REMOTE RESOURCE MAINTENANCE WITH DATA INTEGRITY BASED ON CODE REGENERATION SCHEME

S. Nancy Priya[1], D.Elavarasi[2]

[1, 2] *Dept of CSE, Mount Zion College of Engineering and Technology,*

Anna University, Chennai, (India)

ABSTRACT

With cloud computing, users can remotely store their data into the cloud and use on-demand high-quality applications. Data outsourcing: users are relieved from the burden of data storage and maintenance When users put their data (of large size) on the cloud, the data integrity protection is challenging enabling public audit for cloud data storage security is important Users can ask an external audit party to check the integrity of their outsourced data. Purpose of developing data security for data possession at un-trusted cloud storage servers we are often limited by the resources at the cloud server as well as at the client. Given that the data sizes are large and are stored at remote servers, accessing the entire file can be expensive in input output costs to the storage server. Also transmitting the file across the network to the client can consume heavy bandwidths. Si0nce growth in storage capacity has far outpaced the growth in data access as well as network bandwidth, accessing and transmitting the entire archive even occasionally greatly limits the scalability of the network resources. Furthermore, the input output to establish the data proof interferes with the on-demand bandwidth of the server used for normal storage and retrieving purpose. The Third Party Auditor is a respective person to manage the remote data in a global manner.

Keywords: Cloud Computing, Third Party Auditor (TPA), Cloud Servers, Functional Minimum Storage Regeneration, Proxy Cloud.

I. INTRODUCTION

Cloud Computing has been envisioned as the next-generation architecture of IT enterprises. Enterprise data typically resided in the organization's physical infrastructure, on its own servers in the enterprise's data center, where one could segregate sensitive data in individual physical servers. Today, with virtualization and the cloud, data may be under the organization's logical control, but physically reside in infrastructure owned and managed by another entity. It also provides long list of unprecedented advantages in the IT history such as on-demand self-service, ubiquitous network access, location independent resource pooling, rapid resource elasticity, usage-based pricing and transference of risk. It also enables the enterprises to become more agile and introduce new business models, provide more services, and reduce IT costs.

Being a disruptive technology with profound implications, Cloud Computing provides a paradigm for the data that are being centralized or outsourced into the Cloud. From users perspective, including both individuals and enterprises, storing data remotely into the cloud in a flexible on-demand manner brings appealing benefits such as relief of the burden for storage management, universal data access with independent geographical locations, and avoidance of capital expenditure on hardware, software, and personnel maintenances etc. While these advantages of using clouds are unarguable, due to the opaqueness of the Cloud as separate administrative entities, the internal operation details of cloud service providers (CSP) may not be known by cloud users. Thus maintaining control over the data is paramount to cloud success.

Also, the correctness of the data in the cloud is being put at risk due to the following reasons. First of all, although the infrastructures under the cloud are much more powerful and reliable than personal computing devices, they are still facing the broad range of both internal and external threats for data integrity. Examples of outages and security breaches of noteworthy cloud services appear from time to time. Secondly, for their own benefits, there do exist various motivations for cloud service providers to behave unfaithfully towards the cloud users regarding the status of their outsourced data. Examples include cloud service providers, for monetary reasons, reclaiming storage by discarding data that has not been or is rarely accessed or even hiding data loss incidents so as to maintain a reputation. In short, although outsourcing data into the cloud is economically attractive for the cost and complexity of long-term large-scale data storage, it does not offer any guarantee on data integrity and availability. This problem, if not properly addressed, may impede the successful deployment of the cloud architecture.

Although cloud computing certainly gives organizations with significant cost savings and operational efficiencies, it also brings new security risks and uncertainties. The increased attack surface in a Cloud environment allows for other vulnerabilities to be exploited, thereby increasing the organization's risk. The risk is defined as a given threat that exploits vulnerabilities of an asset or group of assets and thereby cause harm to the organization. The increased attacks in cloud environment, virtual switches and hypervisor that are not present in the traditional data center, allows for other vulnerabilities to be exploited, thereby increasing the organization's risk. The most important threats facing cloud computing are identified as follows:

- Data breaches
- Data Loss/Leakage
- Account or Service hijacking
- Insecure Application Programming Interfaces APIs
- Malicious insiders
- Unknown risk Profile
- Cloud abuse
- Shared Technology Issues
- Changes the business model

