International Journal of Emerging Engineering Research and Technology
Volume 2, Issue 7, October 2014, PP 82-89
ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

Intelligent Cloud Back-Up System

Tanay Kulkarni [1], Krupali Dhaygude[2], Sumit Memane[3], Onkar Nene[4]

Department of Computer Engineering, RMD Sinhgad School of Engineering,
Pune, Maharashtra, India

Abstract: *In cloud computing, data generated in electronic form are in large 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 plan using Seed Block Algorithm (SBA) with Advance Encryption Standard (AES) algorithm. In this paper we are proposing a procedure which allows users to store their data onto the cloud, as soon as the file is stored at the first cloud server it gets encrypted using AES. In case if the certain file gets deleted due to any reason, AES helps to recover that file from a backup file which is stored at a remote location. The time related issues are also being solved by proposed method such that it will take minimum time for the recovery process. Proposed method also focuses on the security concept for the back-up files stored at remote server using AES encryption algorithm.*

Keywords: *Seed Block Algorithm, AES, Cloud back-up, Remote cloud, Main Cloud*

1. INTRODUCTION

Today, Cloud Computing is itself a gigantic technology because of its advantages over previous systems like grid or cluster computing. 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 amount of data on big data centre and according to the requirements of the customers these data centre virtualized the resources and expose them as the storage pools that helps user to store files or data objects. Number of users share the same cloud storage provided by a certain service provider.

A human error, faulty equipments, network connectivity, a bug or any criminal intent may put our cloud storage's security at stake. Cloud service provider may also make some changes in the configuration; this may lead to loss of alteration of the information stored by user. As the data modification, deletion and insertion of new data is permitted, there is possibility of data loss. To solve these difficulties we need to provide data integrity for our cloud. In literature many techniques have been proposed PCS[1], HSDRT[2], Linux Box [3], ERGOT[4], Cold/Hot backup strategy [5] etc. that, discussed the data recovery process. However, still various successful techniques are lagging behind in some critical issues like implementation complexity, low cost, security and time related issues.

To overcome the disadvantages of previously proposed systems we are proposing a new method based on Seed Block Algorithm (SBA) and Advance Encryption Standards (AES). The proposed procedure works in following manner: in first step it allows users to collect and store their files onto the main cloud. As soon as the files get stored at the cloud, those get encrypted using AES algorithm. In step two, in case of file deletion it helps user to recover the files. This paper is organized as follows: Section II focuses on the related literature of existing methods those 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 proposed method based on AES and SBA algorithms and section V shows the results and experimentation analysis of the proposed method. In the last section VI conclusion is mentioned of the proposed method.

2. 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[1], HSDRT[2], Linux Box [3], ERGOT[4], Cold/Hot backup strategy [5] 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 [2] 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[3] 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) [4] 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 [4]. Hence, ERGOT combines both these network Concepts. By implementing a Semantic

Overlay Network over a Distributed Hash Table, ERGOT proposed semantic-driven query answering in DHT-based systems. However it does not go well with semantic similarity search models.

Table1. *Comparison between Various cloud data back-up techniques*

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

All the existing solutions for cloud back-up system somehow fail in various aspects. The pros and cons of all these foresaid techniques are described in the Table-I. The role of a remote data back –up server is very crucial and hot research topic due to the high applicability of backup process in the companies.

3. REMOTE DATA BACKUP SERVER AND ITS ARCHITECTURE

When we think about Backup server of main cloud, we only talk about the replica of main cloud server. When this Backup server is at remote location and having the complete state of the main cloud, then this remote location server is identified as Remote Data Backup Server.

And in case if the central repository loses its data under any scenarios both of any natural calamity or by human attack or deletion that has been done by mistake and then it uses the information from the remote server. The main purpose of the remote backup facility is to help user to collect information from any remote location even if data not available on main cloud. As shown in Fig-1 clients can access the files from remote repository even if the data is not available on central repository.

