Available online at www.scientiaresearchlibrary.com

Scientia Research Library

ISSN 2348-0424
USA CODEN: JETRB4

Journal of Engineering And Technology Research, 2015, 3 (2):1-9
(*http://www.scientiaresearchlibrary.com/arhcive.php*)

AN INTELLIGENT DATA BACK-UP AND RETRIEVING TECHNIQUE FOR CLUSTER ENVIRONMENT

S.Agalya[1], S.Bhavithra [2], S.Sherin Benitta[3]

[1]*Department of Computer Science and Engineering, KSK College of engineering & Technology, Kumbakonam,India*

[2]*Department of Computer Science and Engineering, KSK College of engineering & Technology, Kumbakonam,India*

[3]*Department of Computer Science and Engineering, KSK College of engineering & Technology, Kumbakonam,India*

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 files 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 backup files stored at remote server, without using any of the existing encryption techniques.

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

.

INTRODUCTION

National Institute of Standard and Technology defines as 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 manage meant effort or services provider . 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 datacenter 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 areal so 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, PCS, ERGOT, Linux Box , Cold/Hot backup strategy 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 is 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.

MATERIALS AND METHOD

LITERATURE SURVEY

In literature survey, we have studied the most recent back-up and recovery techniques that have been developed in cloud computing domain such as PCS, HSDRT, Linux Box , ERGOT, Cold/Hot backup strategy etc. When we studied the existing methods in detail we found that, performance of the system is not satisfactory with respect to cost, security, low implementation complexity, redundancy and recovery in short span of time. We inferred after study of various present techniques that PCS is comparatively reliable, simple, easy to use and more convenient for data recovery totally based on parity recovery service. It has higher probability and efficiency of recovering among present techniques. 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 to recover the data. It makes use of the Exclusive–OR functionality for creating Parity information. However, there are some problems associated with this method. This method is unable to control the implementation complexities. On the other side, HSDRT method ensures as a powerful technique for the movable clients such as laptop, smart devices, palmtops etc. However it is not economical for the implementation of the recovery and also unable to control the data replication. 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 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 system follows two sequences one is Backup sequence and second is Recovery sequence. In Backup sequence, it accepts the data to be backed-up and in Recovery Sequence, when some calamities occur or periodically, the Supervisory Server starts the recovery sequence. However there are some limitations in this model and therefore, this model somehow fails to declare as perfect solution for back-up and recovery. We also observed that Linux Box model is having very simple concept of data back-up and recovery with very low cost.

But in this model protection level is very low. Process of migration from one cloud service provider to other seems to be very easy. It is economical for all consumers and Small and Medium Business. This solution removes consumer's dependency on the internet service provider and its associated backup cost. It incorporates an application on Linux box that will perform backup of the cloud onto local drives. The data transmission will be secured and encrypted. The limitation we found that a consumer can backup not only the Data but Sync the entire Virtual Machine which somehow waste the bandwidth because every time when backup takes place it will do back-up of entire virtual machine.

Moreover, Efficient Routing Grounded on Taxonomy (ERGOT) features the semantic analysis and fails to focus on time constraints and implementation complexity. It is a Semantic-based System which helps for Service Discovery in cloud computing. Similarly, we found a unique technique for data retrieval. We observed this technique as it is not a back-up technique but it provides 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 viz. 1) A DHT (Distributed Hash Table) protocol 2) A SON (Semantic Overlay Network), 3) A measure of semantic similarity among service description . Hence, ERGOT combines both these network Concepts.

Sr No.	Approach	Pros	Cons
1	Parity Cloud Service [1]	Privacy Economical	Complexity is high Implementation
2	HSDRT [2]	Used for Movable clients Like laptop, Smart Phone	Increase redundancy Costly
3	Linux Box [3]	Economical Implementation Simple	Complete server Backup at a time Required higher bandwidth
4	ERGOT [4]	Privacy Perform exact-match retrieval	Implementation complexity Time complexity
5	Cold Hot Back-up Strategy [5]	Triggered only when failure detected	Cost increases as data increases gradually
6	Shared backup router resources (SBBR) [6]	Works even if router fails It concerns with cost reduction	Unable to include optimization concept with Cost reduction Inconsistencies between logical and physical configurations
7	Rent Out the Rented Resources [7]	Cost depends on the infrastructure utilization	Resources must kept under special attention due to rented concepts

The Remote backup services should fulfill the following aspects:

Data Confidentiality

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.

