Secure Data Back-up Technique for Cloud Computing using Seed Block Algorithm with RSA Encryption

Mrs. Priyanka Patil
SIT Lonavala, SPPU
Maharashtra, INDIA

Prof. Vikas Kadam
SIT Lonavala, SPPU
Maharashtra, INDIA

***Abstract*— Today, in cloud computing, data has been generated in electronic form are large in amount. It required the data recovery services to maintain the data efficiently. To cater this, we propose a smart remote data backup algorithm, Seed Block Algorithm (SBA) with RSA encryption in this paper. Proposed algorithm has important objectives; first, in the absence of network connectivity, it helps the users to collect information from any remote location. Second, if the cloud gets destroyed due to any reason or if the file gets deleted, it helps to recover the files. Proposed SBA, which is based on RSA encryption, will take minimum time for the recovery process. So, the time related issues are solved. It also focuses on the security of the back-up files stored at remote server. Data stored on cloud is thus secure using RSA security. The proposed system is based on the Seed Block Algorithm (SBA) and RSA Encryption.**

***Index Terms*— Cloud Computing, Seed Block Algorithm (SBA), RSA, Backup, Recovery**

I. INTRODUCTION

Cloud Computing is surpassing all the previous technology of computing (like grid, cluster, distributed etc.) of this challenging and competitive IT world, which is itself a gigantic technology, today. Its advantages overcome the disadvantage of various early computing techniques. So, the need of cloud computing is increasing day by day. Cloud storage supports online storage where data is stored in form of pool and it is hosted by third parties. The hosting party operates large amount of data on large data center. These data centers provide the resources and expose them as the storage pools according to the customer's requirements that help user to store data objects or files.

It is possible for other customers to access your data, as number of user shares the storage and other resources. Some critical issues may put our cloud storage on the risk and danger like a bug, human error, faulty equipment's, network connectivity, or any criminal intent. We introduce here a term data dynamics which means changes in the cloud are made very frequently. It includes various operations such as deletion, insertion and block modification. Remote data integrity is also important along with archiving and data back-up services. Because integrity plays an important role in back-up and recovery services.

There is a huge increase in the electronic data, in today's world. To store this large amount of data, it needs large volume of data storage devices. This requirement introduces 3 Tera Byte HDD. Usually, consumer stores large volume of private data in cloud. If cloud gets damaged or corrupted unfortunately, it results in the loss of all private and important data. So, there must be data back-up techniques in cloud computing that provide the data at the time of loss of data or cloud failure.

Many data back-up techniques have been proposed HSDRT[1], PCS[2], ERGOT[4], Linux Box [5], Cold/Hot backup strategy [6] etc. in literature. These systems provide reliability and high privacy protection however still various techniques are lagging behind low cost, low implementation complexity, time related issues and security. To cater this, we propose a remote data backup algorithm, Seed Block Algorithm (SBA), in this paper. Its contribution is threefold; first, in the absence of network connectivity, it help the users to collect information from any remote location. Second, if the cloud gets destroyed due to any reason or if the file gets deleted, it helps to recover the files. Third, It also supports file sharing application such that system user can share his file to another authenticated user.

This paper is organized as follows: The existing methods that are successful to some extent in the cloud computing domain are described in Section II as literature survey. In Section III, we focus on the remote data backup server. The detailed description of the proposed seed block algorithm (SBA) using RSA encryption is given in Section IV and discussion and results of the proposed SBA is shown in Section V. Finally, in Section VI conclusions are discussed.

II. LITERATURE SURVEY

In literature, we found recent back-up and recovery techniques in cloud computing such as HSDRT[1], PCS[2], ERGOT[4], Linux Box [5], Cold/Hot backup strategy [6] etc. Detail study reveals that no techniques among all these techniques is capable of providing best performances under all uncontrolled circumstances such as security, redundancy, cost, low implementation complexity, and recovery in short span of time.

HS-DRT 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, this model is somehow unable to declare as perfect solution for back-up and recovery.

Parity Cloud Service technique (PCS) [2] is a reliable, very simple, easy to use and more convenient for data recovery and is based on parity recovery service. PCS can

ISSN: 2278 – 7798

recover data with very high probability and it has low cost. It uses a new technique of generating virtual disk in user system for data backup, make parity groups across virtual disk, and store parity data of parity group in cloud for data recovery. It uses the Exclusive–OR (⊕) for creating Parity information. However, PCS somehow lags behind in providing perfect solutions to backup and recovery due to some limitations and is unable to control the implementation complexities.

