Volume 5, Issue 2, February 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Disaster Recovery System Using Seed Block Algorithm in Cloud Computing Environment

S. Deepa, Dr. G. Ramachandran
Dept of Computer Science and Engineering
Annamalai University, Annamalai Nagar
Chidambaram, Tamil Nadu, India

Abstract*— *In cloud environment, large amount of data stored in the server. To maintain the efficiency of those data needs recovery services. In cloud computing large amount of private data stored on the main cloud. Therefore the necessity of the recovery services growing day- by -day and it requires a development of an efficient and effective data recovery service technique. The purpose of the recovery technique is to help user to collect information from any backup server when the server fails to provide the data to the user. There are lots of recovery mechanisms are used to recover the data in the cloud such as HSDRT, ERGOT, LINUX BOX, PCS, COLD and COLD/HOT backup strategy. But there are some limitations in those techniques such as implementation complexity, security issues and retrieval time is high. Hence, propose a smart data backup algorithm called seed block algorithm. The objective of proposed SBA is to recover the files in case of cloud get destroyed or file may be deleted from the cloud. The major advantage of SBA is to take minimum time for the recovery process.

Keywords-Main cloud; Backup; Parity Cloud Service; Seed Block; Complexity; HSDRT;

I. INTRODUCTION

Cloud computing provides on demand resources to the consumer/user. It requires the management of resources among each and every client/user. Such management includes various aspects of proper utilization of the resources. The resources can be any hardware or software. The software like any application programming interface, application development kit and any type of data file etc. Various choices are there among various implementations for back up of the data and 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. 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. In advantages, we found that the cloud is capable enough to store the huge amount of data of various different clients with complete security such that Internet Service Provider (ISP) provides a huge storage in a cloud to the user. And users are allow to upload there private and important data to the main cloud. And at the same time we found critical issue regarding this storage i.e. if any of the client's data file is missing or disappeared for some reason or the cloud get 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 problem of. such scenario, it requires an efficient technique for data backup and recovery so that the client can able to contact the backup server where private data is stored with high reliability and whenever a main cloud fails to provide the user's data. These techniques must possess lower cost as well for implementation of the recovery problem's solution and can easily recover the data after any disaster. That's why, the need of the backup and recovery techniques for cloud computing arises due to heavy storage of its clients.

A number of user shares the storage and other resources, it is possible that other customers can access your data.Either the human error, faulty equipment's, network connectivity, a bug or any criminal intent may put our cloud storage on the risk and danger. And 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 to, archiving and taking backup of data; remote data integrity is also needed. Because 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.

This paper is organized as follows: Section II focuses on the related literature of existing methods that are successful to some extent in the cloud computing domain. In Section III, we discuss about the remote data backup server. Section IV describes the detailed description of the proposed seed block algorithm (SBA) and Section V shows the results and experimentation analysis of the proposed SBA. Finally, in Section VI conclusions are given.

II. REMOTE DATA BACKUP SERVER

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). And if the central repository lost its data, then it uses the information from the remote repository.

The purpose is to help clients to collect information from remote repository either if network connectivity is not available or 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).

When we talk about Backup server of main cloud, we only think about the copy of main cloud. When this Backup server is at remote location (i.e. far away from the main server) and having the complete state of the main cloud, then 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.

