AGILE DATA BACK-UP TECHNIQUE TO IMPROVE DATA SECURITY IN CLOUD COMPUTING

Kesana Praveen[1], Paparao Rapuri[2], Betam Suresh[3]

[1]M.Tech (CSE) Scholar, Vikas Group of Institutions, Nunna, Vijayawada, A.P, (India)
[2]Asst. Professor, Department of CSE, Vikas Group of Institutions, Nunna, Vijayawada, A.P, (India)
[3]HOD, Vikas Group of Institutions, Nunna, Vijayawada, A.P, (India)

ABSTRACT

In today's modern technology everything is possible in terms of storage of data and how to use of that data securely. In cloud computing, data generated in electronic form are large in amount. In cloud the data is used to store, manage and process in a network which is located rather than a local server or a desktop system. To manage this data effectively, there is an essential of data recovery services while working with network transactions. To handle this we propose a technique that, partitioning of data and then making it store at different clouds. In this paper we propose a smart data back-up technique. The main purpose of the proposed technique is to help the users to collect information from any remote location while the required network connectivity was not present or loss due to network problems. The proposed technique solves the time related problems so that the time taken to recovery a data back-up from network failure is minimum. The proposed technique focuses on the data security concept for the back-up files stored at remote server, and it doesn't dependents' on the present 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: Smart Data Back-Up, Remote Server, Data Security, Data Backup, Data Recovery.

I. INTRODUCTION

Cloud computing is a term used to refer to a model of network computing where a program or application runs on a connected server or servers rather than on a local computing device such as a PC, tablet or smart phone. Like the traditional client-server model or older mainframe computing, a user connects with a server to perform a task. The difference with cloud computing is that the computing process may run on one or many connected computers at the same time, utilizing the concept of virtualization. With virtualization, one or more physical servers can be configured and partitioned into multiple independent "virtual" servers, all functioning independently and appearing to the user to be a single physical device. Such virtual servers are in essence disassociated from their physical server, and with this added flexibility, they can be moved around and scaled up or down on the fly without affecting the end user. The computing resources have become "granular", which provides end user and operator benefits including on-demand self-service, broad access across multiple devices, resource pooling, rapid elasticity and service metering capability. Cloud computing services are used both by consumers as well as by organisations and companies. Offers in cloud computing comprise, among other things, the provision of calculating and storage capacity; the provision and operation of development environments and of operating and database management systems; of web hosting; of web mail services; and of a growing number of different types of application software; for word processing and other office applications; customer relationship management; supply chain management; or for the storage and management of photos or personal health related data (electronic health records), to name a few. As number of users and resources are increasing to use the cloud to store and access the data. It is feasible that other customers can access your data. There may be 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 these issues, in this paper we propose a smart remote data backup algorithm with secure data backup. The contribution of the proposed algorithm is twofold; first it 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 literature survey studies previous back-up and recovery techniques that have been developed in cloud computing domain such as Linux Box, PCS, HSDRT,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. The HSDRT is an innovative file backup approach; it 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 receives the data to be backup and in recovery sequence, when some disasters occurs or periodically, the supervisory server starts the recovery sequence. However there are some limitation in this model and therefore, this model is somehow unable to declare as perfect solution for backup and recovery. In addition, Linux Box model is having very simple concept of data backup technique and recovery with very low cost. However, in this model protection level is very low cost. It also makes the process of migration from one cloud service provider to other very easy. This approach will remove consumer's dependency on the ISP and its associated backup cost. The main drawback of this approach is that a consumer can backup not only the data but synchronous the entire server which somehow waste the bandwidth because every time when backup takes place it will do backup of entire cloud server or virtual machine. 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.

S no	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. DATA BACKUP/ DATA RECOVERY

In cloud computing the legal responsibility for data processing is borne by the user, who enlists the services of a cloud service provider. The user is the data collector. As in all other cases in which a third party is given the task of processing personal data, the user or data controller is responsible for ensuring that data protection requirements are met. This applies to consumers (for example, if they use a web mail service or manage their photos over the internet), and to companies and organizations (using, for example, the solution of a cloud service provider for the CRM). The data collector in Switzerland who wants to enlist the services of a cloud computing provider has to ensure that the data protection requirements are considered binding in its contract with the provider. This can be either by means of individually negotiated clauses or through the security and data protection policy of the provider being declared part of the contract, insofar as these fulfill statutory requirements. Particular attention must be paid to the following points in a contract with a cloud service provider:

3.1 Scope of Processing

The type of data processing permissible by the provider is to be clearly specified, and the purpose for which the data may be processed.

3.2 Subcontractors

The conditions under which the provider may for his part pass on the data to subcontractors have to be defined, for example, to a provider of storage capacity. It must be ensured that the user is informed to which subcontractor data is forwarded and that the regulations regarding data protection in the contract between the user and the provider are also binding on subcontractors.

3.3 Deletion of Data

An essential point is that data that has to be deleted by the user because he or she no longer needs it or may no longer process it for another reason is also deleted by the provider and no more copies of data are available. This can lead to problems, in particular in connection with backups that are created by the provider if these contain data belonging to a number of his customers and targeted deletion of individual data items proves financially unreasonable or technically inappropriate in terms of feasibility. Data deletion is also of prime importance when terminating the contract with the provider.