Efficient Routing Grounded on Taxonomy (ERGOT) [4] is based on the semantic analysis for Service Discovery in Distributed Infrastructures in cloud computing and is and unable to focus on time and implementation complexity. We found a unique way of data retrieval. We consider ERGOT 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. It also exploits both coarse-grain service functionality descriptions and at a finer level. 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]. DHTs and SONs both networks architectures have some shortcomings. Hence, ERGOT combines both these network Concept. By building a SON over a DHT, ERGOT proposed semantic-driven query answering in DHT-based systems. An extensive evaluation of the system in different network scenarios demonstrated its efficiency both in terms of accuracy of search and network traffic. However, it does not go well with semantic similarity search models.

Each backup technique in cloud computing is unable to achieve all the issues of remote data back-up server. All these approaches have its own advantages and disadvantages which are described in the Table-1. 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

Backup server of main cloud means the copy of main cloud. When this Backup server is far away from the main server i.e. at remote location and having the complete state of the main cloud, then it is called as Remote Data Backup Server. The main cloud is called as the central repository. Remote backup cloud is termed as remote repository.

If the central repository lost its data due to any natural calamity or by human attack or deletion that has been done mistakenly, then it uses the information from the remote repository. Its main purpose is to help clients to collect information from remote repository even if there is no network connectivity or if data not found on main cloud. As shown in Fig 1, if clients didn't find data on central repository, then clients can access the files from remote location (i.e. indirectly).

The Remote backup services should cover the following issues:

1) Data Integrity
2) Privacy and ownership.
3) Relocation of servers to the cloud.
4) Data security
5) Data Confidentiality
6) Reliability or Trustworthiness
7) Cost effectiveness or Cost efficiency
8) Appropriate Timing

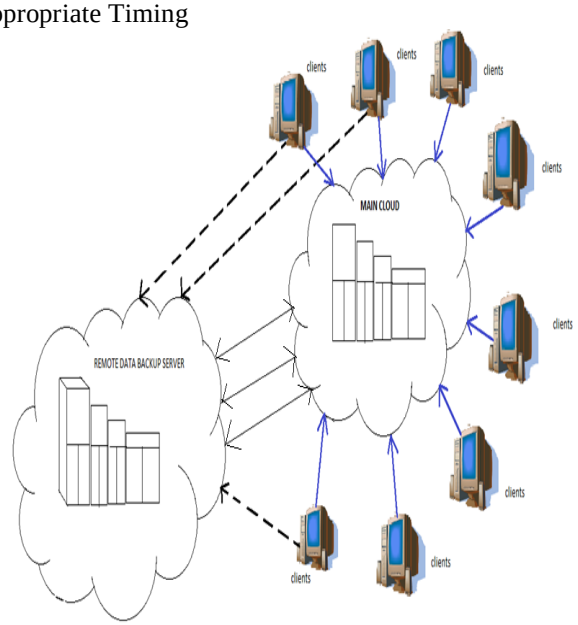

Fig.1 : Remote data Backup Server and its Architecture

IV. SYSTEM ARCHITECTURE

A. Proposed Algorithm Design

This algorithm basically uses the concept of Exclusive–OR (XOR) operation of computation. For ex. Consider two data files: P and Q. When we XOR P and Q, it produced X i.e. X = P ⊕ Q . If suppose P data file is deleted and we want our P data file back then we can get it with the help of Q and X data file .i.e. P = X ⊕ Q.

In this algorithm, 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 (⊕) 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. It is also encrypted using public key RSA and that output 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. For that, first, using the private key of the user, file' is decrypted and then by EXORing file′ with the seed block of the corresponding client, user can produce the original file and return the resulted file i.e. original file back to the requested client. This encryption operation is used to support security as well as file sharing application such that system user can share his file to another authenticated user providing him private key via email to decrypt file'. The architecture representation of the Seed Block Algorithm is shown in the Fig.2.