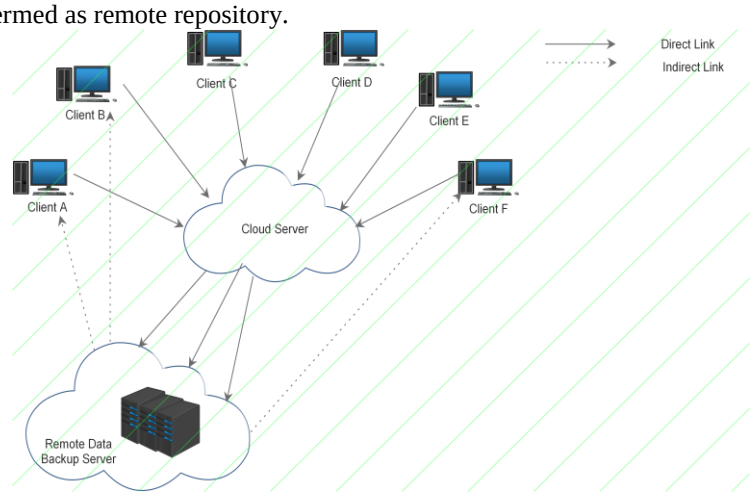

Fig.1.Remote Databackup Server Architecture

The architecture of remote data backup server is shown in Fig.1.It contains various clients, repository (web service), main database, users and architecture is explained as follows. The client application can be ported to any other machine like laptop or handheld devices. The stored data is platform independent that are sent to a central repository. When connected to network, the client application is authenticate into a central repository using a web service and submit all collected information. And if the central repository lost its data under any circumstances either of any natural calamity (for ex - earthquake, flood, fire etc.) or by human attack or deletion that has been done mistakenly and then it uses the information from the remote repository. The main objective of the remote backup facility is to help user to collect information from any remote location even if network connectivity is not available or if data not found on main cloud. As shown in Fig-1 clients are allowed to access the files from remote repository if the data is not found on central repository (i.e. indirectly).

The Remote backup services should cover the following issues:

1. Data Integrity: Data Integrity is concerned with complete state and the whole structure of the server. It verifies that data such that it remains unaltered during transmission and reception. It is the measure of the validity and fidelity of the data present in the server.

2. Data Security: Giving full protection to the client's data is also the utmost priority for the remote server. And either intentionally or unintentionally, it should be not able to access by third party or any other users/client's.

3. Data Confidentiality: Sometimes client's data files should be kept confidential such that if no. of users simultaneously accessing the cloud, then data files that are personal to only particular client must be able to hide from other clients on the cloud during accessing of file.

4. Trustworthiness: The remote cloud must possess the Trustworthiness characteristic. Because the user/client stores their private data; therefore the cloud and remote backup cloud must plays a trustworthy role.

5. Cost Efficiency: The cost of process of data recovery should be efficient so that maximum no. of company/clients can take advantage of back-up and recovery service.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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.

III. DESIGN OF THE PROPOSED SEED BLOCK ALGORITHM

As discussed in literature, many techniques have been proposed for recovery and backup such as HSDRT[1], PCS[2], ERGOT[4], Linux Box[5], Cold/Hot backup strategy[6] etc. As discussed above low implementation complexity, low cost, security and time related issues are still challenging in the field of cloud computing. To tackle these issues we propose SBA algorithm and in forthcoming section, we will discuss the design of proposed SBA in detail.

A. Seed Block Algorithm (SBA) Architechture

This algorithm focuses on simplicity of the back-up and recovery process. It basically uses the concept of Exclusive–OR (XOR) operation of the computing world. For ex: - Suppose there are two data files: A and B. When we XOR A and B it produced X i.e. X =$A \oplus B$. If suppose A data file get destroyed and we want our A data file back then we are able to get A data file back, then it is very easy to get back it with the help of B and X data file .i.e. A = $X \oplus B$.

Similarly, the Seed Block Algorithm works to provide the simple Back-up and recovery process. Its architecture is shown in Fig-2 consists of the Main Cloud and its clients and the Remote Server. Here, first we set a random number in the cloud and unique client id for every client. Second, whenever the client id is being register in the main cloud; then client id and random number is getting EXORed ($\oplus$)with each other to generate seed block for the particular client. The generated seed block corresponds to each client is stored at remote server.

Whenever client creates the file in cloud first time, it is stored at the main cloud. When it is stored in main server, the main file of client is being EXORed with the Seed Block of the particular client. And that EXORed file is stored at the remote server in the form of file' (pronounced as File dash). If either unfortunately file in main cloud crashed / damaged or file is been deleted mistakenly, then the user will get the original file by EXORing file' with the seed block of the corresponding client to produce the original file and return the resulted file i.e. original file back to the requested client. The architecture representation of the Seed Block Algorithm is shown in the Fig.2.

B. SBA Algorithm

The proposed SBA algorithm is as follows:

Initialization: Main Cloud: M_c ; Remote Server: R_s

Clients of Main Cloud: C_i ; Files: $a_1$and $a_1^{'}$

Seed Block: S_i ; Random Number : r;

Client's ID: $Client\_Id_i$

Input: a_1 created by C_i ; r is generated at M_c;

Output: Recovered File a_1 after deletion atM_c;

Given: Authenticated Clients could allow uploading, downloading and do modification on its own the files only.