Data Integrity

Data integrity verifies the 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.

Data Security

Data Security deals with protecting the client's data is also the utmost priority for the remote server. And either intentionally or unintentionally, any user's data should be not able to access by third party or any other client's.

Trustworthiness

The cloud should be trustworthy. Clients may store their private data on the main cloud, so the main as well as the remote back-up cloud should be trustworthy.

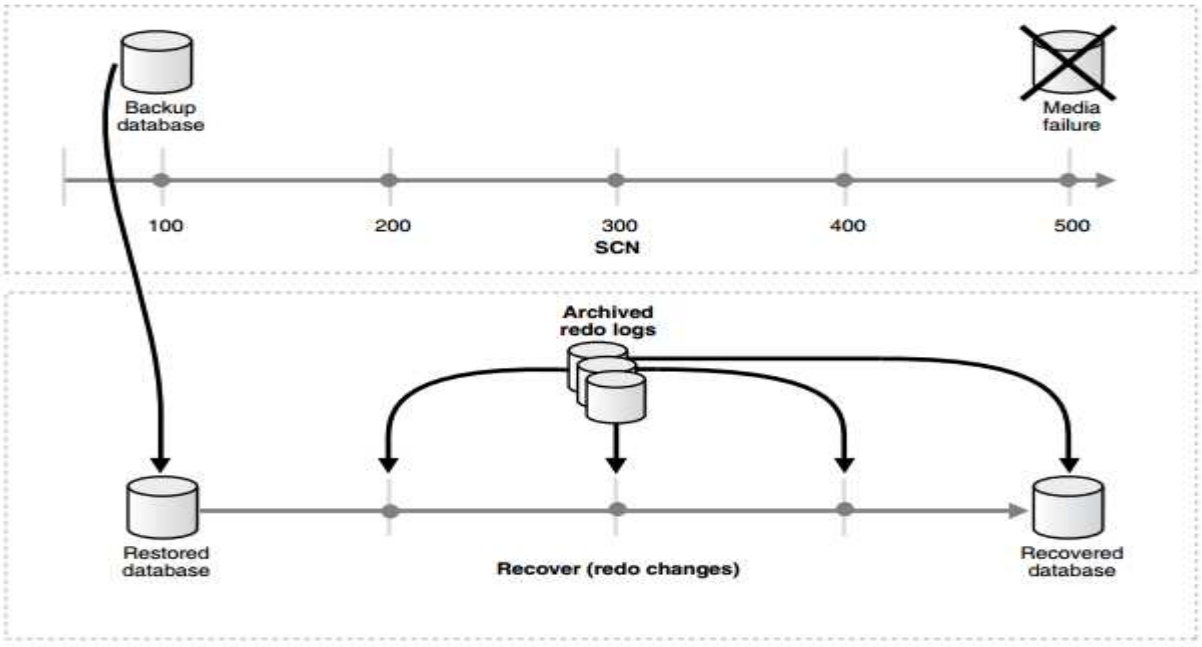
Cost Efficiency

The recovery cost should be lesser. Lesser the cost of recovery, better the system's rating will be.

