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

Rani Dasari[1] Sujata Shinde[2] Priyanka Domal[3] Manali Dhanawade[4]

[1,2,3,4]Department of Information Technology

Abstract— Cloud computing is an emerging computing technology that uses the internet and central remote server to maintain data and application. In cloud computing, data is generated and stored in large amount. In order to maintain its data efficiency, there is a necessity for recovery services. To cater this, in this paper we propose a remote smart data backup algorithm, Seed Block Algorithm. The objective of proposed algorithm works in two parts, firstly it allows the users to gather information from any remote location and secondly to restore the files in case of the file deletion or if the cloud gets crash due to any reason. The issues related to time and are also being solved by SBA such that it will take least amount of time for the recovery process.

Key words: SBA, Seed Block Algorithm

I. Introduction

The term Cloud refers to a Network or Internet. In other words, we can say that Cloud is something, which is present at remote location. Cloud can provide services over public and private networks, i.e., WAN, LAN or VPN. Applications such as e-mail, web conferencing, customer relationship management (CRM) execute on cloud. Cloud Computing refer to manipulating, configuring, and accessing the hardware and software resources remotely. It offers online data storage, infrastructure, and application. cloud computing offers platform independency, as the software is not required to be installed locally on the PC. Hence, the Cloud Computing is making our business applications mobile and collaborative. Security and Privacy of information is the biggest challenge to cloud computing. Security and privacy issues can be overcome by employing encryption, security hardware and security applications.

This is another challenge to cloud computing that applications should easily be migrated from one cloud provider to another. There must not be vendor lock-in. However, it is not yet made possible because each of the cloud provider uses different standard languages for their platforms.

It means the application on one platform should be able to incorporate services from the other platforms. It is made possible via web services, but developing such web services is very complex.

It is necessary for cloud systems to be reliable and robust because most of the businesses are now becoming dependent on services provided by third-party.

Either the human error, fault 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.

In order to overcome some of the security issues and recovery of files if deleted, we propose a remote smart data backup algorithm, Seed Block Algorithm (SBA).

II. Literature survey

Cloud storage provides online storage where data stored in form of virtualized pool that is usually hosted by third parties. 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. Due to which a lot of study and white papers have been published we made our literature survey on A Remote smart Data Backup Technique for cloud computing.

In[1], Yoichiro Ueno, NoriharuMiyaho, Shuichi Suzuki,MuzaiGakuendai, Inzaishi, Chiba,Kazuo Ichihara, proposed the innovative file back-up concept HS-DRT, that uses an effective ultra-widely distributed data transfer mechanism and a high-speed encryption technology.

It consists of two series one is Backup and other is Recovery. The data to be backed-up is obtained in Backup sequence. The recovery sequence is used when there is loss of data occurs the Supervisory Server (one of the components of the HSDRT) begins the recovery sequence. There are some disadvantages to this approach and thus this model cannot be state as a perfect technique for Cloud back-up and recovery. This model can be used for movable clients such as laptops Smart phones etc. the data recovery cost is comparatively costly and there is increased redundancy.

In[2], Chi-won Song, Sungmin Park, Dong-wook Kim, Sooyong Kang, have proposed a data recovery service framework, the Parity Cloud Service (PCS) provides a privacy-protected personal data recovery service. In this infrastructure, user data is not required to be uploaded on to the server for data retrieval. All the server-side resources provide the recovery services within a reasonable bound. The advantages of Parity Cloud Service are it provides a reliable data retrieval at a low cost but the disadvantage is that its implementation complexity is higher.

In[3], VijaykumarJavaraiah introduced a mechanism for online data backup technique for cloud along with disaster recovery. In this approach the cost of having the backup for Cloud platform has been reduced and also it protects data loss at the same time the process of migration from one cloud service provider to another becomes easier and much simpler. In this approach the consumers' are not dependent on the service provider and it also eliminates the associated data recovery cost. A simple hardware box is used that achieves all these at little cost.