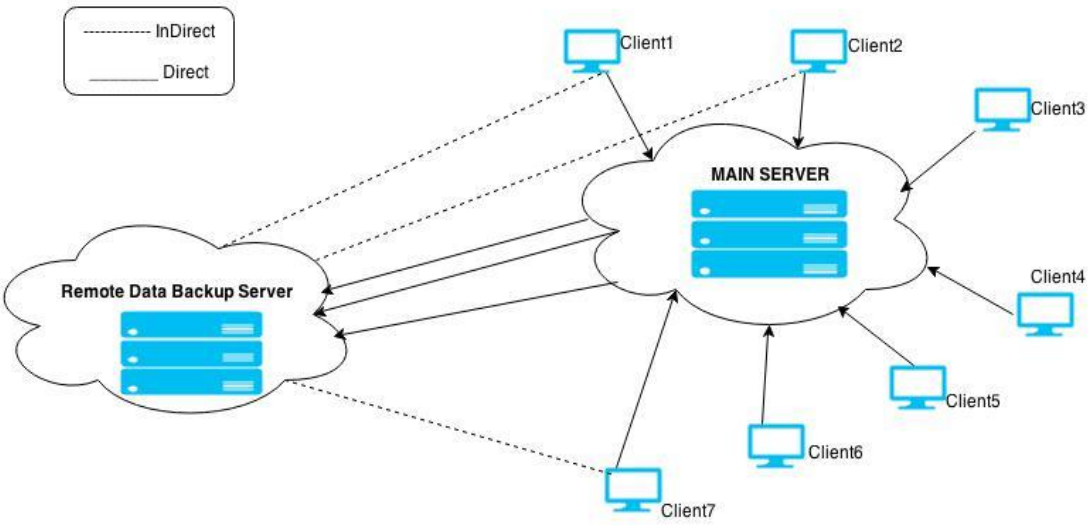

Fig1. *Remote data Backup Server and its Architecture*

The Remote backup services should fulfill the following aspects:

3.1. 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.

3.2. 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.

3.3. 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.

3.4. 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.

3.5. Cost Efficiency

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

In the next section we will be discussing a technique of back-up and recovery in cloud computing domain that will cover the issues mentioned before.

4. DESIGN OF THE PROPOSED SYSTEM

As discussed in previous sections, to overcome the shortcomings of the existing system we are proposing a new system which makes use of Seed Block Algorithm with Advance Encryption Standard algorithm. In this section complete analysis of this method is mentioned.

When client uploads the files on the main cloud server, the file gets encrypted in the first stage itself. For encryption purpose we are using Advance Encryption Standard (AES) algorithm.

4.1. Advance Encryption Standard

Phase 1 and 4:

Advanced Encryption Standard (AES): The Advanced Encryption Standard (AES) is a symmetric key encryption standard. The standard consists of three block ciphers AES-128 AES-192 and AES-256 adopted from Rijndael. Each of these ciphers has a 128-bit block size with key sizes of 128, 192 and 256 bits respectively. The key size used for an AES cipher denotes the number of repetitions of transformation rounds that converts the input called the plaintext into the final output called as cipher text. Numbers of cycles of repetition are as follows: 10 cycles for 128-bit, 12 cycles for 192-bit and 14 cycles for 256-bit.

Phase 2:

Secured searching of valuable data in database:

The objective is to collect the similar queries from various users and store in the database. We can maintain data confidentiality against untrusted parties. Data owner, service provider and trusted client are participants. Data owner is one who stores the data in the database. Here server is the database or

third party who maintains the data in the database. Trusted client is one who needs the data from the database. In this project the data owner provide the privacy to the sensitive information [8]. We use DNA Test related information so we collected the DNA Test related data and stored in my database. Only authorized users are allowed to access the data. Nobody else should be able to view the data. So that data will be kept private and secured. Based on the queries it will be accessible to the trusted users alone. The transformation technique used in AES offers perfect data privacy for the data owner but it gives the final result at multiple rounds of communication.

Phase 3:

Apriori Algorithm:

Apriori is designed to operate on databases containing transactions [9]. The purpose of the Apriori Algorithm is to find associations between different sets of data. The output of Apriori algorithm is sets of rules that tell us how often items are contained in certain sets of data.

The file gets encrypted after the first stage by Advance Encryption Standard (AES) algorithm. After encrypting the file we will be applying Seed Block Algorithm (SBA) in the next phase of the process. This algorithm will be used to create a back-up file on the remote server.

4.2. Seed Block Algorithm (SBA)

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. Seed Block has two sequences first is back-up sequence and second is restore sequence. In the main cloud we set a random number and seed for each client that wants to store data onto the cloud. In the next step, previously generated random number and seed are EXORed to form the seed. The seed which is generated by this process is unique for each client. Whenever client creates the file in cloud first time, it is stored at the main cloud and gets encrypted using AES. After that the main file of client is being EXORed with the Seed Block of the particular client. In this manner a back-up file for the main file is created by the server and the back-up file is stored at the back-up cloud which is at remote location. Unfortunately if the file in the main cloud server gets deleted due to any reason, user can retrieve the lost file from remote back-up server with the help of Seed Block which is unique for each user.

The Seed Block Algorithm is as follows:

Initialization: Main Cloud Server: - M_c; Remote Cloud Server: - R_s;

Clients of Main Cloud: - C_i; Files:-a_1, a_1’ and a_1’’;

Seed Block: - $Seed_i$;

Random Number: - Ran_i; Client’s Id: - $Client_{id}$

Input: Client c_i is created at main server; a_1 created by client; a_1’ is generated after applying AES to a_1; Ran_i is generated at M_c.

Output: Recovered File a_1’ after deletion at main cloud M_c.

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

Step 1: Generate a random number. int Ran_i =ran_no( );

Step 2: Create a Seed Block for each C_i and Store $Seed_i$ at R_s.

$Seed_i = Ran_i$ XOR $Client_{id}$ (Repeat Step2 for all clients).

Step 3: If C_i modifies a_1 and stores at M_c, then a_1' is created after applying AES and a_1'' is created as a_1''=a_1' XOR $Seed_i$;

Step 4: Store a_1'' at remote cloud server 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'' EXOR $Seed_i$;

Step 6: Return a_1 to C_i after decrypting a_1'.

