2013 International Conference on Communication Systems and Network Technologies

Seed Block Algorithm: A Remote Smart Data Back-up Technique for Cloud Computing

Ms. Kruti Sharma
Department of Computer Technology,
YCCE, Nagpur (M.S),
441 110, India
kruti.sharma1989@gmail.com

Prof. Kavita R Singh
Department of Computer Technology,
Nagpur, YCCE, Nagpur (M.S),
441 110, India
singhkavita19@yahoo.co.in

***Abstract*— In cloud computing, data generated in electronic form are large in amount. To maintain this data efficiently, there is a necessity of data recovery services. To cater this, in this paper we propose a smart remote data backup algorithm, Seed Block Algorithm (SBA). The objective of proposed algorithm is twofold; first it help the users to collect information from any remote location in the absence of network connectivity and second 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. Proposed SBA also focuses on the security concept for the back-up files stored at remote server, without using any of the existing encryption techniques.**

Keywords-Central Repository; Remote Repository; Parity Cloud Service; Seed Block;

I. INTRODUCTION

National Institute of Standard and Technology defines as a model for enabling convenient, on-demand network access to a share pool of configurable computing service (for ex-networks, servers, storage, applications and services) that can be provisioned rapidly and released with minimal management effort or services provider [1]. Today, Cloud Computing is itself a gigantic technology which is surpassing all the previous technology of computing (like cluster, grid, distributed etc.) of this competitive and challenging IT world. The need of cloud computing is increasing day by day as its advantages overcome the disadvantage of various early computing techniques. Cloud storage provides online storage where data stored in form of virtualized pool that is usually hosted by third parties. The hosting company operates large data on large data center and according to the requirements of the customer these data center virtualized the resources and expose them as the storage pools that help user to store files or data objects.

As 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 for 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.

In literature many techniques have been proposed HSDRT[1], PCS[2], ERGOT[4], Linux Box [5], Cold/Hot backup strategy [6] etc. that, discussed the data recovery process. However, still various successful techniques are lagging behind some critical issues like implementation complexity, low cost, security and time related issues. To cater this issues, in this paper we propose a smart remote data backup algorithm, Seed Block Algorithm (SBA). The contribution of the proposed SBA is twofold; first SBA helps the users to collect information from any remote location in the absence of network connectivity and second to recover the files in case of the file deletion or if the cloud gets destroyed due to any reason.

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. RELATED LITERATURE

In literature, we study most of the recent back-up and recovery techniques that have been developed in cloud computing domain such as HSDRT[1], PCS[2], ERGOT[4], Linux Box [5], Cold/Hot backup strategy [6] etc. Detail review shows that none of these techniques 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, simple, easy to use and more convenient for data recovery totally based on parity recovery service. It can recover data with very high probability. For data recovery, it generates a virtual disk in user system for data backup, make parity groups across virtual disk, and store parity data of parity group in cloud. It uses the Exclusive–OR ($\oplus$) for creating Parity information. 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. It an innovative file back-up concept, which makes use of an effective ultra-widely distributed data transfer mechanism and a high-speed encryption technology

The HS-DRT [1] is an innovative file back-up concept, which makes use of an effective ultra-widely distributed data transfer mechanism and a high-speed encryption technology. This proposed system follows two sequences one is Backup sequence and second is Recovery sequence. In Backup sequence, it receives the data to be backed-up and in Recovery Sequence, when some disasters occurs or periodically, the Supervisory Server (one of the components of the HSDRT) starts the recovery sequence. However there are some limitation in this model and therefore, this model is somehow unable to declare as perfect solution for back-up and recovery.

Rather, Efficient Routing Grounded on Taxonomy (ERGOT) [4] is totally based on the semantic analysis and unable to focus on time and implementation complexity. It is a Semantic-based System which helps for Service Discovery in cloud computing. Similarly, we found a unique way of data retrieval. We made a focus on this technique as it is not a back-up technique but it provide an efficient retrieval of data that is completely based on the semantic similarity between service descriptions and service requests. ERGOT is built upon 3 components 1) A DHT (Distributed Hash Table) protocol 2) A SON (Semantic Overlay Network), 3) A measure of semantic similarity among service description [4]. Hence, ERGOT combines both these network Concept. By building a SON over a DHT, ERGOT proposed semantic-driven query answering in DHT-based systems. However does not go well with semantic similarity search models.