In[4], Giuseppe Pirr´o, Paolo Trunfio, Domenico Talia, Paolo Missier and Carole Goble proposed Efficient Routing Grounded on Taxonomy (ERGOT) which is fully based on the semantic analysis and does not focuses on time and implementation complexity. This system is based on the

Semantics that provide support for Service Discovery in cloud computing. This model is built upon 3 components one A DHT (Distributed Hash Table) protocol second A SON (Semantic Overlay Network), and third A measure of semantic similarity among service description We makes a focus on this technique because it is not a simple back-up technique rather it provides retrieval of data in an efficient way that is totally based on the semantic similarity between service descriptions and service requests. ERGOT proposes a semantic-driven query answering in DHT-based systems by building a SON over a DHT but it does not go well with semantic similarity search models. The disadvantage of this model is an increased time complexity and implementation complexity.

In [5], Sheheryar Malik, FabriceHuet, proposed the lowest cost point of view a model “Rent out the Rented Resources”. This technique focuses on minimizing the cloud service’s monetary cost. It proposed a model for cross cloud federation which includes of three phases that are 1) Discovery, 2) Matchmaking and 3) Authentication. This model is based on the concept of cloud vendors that rent the resources from different venture(s) and after virtualization, rents it to the clients as cloud services.

In [6], Lili Sun, Jianwei An, Yang Yang, Ming Zeng, propose a technique in which there is a moderately increase in cost with the increase in data i.e. The Cold and Hot back-up strategy that performs backup and recovery on trigger basis of failure detection. In CBSRS (i.e. Cold Backup Service Replacement Strategy) recovery process, it is triggered when a loss of service is detected and it will not be triggered when there is no loss i.e. when the service is available.

The HBSRS (i.e. Hot Backup Service Replacement Strategy), is a transcendental recovery strategy for service constitution that is used for dynamic network. During the implementation of process, the backup services remains in the activated state and the first returned results of services will be used to ensure the successful implementation of service composition.

In[7]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 depository lost its data, then it uses the information from the remote depository. The purpose is to help clients to gather information from remote depository either if network connectivity, 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 depository, then clients are allowed to access the files from remote depository (i.e. indirectly).

In[8] Xi Zhou, Junshuai Shi, YingxiaoXu, Yinsheng Li and Weiwei Sun, 2008, "A backup restoration algorithm of service composition in MANETs," This paper proposes a backup restoration algorithm of service framework in ad hoc networks. This algorithm is based on the precondition of setting up backup service framework before disturbance service composition. Backup service framework is only set up for services provided by untrustworthy nodes which are predicted.

Disturbance service framework can be replaced by backup service composition other than reorganizing service composition.

In[9] M. Armbrust et al, “Above the clouds: A berkeley view ofcloudcomputing,”http://www.eecs.berkeley.edu/Pubs/TechRpts/2009//EEC
S-2009-28.

Obstacle Opportunity

1) Availability of Service Use Multiple Cloud Providers; Use Elasticity to Prevent DDOS
2) Data Lock-In Standardize APIs; Compatible SW to enable Surge Computing
3) Data Confidentiality and Auditability Deploy Encryption, VLANs, Firewalls; Geographical Data Storage
4) Data Transfer Bottlenecks FedExing Disks; Data Backup/Archival; Higher BW Switches
5) Performance Unpredictability Improved VM Support; Flash Memory; Gang Schedule VMs
6) Scalable Storage Invent Scalable Store
7) Bugs in Large Distributed Systems Invent Debugger that relies on Distributed VMs
8) Scaling Quickly Invent Auto-Scaler that relies on ML; Snapshots for Conservation
9) Reputation Fate Sharing Offer reputation-guarding services like those for email
10) Software Licensing Pay-for-use licenses; Bulk use sales

In[10]The cloud hook formation provides a useful similarity for cloud computing, in which the most severe obstacles with outsourced services (i.e., the cloud hook) are security and privacy issues. This paper recognizes key issues, which are assume to have long-term importance in cloud computing security and privacy, based on registered problems and exhibited weaknesses.