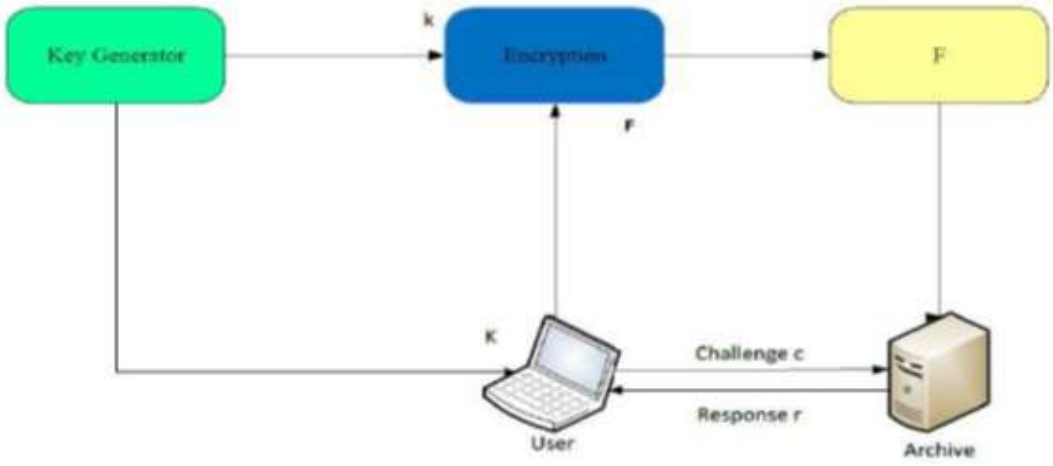

Figure.1 Schematic View of a Proof of Retrievability Based On Inserting Random Sentinels in the Data File

Data security is also an issue in cloud. Since the users no longer physically possess the storage of their data, traditional cryptographic primitives for the purpose of data security protection cannot be directly adopted. Thus, how to efficiently verify the correctness of outsourced cloud data without the local copy of data files becomes a big challenge for data storage security in Cloud Computing. Note that simply downloading the data for its integrity verification is not a practical solution due to the expensiveness in I/O cost and transmitting the file across the network. Besides, it is often insufficient to detect the data corruption when accessing the data, as it might be too late for recover the data loss or damage. The files are encrypted to preserve the data integrity.

The Repair operation in Multiple Cloud Storage occurs when the system or server fails. During Transient failure, the cloud may fail and will return to normal state after some time and no outsourced data will be lost. But during Permanent failure the outsourced data on a failed cloud will become permanently unavailable.

Examples of causes of system failure are

- Data center outages in disasters.
- Data loss and corruption.
- Malicious attacks

This recovery operation creates traffic when it tries to repair the data from multiple servers. The regeneration scheme addressed in this project will reduce the cost of repairing the traffic greatly from the earlier regeneration schemes.

II. LITERATURE SURVEY

2.1. Cloud Computing Security Threats and Responses

The IT organizations have expresses concerns about critical issues (such as security) that exist with the widespread implementation of cloud computing. These types of concerns originate from the fact that data is stored remotely from the customer's location; in fact, it can be stored at any location. Security, in particular [1], is one of the most argued-about issues in the cloud computing field; several enterprises look at cloud computing

warily due to projected security risks. The risks of compromised security and privacy may be lower overall, however, with cloud computing than they would be if the data were to be stored on individual machines instead of in a so - called "cloud" (the network of computers used for remote storage and maintenance). Comparison of the benefits and risks of cloud computing with those of the status are necessary for a full evaluation of the viability of cloud computing. Consequently, some issues arise that clients need to consider as they contemplate moving to cloud computing for their business. There are solutions to solve reliability, availability, and security issues for cloud computing (RAS issues).

2.2 Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing

This keynote paper: In Cloud Computing moves the application software and databases to the centralized large data centers, where the management of the data and services may not be fully trustworthy. This unique paradigm brings about many new security challenges, which have not been well understood. This paper addressed the problem of ensuring the integrity of data storage in Cloud Computing. In particular, this paper considered the task of allowing a third party auditor (TPA), on behalf of the cloud client, to verify the integrity of the dynamic data stored in the cloud. The introduction of TPA eliminates the involvement of the client through the auditing of whether his data stored in the cloud is indeed intact, which can be important in achieving economies of scale for Cloud Computing. In particular [2], to achieve efficient data dynamics, the paper improves the existing proof of storage models by manipulating the classic Merkle Hash Tree construction for block tag authentication. The efficient handling of multiple auditing tasks explore the technique of bilinear aggregate signature to extend the result into a multi-user setting, where TPA can perform multiple auditing tasks simultaneously. Extensive security and performance analysis show that the proposed schemes are highly efficient and provably secure.