BACKUP AND RECOVERY: BASIC CONCEPTS

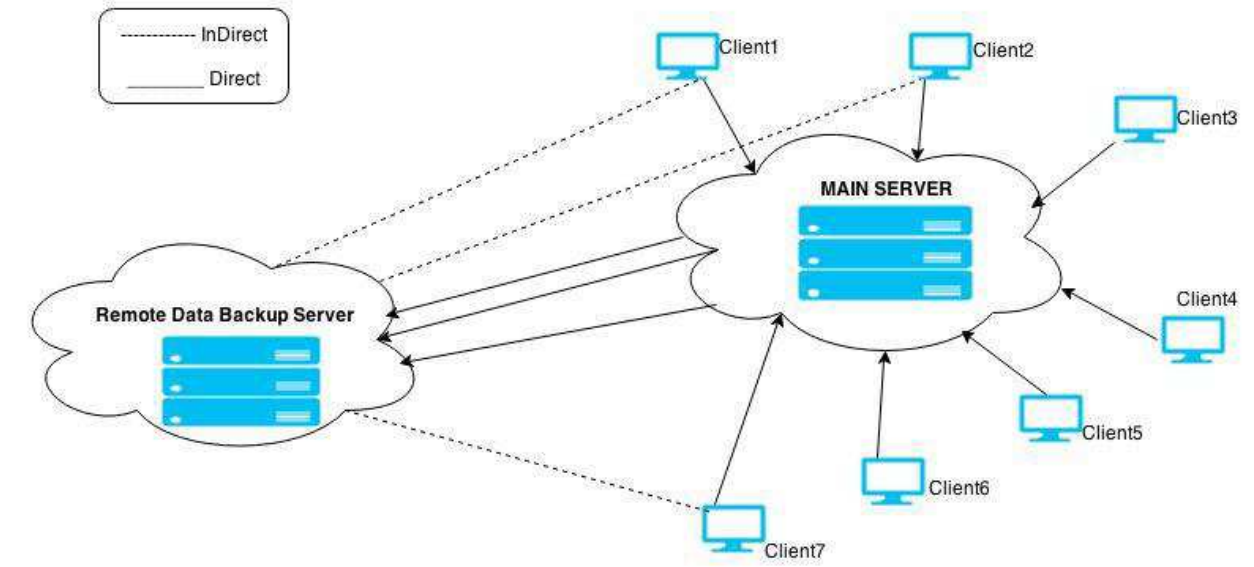
The files and other structures that make up an Oracle database store data and safeguard it against possible failures. This discussion introduces each of the physical structures that make up an Oracle database and their role in the reconstruction of a database from backup. This section contains these topics:

■Data files and Data Blocks
■ Redo Logs
■ Undo Segments
■ Control Files



REMOTE DATA BACK-UP 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.



The Remote backup services should cover the following issues:

1) Privacy and ownership.
2) Relocation of servers to the cloud.
3) Data security.
4) Reliability.
5) Cost effectiveness.
6) Appropriate Timing.

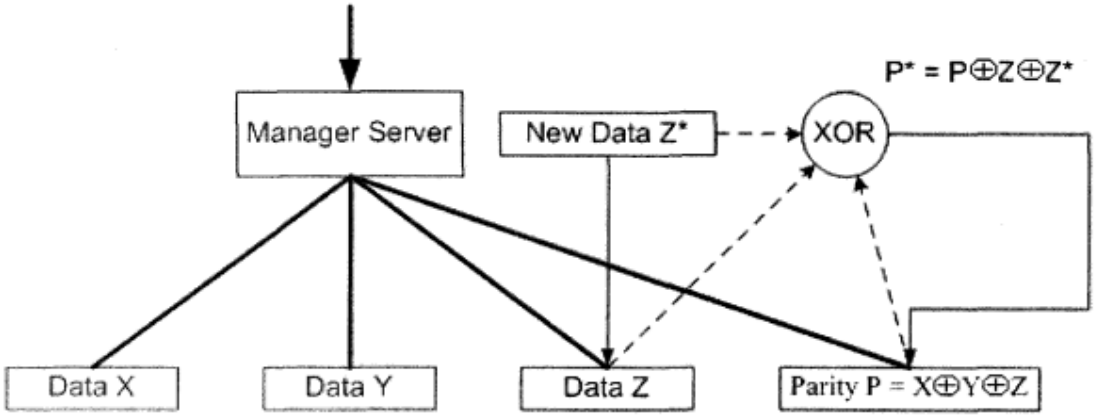
DESIGN OF THE PROPOSED SYSTEM

Seed block Algorithm focuses on simplicity of the back-up and recovery process. It uses the concept of Exclusive– OR (XOR) operation of the computing world. For ex: - Take two data files: C and D. When the XOR of the C and D it produced Z i.e. Z = C⊕D. If C data file get destroyed and we want our C data file back then we can get C data file back, by using Exclusive– OR (XOR) operation the data file and Z.

Seed Block Algorithm consists of the Main Cloud and its clients and the Remote backup Server. Here, first Random number in the cloud and unique client id for each client. Then, whenever the client id is register in the main cloud; then client id and random number is getting EXORed (⊕) with each other to generate seed block for the particular client. Whenever client creates the file then it is stored at the main cloud. When it is stored at main server, then 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'.

If either due to some reason files in main cloud crashed / damaged, then the user can get the original file back by EXORing file with the seed block of the corresponding client to produce the original file and return the original file back to the requested client.