Step 1: Generate a random number.

int r= rand( );

Step 2: Create a Seed Block S_i for each C_i and store S_i at R_s,

S_i= r$\oplus$ $Client\_Id_i$ (Repeat step 2 for all clients)

Step 3: If C_i/ Admin Creates /modifies a $a_1$and Stores at M_c , then $a_1^{'}$ create $a_1^{'}$=$a_1 \oplus S_i$

Step 4: Store $a_1^{'}$ at R_s

Step 5: If Server crashes a_1 deleted from M_c, then we do EXOR to retrieve the original a_1 as: $a_1 = a_1^{'} \oplus S_i$

Step 6: Return a_1 to C_i

Step 7: END

Fig.2 Seed Block Algorithm Architecture

IV. EXPERIMENTATION AND RESULT ANALYSIS

In this section, we discuss the experimentation and result analysis of the SBA algorithm. During experimentation, we found that size of original data file stored at main cloud is exactly similar to the size of Back-up file stored at Remote Server as depicted in Table-I. In order to make this fact plausible, we perform this experiment for different types of files. Results tabulated in Table-I for this experiment shows that proposed SBA is very much robust in maintaining the size of recovery file same as that the original data file. From this we conclude that proposed SBA recover the data file without any data loss.

Table-I: Performance analysis for different types of files

Type	Size 0f Original File in Main Server	Size of Backup File in Remote Server	Size Of the Recovered File
Text(txt/.doc/.docx/.xl/.pdf)	512 KB	512 KB	512 KB
	5.2 MB	5.2 MB	5.2 MB
Images(.jpeg/.gif/.png/.bitmap)	70 KB	70 KB	70 KB
	8 MB	8 MB	8 MB

Processing Time means time taken by the process when client uploads a file at main cloud and that includes the assembling of data such as the random number from main cloud, seed block of the corresponding client from the remote server for EXORing operation; after assembling, performing the EXORed operation of the contents of the uploaded file with the seed block and finally stored the EXORed file onto the remote server. Performance of this experiment is tabulated in Table-II. We also observed that as data size increases, the processing time increases. On other hand, we also found that performance which is megabyte per sec (MB/sec) being constant at some level even if the data size increases as shown in Table-II

Table-II Effect of data size on processing time

Data Size[GB]	Main Cloud Processing Time(in sec)(Approx).	Remote Cloud processing Time(in sec)(Approx).	Performance(MB/sec)
1	6.76	2	151
2	12.8	3	160
4	25	5	164
8	49.3	8	166
12	73.9	15	166
16	97.9	35	167
24	146	45	168
32	195	63	168
48	292	73	168
64	390	80	168

In this paper, we presented detail review of most recent back-up and recovery techniques that have been developed in cloud computing domain. Detail review of this paper shows that these techniques have its own advantages and disadvantages which are summarized in the Table-1. All these approaches are able to provide best performances under all uncontrolled circumstances such as cost, security, low implementation complexity, redundancy and recovery in short span of time. Among all the techniques reviewed PCS is comparatively reliable; maintain its privacy for each resource and also it try to minimize the cost of infrastructure. However, it is unable to control the implementation complexities. On the contrary, HSDRT has come out an efficient technique for the movable clients such as laptop, smart phones etc. nevertheless it fails to manage the low cost for the implementation of the recovery and also unable to control the data duplication. Rather, ERGOT is totally based on the semantic analysis and unable to focus on time and implementation complexity. In addition, Linux Box model is having very simple concept of data back-up and recovery with very low cost. However, in this model protection level is very low. Similarly, in the list of techniques maintaining the cost of implementation, SBBR focuses on the cost reduction; however fails to concentrate on the optimization concept and redundancy. With entirely new concept of virtualization REN cloud also focuses on the low cost infrastructure with the complex implementation and low security level. All these techniques tried to cover different issues maintaining the cost of implementation as low as possible. However there are some techniques in which cost increases gradually as data increases. For example, Cold and Hot back-up strategy that performs backup and recovery on trigger basis of failure detection

Although each one of the backup solution in cloud computing is unable to achieve all the aforesaid issues of remote data back-up server. Therefore, due to the high applicability of backup process in the companies, the role of a remote data back –up server is very crucial and hot research topic.

The Fig-3 shows the CPU utilization at Main Cloud and Remote Server. As shown in Fig-3 the Main Cloud's CPU utilization starts with 0% and as per the client uploads the file onto it then utilization increases; such that it has to check whether the client is authenticated or not, at the same the time it send request to Remote Server for the corresponding Seed Block. When request reached to Remote Server it started collecting the details as well as the seed Block and gives response in form of the seed Block and during this period, load at Main Cloud decreases which in return cause for gradual decreases in CPU utilization at main cloud. After receiving the requested data, CPU utilization at main cloud increases as it has to perform the EXORed operation. Again the Final EXORed file sends to Remote Server. As compared to Table-IV the processing time given can be compare with the time showing in Fig-3.