2.3 Analysis and construction of functional regenerating codes with uncoded repair for distributed storage systems

The distributed storage systems apply redundancy coding techniques to stored data. One form of redundancy is based on regenerating codes, which can minimize the repair bandwidth, i.e., the amount of data transferred when repairing a failed storage node. Existing regenerating codes mainly require surviving storage nodes encode data during repair. This paper [4] showed the functional minimum storage regenerating (FMSR) codes, which enable uncoded repair without the encoding requirement in surviving nodes, while preserving the minimum repair bandwidth guarantees and also minimizing disk reads. Under double-fault tolerance settings, they formally prove the existence of FMSR codes, and provide a deterministic FMSR code construction that can significantly speed up the repair process.

2.4 A Review of Approaches to Achieve Data Storage Correctness in Cloud Computing Using Trusted Third Party Auditor

In this approach [5], cloud computing is to avail all the resources at one place in the form a cluster and to perform the resource allocation based on request performed by different users. They defined the user request in the form of requirement query. Cloud Computing devices being able to exchange data such as text files as well as business information with the help of internet. Technically, it is completely distinct from an infrared. Using new models Iaas, Paas and Saas. The transmission and storage of large amounts of information, and become

propulsion of fiber-optic accelerating towards 40G/100G. Its foreground is to provide secure, quick, convenient data storage and net computing service centered by internet.

2.5. NCCloud: A Network-Coding-Based Storage System in a Cloud-of-Clouds

In this paper [6], to provide fault tolerance for cloud storage to stripe data across multiple cloud vendors. However, if a cloud suffers from a permanent failure and loses all its data, it is necessary to repair the lost data with the help of the other surviving clouds to preserve data redundancy. This paper presented a proxy-based storage system for fault-tolerant multiple-cloud storage called NCCloud, which achieves cost-effective repair for a permanent single-cloud failure. NCCloud is built on top of a network-coding-based storage scheme called the functional minimum-storage regenerating (FMSR) codes, which maintain the same fault tolerance and data redundancy as in traditional erasure codes (e.g., RAID-6), but use less repair traffic and hence incur less monetary cost due to data transfer. FMSR codes validation provide significant monetary cost savings in repair over RAID-6 codes, while having comparable response time performance in normal cloud storage operations such as upload/download.

III. SYSTEM DESIGN

3.1 System Architecture

Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations. As storage-outsourcing services and resource-sharing networks have become popular, the problem of efficiently proving the integrity of data stored at untrusted servers has received increased attention. In the provable data possession (PDP) model, the client preprocesses the data and then sends it to an untrusted server for storage, while keeping a small amount of Meta data. The client later asks the server to prove that the stored data has not been tampered with or deleted. The integrity of the cloud infrastructure is ensured through the use of Trusted Computing. In addition, we advocate the seamless extension of control from the enterprise into the cloud through the powerful combination of high-assurance remote server integrity, and cryptographic protocols supporting computation on cipher text. With our approach, content is protected in a manner consistent with policies, whether in the enterprise or the cloud.

