cloud. The cost for establishing the remote setup and for implementing its technique must be minimum such that small business can afford such system and large business can spend minimum cost as possible.

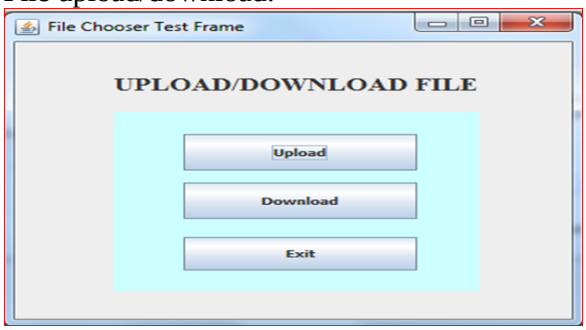
Appropriate Timing:
The process of data recovery takes some time for retrieval of data from remote repository as this remote repository is far away from the main cloud and its clients. Therefore, the time taken for such a retrieval must be minimum as possible such that the client can get the data as soon as possible without concerning the fact that remote repository is how far away from the client. There are many techniques that have focused on these issues. In forthcoming section, we will be discussing some of recent techniques of back-up and recovery in cloud computing domain.

VI. RESULTS

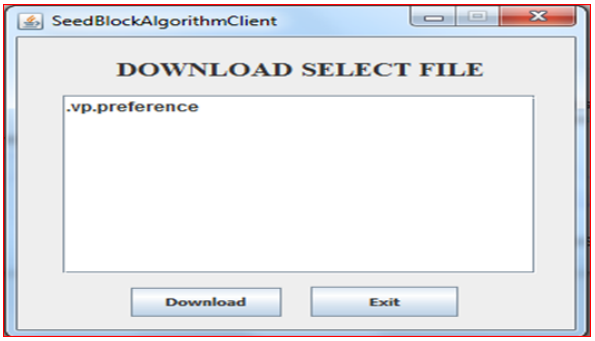
Registration:



File upload/download:



Download file selection:



CONCLUTION

This implementation paper proposes a technique to perform a smart remote data backup using SEED Block Algorithm. The Seed Block Algorithm is time efficient technique to recover lost files. It maintains the data integrity and solves the issues like cost, implementation complexity, data storage. SBA also focuses on the security concept for the back-up files stored at remote backup server.

REFERENCES

[1] 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,2013

[2] Yoichiro Ueno, Noriharu Miyaho, Shuichi Suzuki,Muzai Gakuendai, Inzai-shi, Chiba,Kazuo Ichihara, "Performance Evaluation of a Disaster Recovery System and Practical Network System Applications," Fifth International Conference on Systems and Networks Communications, pp 256-259,2010.

[3] Chi-won Song, Sungmin Park, Dong-wook Kim, Sooyong Kang, "Parity Cloud Service: A Privacy-Protected Personal Data Recovery Service," International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11, 2011.

[4] Y.Ueno, N.Miyaho, and S.Suzuki, "Disaster Recovery Mechanism using Widely Distributed Networking and Secure Metadata Handling Technology", Proceedings of the 4th edition of the UPGRADE-CN workshop, 2009, pp. 45-48.

[5] Vijaykumar Javaraiah Brocade Advanced Networks and Telecommunication Systems (ANTS), "Backup for Cloud and Disaster Recovery for Consumers and SMBs," IEEE 5th International Conference, 2011.

[6] 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.

[7] Lili Sun, Jianwei An, Yang Yang, Ming Zeng, "Recovery Strategies for Service Composition in Dynamic Network," International Conference on Cloud and Service Computing, 2011.

[8] Xi Zhou, Junshuai Shi, Yingxiao Xu, Yinsheng Li and Weiwei Sun, 2008, "A backup restoration algorithm of service composition in MANETs," Communication Technology ICCT 11th IEEE International Conference, pp. 588-591.

[9] M. Armbrust et al, "Above the clouds: A Berkeley view of cloud computing," http://www.eecs.berkeley.edu/,2009.

[10] F.BKashani, C.Chen, C.Shahabi.WSPDS, 2004, “Web Services Peer-to-Peer Discovery Service,” ICOMP.

[10] Kalyani Bangale1, Nivedita Gupta, 2014, “Remote Data Collection Server: E-Health Care”, IEEE ICC.

[11] Balazs Gerofi, Zoltan Vass and Yutaka Ishikawa, “Utilizing Memory Content Similarity for Improving the Performance of Replicated Virtual Machines”, Fourth IEEE International Conference on Utility and Cloud Computing2011.

[12] P.Demeester et al., 1999. , “Resilience in Multilayer Networks,” IEEE Communications Magazine, Vol. 37, No. 8, p.70-76. S. Zhang, X. Chen, and X. Huo, 2010, “Cloud Computing Research and Development Trend,” IEEE Second International Conference on Future Networks, pp. 93-97.

[13] T. M. Coughlin and S. L. Linfoot, 2010, “A Novel Taxonomy for Consumer Metadata,” IEEE ICCE Conference.

[14] K. Keahey, M. Tsugawa, A. Matsunaga, J. Fortes, 2009, “Sky Computing”, IEEE Journal of Internet Computing, vol. 13, pp. 43-51.

[15] M. D. Assuncao, A.Costanzo and R. Buyya, 2009 , “Evaluating the Cost- Benefit of Using Cloud Computing to Extend the Capacity of Clusters,” Proceedings of the 18th International Symposium on High Performance Distributed Computing (HPDC 2009), Munich, Germany.

[16] Sheheryar Malik, Fabrice Huet, December 2011, “Virtual Cloud: Rent Out the Rented Resources," 6th International Conference on Internet Technology and Secure Transactions,11-14 ,Abu Dhabi, United Arab Emirates.