III. REMOTE DATA BACKUP SERVER

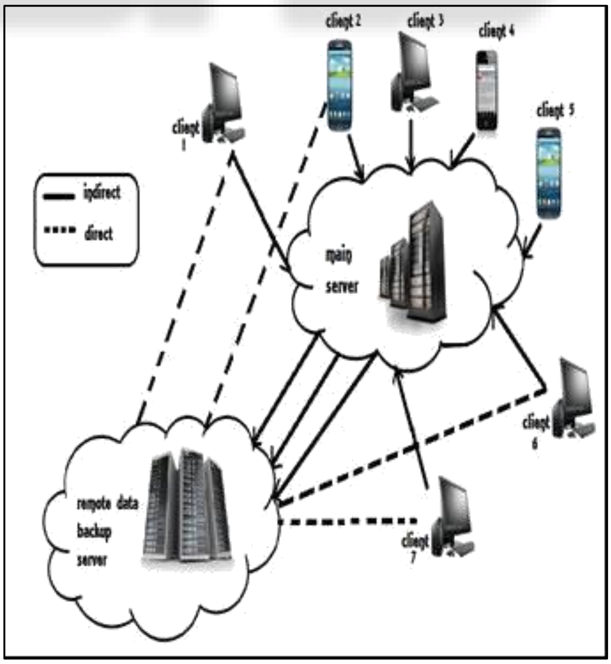

Fig. 1: Remote data Backup Server and its Architecture

When we give a thought on the concept such as Backup server of main cloud, we usually reach to a conclusion that it is a copy of main cloud. This copy of main cloud is stored in Backup server which is kept at any remote location and which knows the entire state of the main cloud, then we can label the remote located server as Remote Data Backup Server. Remote backup is a form of offsite backup technique but the main difference in both is that in remote backup you can access, restore or administer the backups from your source location i.e. no physical presence is required at the backup storage location. The main cloud can also be labeled as central repository and similarly the remote backup cloud is labeled as remote repository. Consider if any natural calamities occur such as earthquake, flood, fire etc. or any human attack such as performing deletion operation on data mistakenly etc. may lead to loss of data from central repository at this time the use of remote repository comes into picture. The main objective of remote backup facility is to provide help to the user for retrieving the data which is not found on main cloud from any remote location. Refer to Fig-1 where it is clearly shown how clients are allowed to access the data from remote repository if it is not found on central repository i.e. indirectly.

The Remote backup service must overcome the following issues:-

1) Data Integrity - It mainly refers to maintain and assure the accuracy and consistency of data over its entire period of life cycle. The main concern of data integrity is with its complete state and whole structure of the server. The complete focus of data integrity during transmission is only that the data is stored exactly as intended and during reception it should retrieve the unaltered data (i.e. the original data as stored in the main cloud)
2) Data Security – It mainly refers to protecting of data from the unauthorized users. In remote backup service it plays an important role such that the full protection to client's data must be provided. And intentionally or unintentionally the data must not be accessible to the third party or any unauthorized client or user
3) Data confidentiality – It mainly refers to managing of sensitive and confidential data of the client against any malpractice. It is important to overcome this issue during simultaneous accessing of cloud is done by the large number of users.
4) Trustworthiness – it mainly refers to the quality of the service that inspires reliability to the client's or users. The remote backup service must maintain its trustworthiness characteristic such that the clients can easily rely on it for storing its private data.
5) Cost efficiency – it mainly refers to good value of money where the benefits and usage are worth. The process of data recovery must be very efficient such that maximum number of clients or users or companies can take its advantage and they must feel that the good value of money as invested.

IV. DESIGN OF THE PROPOSED SEED BLOCK ALGORITHM

As discussed in above contest, to overcome the issues of the existing system we are proposing a advanced system which makes use of Seed Block Algorithm with Advance Encryption Standard algorithm. In this part complete analysis of this method is mentioned. Each cipher encrypts and decrypts data in blocks of 128 bits using cryptographic keys of 128-, 192- and 256-bits, respectively.

A. Seed Block Algorithm (SBA) Architecture:

Seed block algorithm focuses on back-up and recovery process and uses the concept of Exclusive–OR (XOR) operation.