ISSN: 2278 – 7798

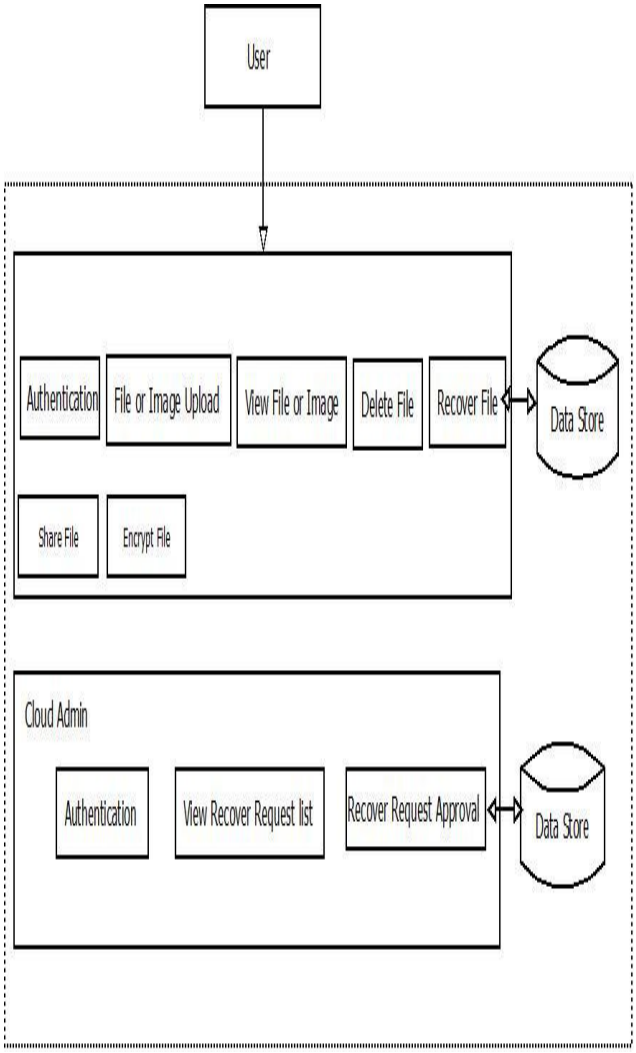

Fig.2 : Proposed Seed Block Algorithm Architecture

B. SBA Algorithm with RSA encryption

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: $Client\_Id_i$

Algorithm:

Step 0 : Start

Step 1: In this algorithm, first we set a random number in the cloud and unique client id for every client.

Step 2: Second, whenever the client id is being 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.

Step 3: The generated seed block corresponds to each client is stored at remote server.

Step 4: 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.

Step 5: It is also encrypted using public key RSA.

Step 6: And that output file is stored at the remote server in the form of file' (pronounced as File dash).

Step 7: If either unfortunately file in main cloud crashed /damaged or file is been deleted mistakenly, then the user will get the original file.

Step 8: For that, first, using the private key of the user, file' is decrypted.

Step 9: And then by EXORing file' with the seed block of the corresponding client, user can produce the original file and return the resulted file i.e. original file back to the requested client. (This encryption operation is used to support file sharing application such that system user can share his file to another authenticated user providing him private key via email to decrypt file')

Step 10: Stop.

C. Mathematical Model

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.
Process:
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_1 and stores at M_c, then a_1' create as
$a_1' = a_1 \oplus S_i$
Step 4: Encrypt data using following mechanism using public key e of C_i.
$E = M^e \pmod n$
where $M = a_1$
Step 5: Store E at M_c and store a_1' at R_s.
Step 6: If server crashes a_1 deleted from M_c , then, we do EXOR on a_1' at R_s to retrieve the original M^e as:
$M^e = a_1' \oplus S_i$
Step 7: Decrypt data using following mechanism using private key d of Owner.
$E^d = (M^e)^d$
where $M^e = M$
Step 8: Return a_1 to C_i **.**
Step 9: END.

D. RSA Methodology

We are using RSA Encryption Security technique in our proposed algorithm. We will explain it below:

ISSN: 2278 – 7798

RSA (Rivest, Shamir, Adleman) algorithm requires keys of at least 1024 bits for good security. The RSA method is based on some principles from number theory. We will now summarize how to use the method.

1. Choose two large primes, *p* and *q* (typically 1024 bits).
2. Compute $n = p \times q$ and $z = (p - 1) \times (q - 1)$.
3. Choose a number relatively prime to *z* and call it *d*.
4. Find e such that e x d = 1 mod *z*.

With these parameters computed in advance, we are ready to begin encryption. Divide the plaintext (regarded as a bit string) into blocks, so that each plaintext message, *P*, falls in the interval $0 < P < n$. Do that by grouping the plaintext into blocks of *k* bits, where *k* is the largest integer for which $2^k < n$ is true.