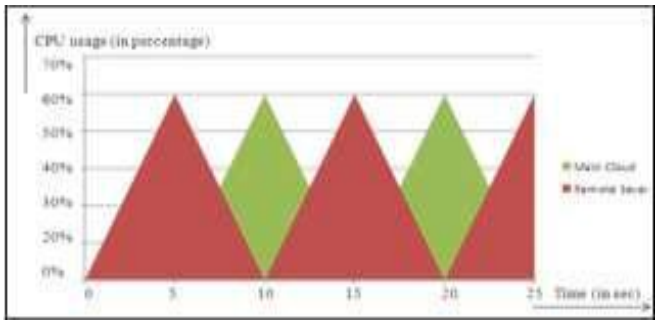

Fig.3 Graph Showing Processor Utilization

The Fig-4 shows the experimentation result of proposed SBA. As fig-4 (a) shows the original file which is uploaded by the client on main cloud. Fig-4 (b) shows the EXORed file which is stored on the remote server. This file contains the secured EXORed content of the original file and seed block content of the corresponding client. Fig-4 (c) shows the recovered file; which indirectly sent to clients in the absence of network connectivity and in case of the file deletion or if the cloud gets destroyed due to any reason.

A) Original File B)XORed File C)Recovered File

Fig.4 Sample output image of SBA Algorithm

VI. CONCLUSION

In this paper, we proposed a remote data backup algorithm called seed block algorithm which helps the user to recover the disaster files from the remote location when the main cloud fails to fetch the files to the client. Experimentation and results shows that there is no modification can be done in the original file so the integrity of the file should be maintained and the time related issues also being solved by the proposed SBA so, it took minimum time to recover the files from remote server.

REFERENCES

[1] 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.

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

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

[4] 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 Computing.

[5] Paris Kitsos and Athanassios N. Skodras, 2009, "An FPGA Implementation and Performance Evaluation of the Seed Block Cipher".

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

[7] Joon-Ho Hwang, 2004, "Efficient Hardware of SEED S-box for smart cards", Journal of semiconductor technology and science, Vol.4,No.4,December,2004.

[8] Timothy wood, Emmanuel Cecchet, K.K.Ramakrishnan, Prashant Shenoy, Jacobus van der Merwe, and Arun Venkatramani, 2011, "Disaster as a Cloud service: Economic Benefits & Deployment Challenges".

[9] Maulik Dave,2013, "Data Storage Security in cloud Computing :A Survey," ijarcsse volume 3,Issue 10, October 2013.

[10] Eleni Palkopoulouy, Dominic A. Schupke, Thomas Bauscherty, 2011, "Recovery Time Analysis for the Shared Backup Router Resources (SBRR) Architecture", IEEE ICC.

[11] Lili Sun, Jianwei An, Yang Yang, Ming Zeng ,2011,"Recovery Strategies for Service Composition in Dynamic Network," International Conference on Cloud and Service Computing, pp. 221–226.

[12] P.Demeester et al., 1999, "Resilience in Multilayer Networks," IEEE Communications Magazine, Vol. 37, No. 8, p.70-76.

[13] S. Zhang, X. Chen, and X. Huo, 2010, "Cloud Computing Research and Development Trend," IEEE Second International Conference on Future Networks, pp. 93-97.

[14] T. M. Coughlin and S. L. Linfoot, 2010, "A Novel taxonomy for Consumer Metadata," IEEE ICCE Conference.

[15] Kazuo Ichihara ,Noriharu Miyaho,Yoichiro Ueno,2013,"Optimized Implementation of the Mutual Authentication Storage System Using HS-DRT".

[16] 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), Germany.

[17] 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.

[18] Wayne A. Jansen, 2011, "Cloud Hooks: Security and Privacy Issues in Cloud Computing, 44th Hawaii International Conference on System Sciences.Hawaii.

[19] Jinpeng et al, 2009, "Managing Security of Virtual Machine Images in a Cloud Environment", CCSW, Chicago, USA.

[20] Ms..Kruti Sharma,Prof K.R.Singh, 2012, "Online data Backup And Disaster Recovery techniques in cloud computing :a review", IJEIT,vol.2,Issue 5.