3.4 Data Security Measures

The organizational and technical data security measures that are to be taken by the provider are to be stipulated in the contract, such as the access rights of the provider's employees to data and the systems used to process them, or the encryption of data during transmission or storage, or both.

3.5 Localization of Data

To enable fulfillment of the requirements in connection with the export of data, the customer must know in which countries the servers are deployed on which the data is processed and stored and the provider is to be under an obligation not to transfer the data to any other countries without prior consultation with the user.

3.6 Service Level Agreements

According to the purpose for which the data is processed it is important to agree on binding service levels for availability and data recovery and if necessary, safeguarded by supporting fixed penalties in the event of non-compliance with the agreed service levels.

3.7 Restitution of Data

Upon termination of the contract, the orderly return of data to the user has to be ensured. This requires sufficiently long periods of notice for the user to be able to take the necessary measures to ensure the availability and constant further processing of data after termination of the contract. The form in which the data is to be delivered to the user by the provider must also be ascertained.

3.8 Audits

By agreeing on information and audit rights, the user establishes the opportunity to verify that the obligations entered into by the provider are being fulfilled. Depending on the sector to which the user belongs, such rights also have to be provided for auditing companies and regulatory authorities to whose control the user is subject.

3.9 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.

3.10 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.

3.11 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.

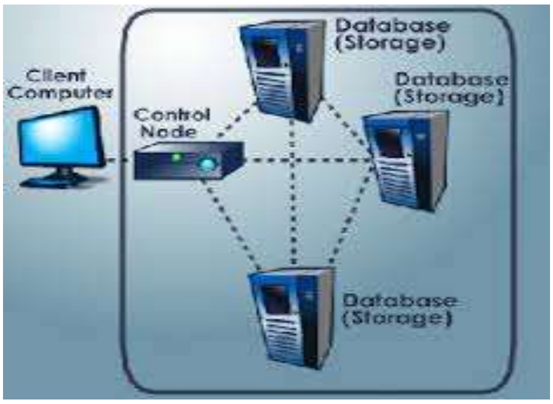

Fig.1 Remote Data Backup Server and Its Architecture

The Cloud Security Issues should cover the following issues:

The benefits of cloud adoption are numerous, including improved efficiency, reduced costs and greater accessibility and flexibility. Cloud computing is one of the fasted growing segments of the IT industry. However, as more information on individuals and companies is placed in the cloud, companies must address cloud computing security issues. As with other major business decisions, an enterprise must evaluate the benefits and be prepared to address any risks and challenges cloud adoption brings. Moving applications to the cloud and accessing the benefits means first evaluating specific data security issues and cloud security issues.

When enterprises move applications from on-premise to cloud-based, challenges arise from data residency, industry compliance requirements, and privacy and third party obligations concerning the treatment of sensitive data. Corporate policies or the regulations of the governing jurisdictions impact the way sensitive data is managed including where it is located, what types of data can be collected and stored and who has access to it. These issues can determine the degree to which organizations can realize the value of cloud computing.

Cloud security issues fall primarily into three areas:

Data Residency - Many companies face legislation by their country of origin or the local country that the business entity is operating in, requiring certain types of data to be kept within defined geographic borders. There are specific regulations that must be followed, centered around data access, management and control.

Data Privacy - Business data often needs to be guarded and protected more stringently than non-sensitive data. The enterprise is responsible for any breaches to data and must be able ensure strict cloud security in order to protect sensitive information.

Industry & Regulation Compliance - Organizations often have access to and are responsible for data that is highly regulated and restricted. Many industry-specific regulations such as GLBS, CJIS, ITAR and PCI DSS, require an enterprise to follow defined standards to safeguard private and business data and to comply with applicable laws.

IV. PROPOSED ALGORITHM FOR SMART DATA BACK-UP

As discussed earlier low cost, low implementation complexity, security and time related issues are still challenging in the field of cloud computing.

4.1 Smart Data Backup Technique Overview

The Proposed 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 = $X \oplus B$.

4.2 Smart Data Backup Algorithm

The 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 data backup algorithm for cloud system that supports the users to collect information from any remote location even when there is a failure of network connectivity. We can prove that the output of our algorithm is robust and provide optimal solution which means any other solutions would definitely cause larger payment cost. 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.

Smart Data Backup 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 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.

AUTHORS PROFILE

	Kesana Praveen, pursuing M.Tech(CSE) from Vikas Group of Institutions, Nunna, Vijayawada. Affiliated to JNTU-Kakinada, A.P., India
	Paparao Rapuri, working as an Asst. Professor of CSE department at Vikas Group of Institutions, Nunna, Vijayawada, Affiliated to JNTU-Kakinada, A.P., India
	Betam Suresh, is working as an HOD, Department of Computer science Engineering at Vikas Group of Institutions, Nunna, Vijayawada, Affiliated to JNTU-Kakinada, A.P., India