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. It also makes the process of migration from one cloud service provider to other very easy. It is affordable to all consumers and Small and Medium Business (SMB). This solution eliminates consumer's dependency on the ISP and its associated backup cost. It can do all these at little cost named as simple Linux box which will sync up the data at block/file level from the cloud service provider to the consumer. It incorporates an application on Linux box that will perform backup of the cloud onto local drives. The data transmission will be secure and encrypted. The limitation we found that a consumer can backup not only the Data but Sync the entire Virtual Machine[5] which somehow waste the bandwidth because every time when backup takes place it will do back-up of entire virtual machine.

Similarly, we also found that one technique basically focuses on the significant cost reduction and router failure scenario i.e. (SBBR). It concerns IP logical connectivity that will be remain unchanged even after a router failure and the most important factor is that it provides the network management system via multi-layer signaling.

Table-I Comparison between various techniques of Back-up and recovery [20]

S.No	Approach	Advantage	Disadvantage
1	HSDRT[1]	• Used for Movable clients like laptop, Smart Phone	• Costly • Increase redundancy
2	Parity Cloud Service[2]	• Reliable • Privacy • Low cost	• Implementation • complexity is high
3	ERGOT[4]	• Perform exact-match retrieval • Privacy	• Time complexity • Implementation complexity
4	Linux Box[5]	• Simple • Low cost for implementation	• Required higher bandwidth • Privacy • Complete server Backup at a time
5	Cold /Hot Back-up Strategy[6]	• Triggered only when failure detected	• Cost increases as data increases gradually
6	Shared backup router resources(SBBR)[10]	• It concerns with cost reduction • Works even if router fails	• Inconsistencies between logical and physical configurations may lead to some performance problem • It is unable to includes optimization concept with cost reduction
7	Rent Out the Rented Resources [17]	• Virtualization, rents it to the clients in form of cloud services • Cost depends on the infrastructure utilization	• Implementation get Complex • Resources must kept under special attention due to rented concept

Additionally [10], it shows how service imposed maximum outage requirements that have a direct effect on the setting of the SBRR architecture (e.g. imposing a minimum number of network-wide shared router resources locations). However, it is unable to include optimization concept with cost reduction.

With entirely new concept of virtualization REN cloud focuses on the low cost infrastructure with the complex implementation and low security level. Another technique we found in the field of the data backup is a REN (Research Education Network) cloud. The lowest cost point of view we found a model "Rent Out the Rented Resources" [17]. Its goal is to reduce the cloud service's monetary cost. It proposed a three phase model for cross cloud federation that are discovery, matchmaking and authentication. This model is based on concept of cloud vendor that rent the resources from venture(s) and after virtualization, rents it to the clients in form of cloud services.

All these techniques tried to cover different issues maintaining the cost of implementation as low as possible. However there is also a technique in which cost increases gradually as data increases i.e. Cold and Hot back-up strategy [6] that performs backup and recovery on trigger basis of failure detection. In Cold Backup Service Replacement Strategy (CBSRS) recovery process, it is triggered upon the detection of the service failures and it will not be triggered when the service is available. In Hot

Backup Service Replacement Strategy (HBSRS), a transcendental recovery strategy for service composition in dynamic network is applied [6]. During the implementation of service, the backup services always remain in the activated states, and then the first returned results of services will be adopted to ensure the successful implementation of service composition.

Although each one of the backup solution in cloud computing is unable to achieve all the issues of remote data back-up server. The advantages and disadvantages of all these foresaid techniques are described in the Table-I. And 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.

III. REMOTE DATA BACKUP SERVER

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.

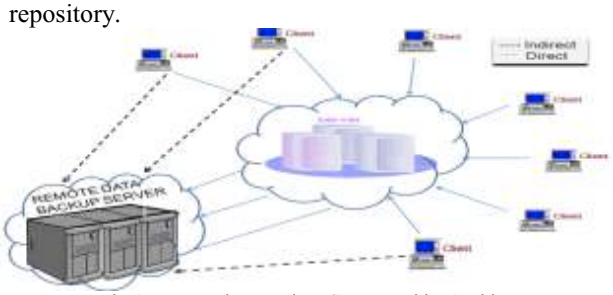

Fig.1 Remote data Backup Server and its Architecture

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 play 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.

There are many techniques that have focused on these issues. In forthcoming section, we will be discussing a technique of back-up and recovery in cloud computing domain that will cover the foresaid issues.

IV. 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:

Algorithm 1:

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 at M_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 as
$a_1' = a_1 \oplus S_i$

Step 4: Store a' 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.

