Smart Data Back-up Technique for Cloud Computing using Secure Erasure Coding

Kolipaka Kiran[1], Janapati Venkata Krishna[2]

[1]pursuing M.Tech (CSE), [2]Associate Professor & HOD (CSE Department)

[1,2]Holy Mary Institute of Technology and Science, Keesara, Affiliated to JNTU- Hyderabad, A.P, India

***Abstract*— In cloud computing, data generated in electronic form are large in amount. To manage this data efficiently, there is an essential of data recovery services. To handle this we propose a smart remote data backup algorithm i.e. Secure Erasure Coding Algorithm. 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 problems are also being solved by proposed SEC such that it will take minimum time for the recovery process. Proposed SEC also focuses on the security concept for the back-up files stored at remote server, without using any of the current encryption techniques. In enhancement approach we will maintain 'n' servers where we can split our file into 'm' parts, while uploading data by data owner into these clouds each file will split into 'm' parts and then after it will place all these files in 'n' clouds in such a way that in every cloud we will maintain 'm-1' broken parts of original file.**

***Keywords*— Secure Erasure Code, Central Repository, Remote Repository, Cloud Server, Data Backup, Data Restore.**

I. INTRODUCTION

Cloud Computing is itself a gigantic technology which is surpassing all the foregoing technology of computing (like distributed, grid, cluster, etc.) of this competitive and challenging Information Technology (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 furnishes 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 large data centre and according to the requirements of the customer these data center convert the resources and expose them as the storage pools that help user to store files or data objects.

As number of user allocates the storage and other resources, it is feasible 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 uncertain. And changes in the cloud are also made very frequently; we can term it as data dynamics. It is supported by various operations such as deletion, block modification and insertion. Hence services are not limited for archiving and taking backup of data; remote data integrity is also needed. As the data integrity always focuses on the validity and constancy of the complete state of the server that takes care of the heavily generated data which remains unchanged during storing at major cloud remote server and transmission. Integrity plays an important function in back-up and recovery services.

However, still various successful techniques are straggle behind some critical issues like implementation complexity, low cost, security and time related issues. To victual this issues, in this paper we propose a smart remote data backup algorithm i.e. Secure Erasure Coding Algorithm (SEC). The contribution of the proposed SEC is twofold; first SEC 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.

II. RELATED WORK

The recent back-up and recovery techniques that have been developed in cloud computing domain such as PCS, HSDRT, Linux Box, ERGOT, Cold/Hot backup approach etc. Detail reviews convey that none of these techniques are able to provide best performances under all uncontrolled circumstances for example security, cost, redundancy and recovery low implementation complexity in short span of time. But we consider performance of PCS it is comparatively stable, 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 groups in cloud. However, it is unable to control the implementation complexities.

These techniques tried to cover various issues maintaining the cost of implementation as low as possible. Although there is also a technique in which cost increases gradually as data increases i.e. Cold and hot back-up strategy 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. Through 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.

However each one of the backup solution in cloud computing is unable to accomplish all the issues of remote data back-up server. The advantages and problems of all these foresaid techniques are described in the Table-I. Because of 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.

Sno	Approach	Advantage	Disadvantage
1	HSDRT	Used for Movable clients like laptop, smart phone	> Costly > Increase **redundancy**
2	Parity Cloud Service	> Reliable > Privacy > Low cost	> Implementation > Complexity high
3	ERGOT	> Perform exact-match retrieval > Privacy	> Time complexity > Implementation complexity
4	Linux Box	> Simple > Low cost for implementation	> Required higher bandwidth >Privacy > Complete server Backup at a time
5	Cold/ Hot Backup Strategy	> Triggered only when failure detected	> Cost increases as data increases gradually
6	Shared backup router resources(SBR)	> 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	> 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

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

III. REMOTE REPOSITORY/ DATA BACKUP

While we focus on the Backup server of major cloud, we only figure about the copy of major cloud. When the Backup server is at remote location (i.e. far away from the major server) and having the complete state of the major cloud, then this remote location server is termed as Remote Data Backup Server. The major cloud is termed as the central repository and remote backup cloud is termed as remote repository.

Cloud Computing

Cloud computing in general can be portrayed as a synonym for distributed computing over a network, with the ability to run a program or application on many connected computers at the same time. It specifically refers to a computing hardware machine or group of computing hardware machines commonly referred as a server connected through a communication network such as the Internet, an intranet, a local area network or wide area network and individual users or user who have permission to access the server can use the server's processing power for their individual computing needs like to run an application, store data or any other computing need. Therefore, instead of using a personal computer every-time to run the application, the individual can now run the application from anywhere in the world, as the server provides the processing power to the application and the server is also connected to a network via internet or other connection platforms to be accessed from anywhere.

Cloud computing is the allocation of dynamically scalable and often virtualized resources as a services over the internet Users need not have knowledge of, expertise in, or control over the technology infrastructure in the "cloud" that supports them. Cloud computing represents a major change in how we store information and run applications. Instead of hosting apps and data on an individual desktop computer, everything is hosted in the cloud an assemblage of computers and servers accessed via the Internet.