To encrypt a message *P*, compute $C = P^e \pmod{n}$. To decrypt *C*, compute $P = C^d \pmod{n}$. It can be proven that for all *P* in the specified range, the encryption and decryption functions are inverses. To perform the encryption, you need *e* and *n*. To perform the decryption, you need *d* and *n*. Therefore, the public key consists of the pair (*e*, *n*), and the private key consists of (*d*, *n*).

E. System Modules

There are following modules in a system:

1. User Login & Registration

This is the authentication module of the system facilitating users to add themselves to the system as well as authenticate and utilize the system, thereby providing access to valid registered users in the system.

- File Upload
- File Encryption
- Image Upload
- Image Encryption
- View File
- Delete File
- View Image
- Download File and Image
- Decryption of File and Image
- Recover File

2. Cloud admin Login & Registration

This is the authentication module of the system facilitating admin to add them to the system as well as authenticate and utilize the system, thereby providing access to valid registered admin in the system.

- File/image Recover Request Approval

*F. Working of System***:**

The total system works on the "Seed Block algorithm". In the System, there are two main modules as 1st is user and 2nd is admin. Now user and admin enters into their authorized area so both of them goes through authentication process. After getting into "seed block" user can do his/her desired utilization as user can upload, view, download, delete and recover. The deleted file gets store on the backup server. When user uploads the file, the file goes through RSA encryption and file gets store into the database where user can view, delete that file. When user wants to download file, he requests system to download where admin alias the system accepts request generates OTP i.e. One Time Password and sends it to user's e-mail box and database. So when user enters that OTP into system then only he can get his file downloaded. If file gets deleted, user can get back his file as system recovers it.

G. Existing System(SBA without Security) Limitations

- High Implementation cost
- Low security
- Reduced Privacy
- High implementation complexity
- Poor Recovery
- Redundancy
- Time Complexity

H. Proposed System(Secure SBA) Advantages

- Recover same size data
- Low cost
- Good Privacy
- Reduced Time Complexity
- High Security
- Good Storage Facilities
- Backup and Recovery in short span of time.

I. Hardware and Software Used

1. Hardware System Configuration

Speed - 1.1 Ghz
RAM - 256 MB(min)
Hard Disk - 20 GB
Floppy Drive - 1.44 MB
Key Board - Standard Windows Keyboard
Mouse - Two or Three Button Mouse
Monitor - SVGA

2. Software System Configuration

Operating System : Windows95/98/2000/XP
Application Server : Tomcat5.0/6.X
Front End : HTML, Java, Jsp
Scripts : JavaScript
Server side Script : Java Server Pages
Database : Mysql
Database Connectivity: JDBC
Tool: Netbeans

V. RESULTS AND DISCUSSION

We discuss the performance evaluation and result analysis of the proposed SBA algorithm in this section. We perform 2 main analysis in proposed SBA likely as :

Time Analysis: It will demonstrate the time required for the uploading mechanism for existing v_s proposed scenario.

Backup & Recovery Time : This is the time required for performing the backup and recovery of a particular file from the cloud.

ISSN: 2278 – 7798

To evaluate performance, it is observed that memory requirement is more in remote server as compared to the main cloud's server because additional information is placed onto remote server (for example- different Seed Blocks of the corresponding client.) 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. From this we conclude that proposed SBA recover the data file without any data loss.

We also observed that as data size increases, the processing time increases. Performance which is megabyte per sec (MB/sec) is constant at some level even if the data size increases. CPU utilization at Main Cloud and Remote Server increases or decreases as per the load on that respective cloud. Proposed algorithm recovers same size of data.

VI. CONCLUSION AND FUTURE SCOPE

In this paper, we presented detailed design of SBA with RSA technique. Proposed SBA focus on the security concept for the back-up files stored at remote server so that security is maintained.. By doing the encryption, we are assuring the safety of the data stored on the cloud. Hence, an efficient way to store and retrieve data in a safe manner has been discussed.

In future work, We can design the prototype to support data like : Video, Sound, Other data formats. Current System can replicate data on one remote server. In future, we can extend it to replicate data on Multiple Remote Servers. Current System may fail to satisfy Mobile and Handheld device users. But we can imply better Mobile users and Handheld devices usage support. We can provide great flexibility in storage supports by using 'Portable Accounts' concepts.

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//EEC S-2009-28.pdf.

[9] F.BKashani, C.Chen,C.Shahabi.WSPDS, 2004, "Web Services Peerto 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.

ISSN: 2278 – 7798