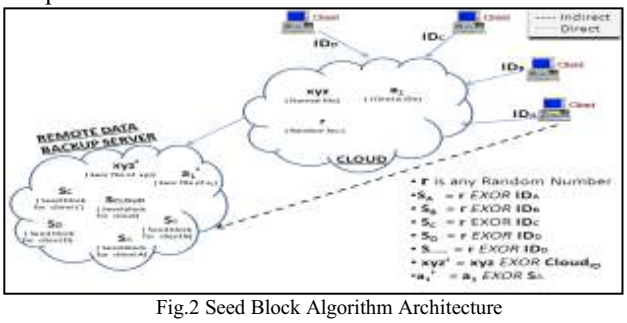

Fig.2 Seed Block Algorithm Architecture

V. Experimentation And Result Analysis

In this section, we discuss the experimentation and result analysis of the SBA algorithm. For experimentation we focused on different minimal system requirement for main cloud's server and remote server as depicted in Table-II. From Table-II, memory requirement is kept 8GB and 12GB for the main cloud's server and remote server respectively, which can be extended as per the necessity. From Table-II, it is observed that memory requirement is more in remote server as compare to the main cloud's server because additional information is placed onto remote server (for example- different Seed Block of the corresponding client shown in Fig-2).

Table-II System Environment

	Main Cloud's server	Remote Cloud Server
CPU	Core2 Quad Q6600 2.40GHz	Core2 Quad Q6600 2.40GHz
Memory	8GB(DDR2-800)	12GB(DDR2-800)
OS	Any Windows/Linux Platform	Any Windows/Linux Platform
HDD	SATA 250GB or more (7200rpm)	SATA 500GB or more (7200rpm)

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-III. In order to make this fact plausible, we perform this experiment for different types of files. Results tabulated in Table-III 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-III: Performance analysis for different types of files

Type	Size Of Original File in main Server	Size Of Back-up File in Remote Server	Size Of Recovered File after recovery Process
Text (.txt/.doc/.docx/.xl/.pdf)	434KB	434KB	434KB
	2.5MB	2.5MB	2.5MB
Image (.jpeg/.gif/.png/.bitmap)	80KB	80KB	80KB
	4MB	4MB	4MB

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-IV. 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-IV.

Table-IV Effect of data size on processing time

Practical Data Size	Processing Time on Main Cloud Time(in sec) (Approx.)	Processing Time on Remote Cloud Time(in sec) (Approx.)	Performance (MB/sec)
1KB	6.76	2	150
64KB	12.8	3	160
2MB	3600	5	164
32 MB	8400	8	250
1GB	16200	15	280
4GB	32100	35	280
6GB	52000	45	280

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.

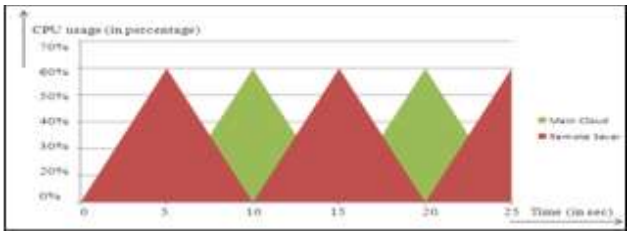

Fig.3 Graph Showing Processor Utilization

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 original file and seed block content of the corresponding client. Fig-4 (c) shows the recovered file; which indirectly sent to client in the absence of network connectivity and in case of the file deletion or if the cloud gets destroyed due to any reason.

Fig.4 Sample output image of SBA Algorithm

VI. CONCLUSION

In this paper, we presented detail design of proposed SBA algorithm. Proposed SBA is robust in helping the users to collect information from any remote location in the absence of network connectivity and also to recover the files in case of the file deletion or if the cloud gets destroyed due to any reason. Experimentation and result analysis shows that proposed SBA also focuses on the security concept for the back-up files stored at remote server, without using any of the existing encryption techniques. The time related issues are also being solved by proposed SBA such that it will take minimum time for the recovery process.

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

[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] 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.

[8] M. Armbrust et al, "Above the clouds: A berkeley view of cloud computing,"http://www.eecs.berkeley.edu/Pubs/TechRpts/2009//EECS-2009-28.pdf.

[9] F.BKashani, C.Chen,C.Shahabi.WSPDS, 2004, "Web Services Peer-to-Peer Discovery Service ," ICOMP.

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

[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.