Cloud computing exhibits the following key characteristics:

- **Agility:** It improves with users' ability to re-provision technological infrastructure resources.

- **Cost:** It is claimed to be reduced and in a public cloud delivery model capital expenditure is converted to operational expenditure. This is purported to lower barriers to entry, as infrastructure is typically provided by a third-party and does not need to be purchased for one-time or infrequent intensive computing tasks. Pricing on a utility computing basis is fine-grained with usage-based options and fewer IT skills are required for implementation. The e-FISCAL project's state of the art repository contains several articles looking into cost aspects in more detail, most of them concluding that costs savings depend on the type of activities supported and the type of infrastructure available in-house.

- **Virtualization**: Is a technology allows servers and storage devices to be shared and utilization is increased. Applications can be easily migrated from one physical server to another.

- **Multi tenancy**: It enables sharing of resources and costs across a large pool of users thus allowing for:

- **Centralization:** The centralization of infrastructure in locations with lower costs (such as real estate, electricity, etc.)

- **Utilization and efficiency:** It improvements for systems that are often only 10–20% utilized.

- **Reliability**: It is improved if multiple redundant sites are used, which makes well-designed cloud computing suitable for business continuity and disaster recovery.

- **Performance**: It is monitored and consistent and loosely coupled architectures are constructed using web services as the system interface.

- **Security**: The security in could improve due to centralization of data, increased security-focused resources, etc., but concerns can persist about loss of control over certain sensitive data, and the lack of security for stored kernels. Security is often as good as or better than other traditional systems, in part because providers are able to devote resources to solving security issues that many customers cannot afford. However, the complexity of security is greatly increased when data is distributed over a wider area or greater number of devices and in multi-tenant systems that are being shared by unrelated users. In addition, user access to security audit logs may be difficult or impossible. Private cloud installations are in part motivated by users' desire to retain control over the infrastructure and avoid losing control of information security.

- **Maintenance**: The managing of cloud computing applications is easier, because they do not need to be installed on each user's computer and can be accessed from different places.

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.

Remote Data Backup Server

The major cloud is termed as the central repository and remote backup cloud is termed as remote repository. 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 major 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 major cloud.

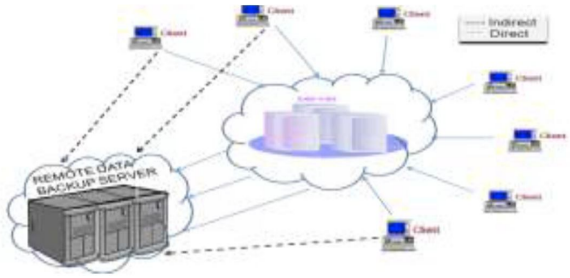

Fig.1 Remote data Backup Server and its Architecture

The Remote backup services should cover the following issues:

- 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.

- 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.

- 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.

- 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.

- 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. SECURE ERASURE CODE ALGORITHM

As discussed earlier low cost, low implementation complexity, security and time related issues are still challenging in the field of cloud computing. To tackle these issues we p r o p o s e SEC algorithm.

Secure Erasure Coding Algorithm (SEC) Architecture:

Secure Erasure Coding algorithm focuses on simplicity of the back-up and recovery process. It basically uses the theory 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 we expect 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 = $X \oplus B$.

Secure Erasure Coding Algorithm:

The Secure Erasure Coding Algorithm works to provide the simple Back-up and recovery process. Its architecture consists of the Major 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 major 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.

Algorithm:

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 ***Ci /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.

V. CONCLUSION

We propose a smart remote data backup algorithm i.e. Secure Erasure Coding Algorithm for cloud system that supports the users to collect information from any remote location in the absence of network connectivity. We can prove that the output of our algorithm is optimal which means any other solutions would definitely cause larger payment cost. In addition, we analyze the approximation ratio for the expanded execution time generated by our algorithm to the user-expected deadline, under the possibly inaccurate task property prediction. When the resources provisioned are relatively

sufficient, we can guarantee task's execution time always within its deadline even under the wrong prediction about task's workload characteristic.

Secure Erasure Coding Algorithm 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. It 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 Secure Erasure Coding Algorithm 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] S. Zhang, X. Chen, and X. Huo, 2010, "Cloud Computing Research and Development Trend," IEEE Second International Conference on Future Networks, pp. 93-97.

[3] P.Demeester et al., 1999, "Resilience in Multilayer Networks," IEEE Communications Magazine, Vol. 37, No. 8, p.70-76.

[4] Wayne A. Jansen, 2011, "Cloud Hooks: Security and Privacy Issues in Cloud Computing, 44th Hawaii International Conference on System Sciences. Hawaii.

[5] 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.

AUTHOR PROFILE

Kolipaka Kiran, pursuing M.Tech (CSE) from Holy Mary Institute of Technology and Science, Keesara, Ranga Reddy Dist., Affiliated to JNTU-HYDERABAD.

Janapati Venkata Krishna, Associate Professor & H O D (CSE Department), Holy Mary Institute Of Technology and Science, Keesara, Ranga Reddy Dist., Affiliated to JNTU-HYDERABAD.