Step 7: End.

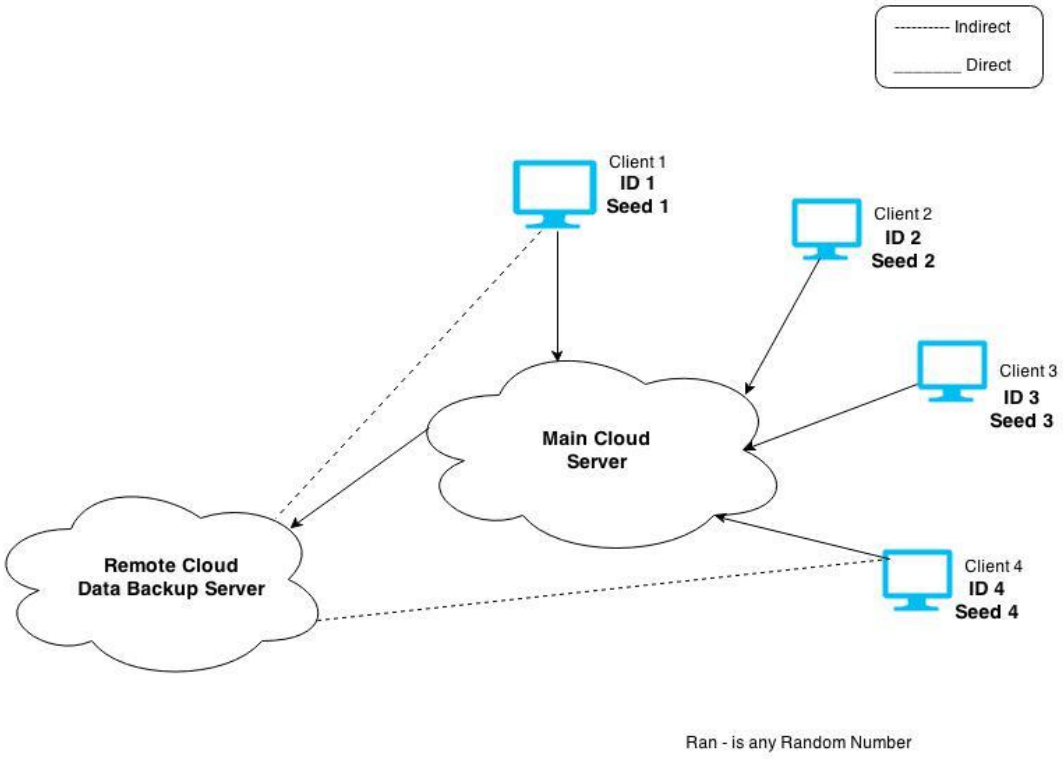

Figure 2. *Proposed System Architecture*

5. EXPERIMENTATION AND RESULT ANALYSIS

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:

Table2. *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. The following table shows that the proposed method preserves the size of the file which is uploaded at the main cloud server by the client.

Table3. *Performance analysis for different types of files*

Type	Size of Original File in Main Cloud Server	Size of Back-up File in Remote Server	Size of Recovery File After Recovery Process
Text(.txt/.docx/.doc/.pdf/.xl)	450KB	450KB	450KB
	2.3MB	2.3MB	2.3MB
Image(.jpeg/.png/.bitmap)	120KB	120KB	120KB
	3MB	3MB	3MB

Processing time means the time taken by the system to process all the requests. Client uploads file on the main cloud server, it gets encrypted using Advance Encryption Standard algorithm. After that the main cloud server takes a back-up of the file uploaded by the client and stores backed-up file on the remote cloud server. Following table shows the processing time of all tasks. We also observed that as the size of file increases the processing speed also increases.

Table4. *Effect of size of data on performance*

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

The above table shows the CPU usage according to the file sizes. We can see a trend in that data, as the file size increases the time taken to process task also increases.

6. CONCLUSION

In this paper, we presented detailed design of SBA and AES algorithms. Proposed system helps clients to collect their files from remote server cloud. Experimentation and result analysis shows that proposed system also focuses on the security concept for the back-up files stored at remote server, by using AES algorithm. The time related issues are also being solved by proposed system 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] r. Rathika1, dr. K. Raja2 "secured searching of valuable data in a metric space based on similarity measure" ijcsmc, vol. 2, issue. 4, april 2013, pg.507 – 512.

[9] A.V. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke. Privacy preserving mining of association rules. In KDD, pages 217–228, 2002.

AUTHORS' BIOGRAPHY

Tanay Kulkarni. Pursuing Bachelor of Engineering degree in Computer Science field from University of Pune; currently in last year.

Krupali Dhaygude. Pursuing Bachelor of Engineering degree in Computer Science field from University of Pune; currently in last year.

Sumit Memane. Pursuing Bachelor of Engineering degree in Computer Science field from University of Pune; currently in last year.

Onkar Nene. Pursuing Bachelor of Engineering degree in Computer Science field from University of Pune; currently in last year.