It has two phases first is back-up phase and other one is recovery phase. We set random number in the main cloud and unique client id for each client that wants to store data on the cloud. In the main cloud ,whenever client id is being register ,we generate seed for that particular client by EXORing client id and random number with each other and we will store it at remote location. Whenever client creates file in cloud for first time, it is stored at the main cloud. After that we will apply AES algorithm between main file of client and seed block of that particular client. And that encrypted file is stored at remote server in the form of backup file. 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.

B. SBA Algorithm:

The proposed SBA algorithm is as follows:
Algorithm 1:

Initialization: Main Cloud Server: - M_C; Remote Server: R_S

Clients of Main Cloud: - C_i; Files:- **A1, A1'** ;

Seed Block: - $seed_i$; Random Number: - ran_i;

Client's Id: - C_id_i

Input: A1 created by C_i; rand is generated at M_C.

Output: Recovered File A1 after deletion at 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 = rand ( );

Step 2: Create a Seed Block for each C_i and Store $seed_i$ at R_S. $\oplus$

$seed_i = ran_i$ C_id_i (Repeat Step2 for all clients).

Step 3: If C_i Admin modifies A1 and stores at M_C, then A1' is created

$$A1' = A1 \oplus seed_i;$$

Step 4: Store A1' at R_S ;

Step 5: If server crashes, A1 deleted from M_C, then we do EXOR to retrieve the original A1 as

$$A1' = A1 \oplus seed_i;$$

Step 6: Return A1 to C_i.

Step 7: End.

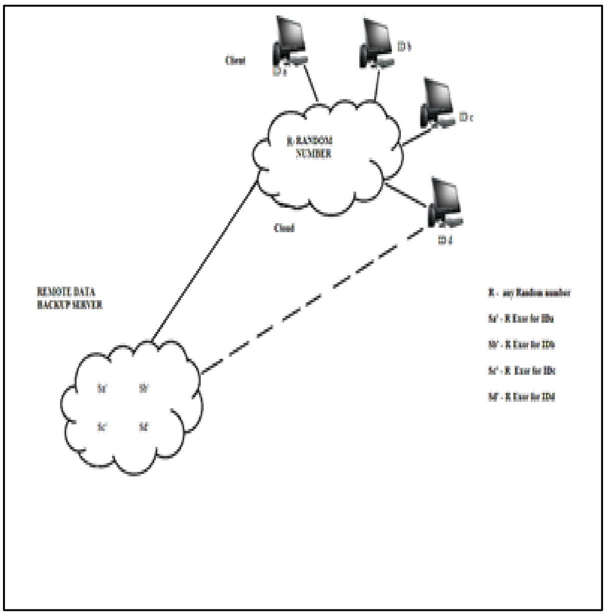

Fig. 2: System Architecture

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, "Disaster Recovery Mechanism using Widely Distributed Networking and Secure Metadata Handling Technology", Proceedings of the 4th edition of the UPGRADE-CN workshop, 2009, pp. 45-48.

[4] Giuseppe Pirr´o, Paolo Trunfio , Domenico Talia, Paolo Missier and Carole Goble, "ERGOT: A Semantic-based System for Service Discovery in Distributed Infrastructures," 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing,2010.

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

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

[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/,2009.

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

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

[11] Balazs Gerofi, Zoltan Vass and Yutaka Ishikawa, "Utilizing Memory Content Similarity for Improving the Performance of Replicated Virtual Machines", Fourth IEEE International Conference on Utility and Cloud Computing2011.

[12] P.Demeester et al., 1999. "Resilience in Multilayer Networks," IEEE Communications Magazine, Vol. 37, No. 8, p.70-76. S. Zhang, X. Chen, and X. Huo, 2010, "Cloud Computing Research and Development Trend," IEEE Second International Conference on Future Networks, pp. 93-97.

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

[14] Kruti Sharma, Kavita R Singh "Online Data Back-up and Disaster Recovery Techniques in Cloud Computing: A Review", IJEIT, Volume 2, Issue 5, November 2012.