EX-OR Client _Id and Random Number



SEED BLOCK ALGORITHM

Initialization: Main cloud:M_c**; Remote Server:**R_s **;**

Clients of Main Cloud:: $\mathbf{C_i}$ **; Files:**$\mathbf{a_1}$ **and:**$\mathbf{a'_1}$ **;**

Seed block: $\mathbf{S_i}$ **; Random Number: r;**

Client's ID : Client _ $\mathbf{Id_i}$

Input:$\mathbf{a_1}$ created by $\mathbf{C_i}$ **; r** is generated at $\mathbf{M_c}$;

Output: Recovered file a_1 after deletion at $\mathbf{M_c}$;

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 $\mathbf{S_i}$ for each $\mathbf{C_i}$ and Store

$\mathbf{S_i}$ **at** $\mathbf{R_s}$

$\mathbf{S_i = r} \oplus$ **Client** $\mathbf{\_Id_i}$**.**(Repeat step 2 for all clients)

Step 3:: If $\mathbf{C_i}$ **/Admin** creates/modifies a $\mathbf{a_1}$ and stores at

$\mathbf{M_c}$**,** then $\mathbf{a_1'}$ create as

$$a'_1 = a_1 \oplus S_i$$

Step 4: Store **a'** at a_s.

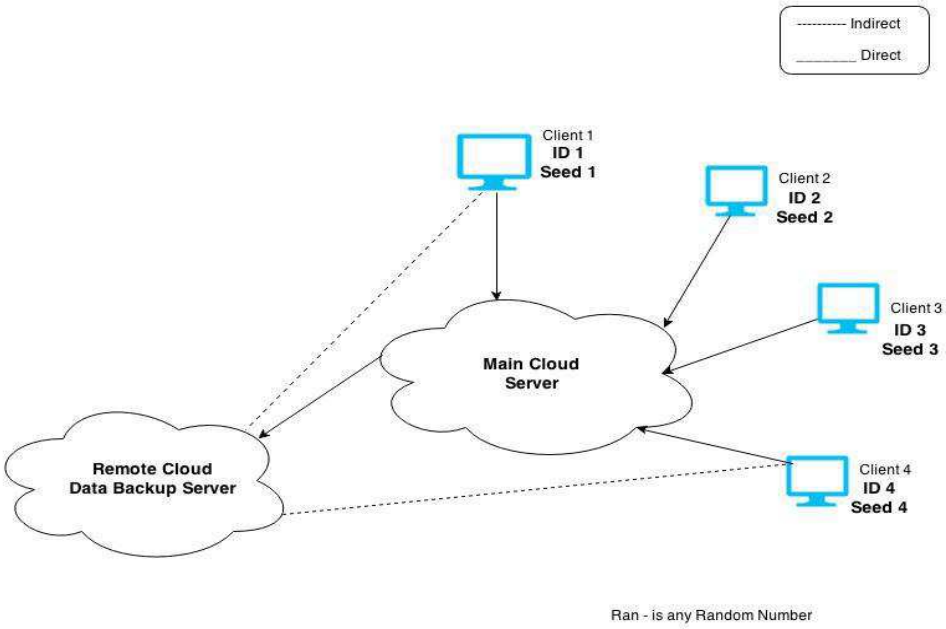
Step 5: If server crashes a_1 deleted from M_c,

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.



RESULTS AND DISCUSSION

In this section we will be discussing the experimentation and result analysis part of the proposed method. For experimentation purpose we will be taking a system with minimal specification for main cloud as well as remote cloud. Minimal specifications are given as per in the following table:

System Environment

	Main Cloud Server	Remote Cloud Server
CPU	Core2 Quad Q660 2.40GHz	Core2 Quad Q660 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 the files stored at the remote cloud server are of the same size that of the files stored at the main cloud server by the client.

Performance analysis for different types of files

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	7.57	2	150
64KB	13.7	3	160
2MB	3700	5	164
32MB	7400	8	250
1GB	16800	15	280

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]. 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.

[2]. 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.

[3]. 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**

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

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

[7]. 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.

[8]. M.Armbrust et al, "Above the clouds: A berkeley view of cloud computing,"http://www.eecs.berkeley.edu/Pubs/TechRpts/**2009**//EEC S-2009-28.pdf.
[9]. F.BKashani, C.Chen, C.Shahabi.WSPDS, **2004**, "Web Services Peer to- Peer Discovery Service ," ICOMP.