1. Database
2. Data Owner
3. Database Server
4. Third Party Auditor(TPA)

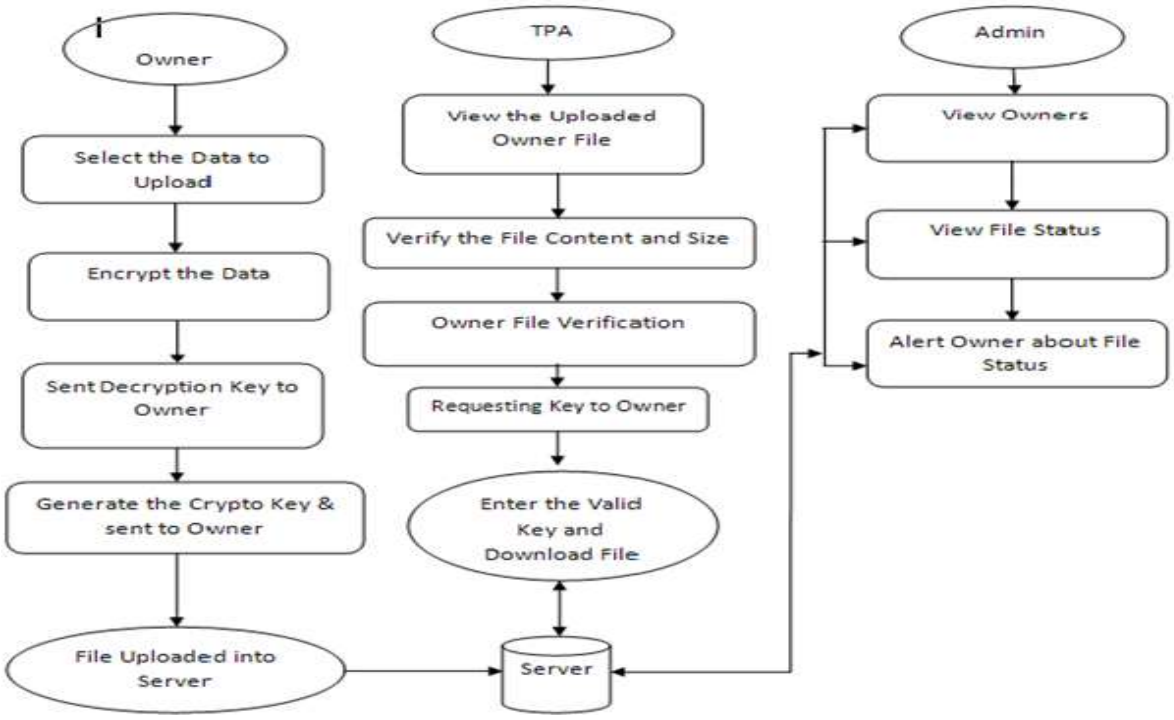

Fig.2 System Architecture

3.1.1 Database

A database is a system intended to organize, store, and retrieve large amounts of data easily. It consists of an organized collection of data for one or more uses, typically in digital form. One way of classifying databases involves the type of their contents, for example: bibliographic, document-text, statistical. Digital databases are managed using database management systems, which store database contents, allowing data creation and maintenance, and search and other access. Database architecture consists of three levels like External, Conceptual and Internal. The external level defines how users understand the organization of the data. A single database can have any number of views at the external level. The internal level defines how the data is physically stored and processed by the computing system. Internal architecture is concerned with cost, performance, scalability and other operational matters. The conceptual is a level of indirection between internal and external. It provides a common view of the database that is uncomplicated by details of how the data is stored or managed, and that can unify the various external views into a coherent whole.

3.1.2 Data Owner

Data owner refers to both the possession of and responsibility for information. Data Owner implies power as well as control. The control of information includes not just the ability to access, create, modify, package, derive benefit from, sell or remove data, but also the right to assign these access privileges to others.

3.1.3 Database Server

It is a computer program that provides database services to other computer programs or computers, as defined by the client–server model. The term may also refer to a computer dedicated to running such a program. Database management systems frequently provide database server functionality, and some DBMSs (e.g., MySQL) rely exclusively on the client–server model for database access. Such a server is

accessed either through a "front end" running on the user's computer which displays requested data or the "back end" which runs on the server and handles tasks such as data analysis and storage.

3.1.4 Third Party Auditor (TPA)

In this module, Auditor views the all user data and verifying data. Auditor directly views all user data without key. Admin provided the permission to auditor. After auditing data, store to the cloud. In the cloud, application and services move to centralized huge data center and services and management of this data may not be trustworthy, into cloud environment the computing resources are under control of service provider and the third party auditor ensures the data integrity over out sourced data. Third party auditor not only read but also may be change the data. Therefore a mechanism should be provided to solve the problem. . TPA checks the integrity of data on cloud on the behalf of users, and it provides the reasonable way for users to check the validity of data in cloud.

IV. MODULES

1. Privacy Safeguard Community Auditing
2. Batch Auditing
3. Data Dynamics
4. Simply Archives
5. Sentinels
6. Verification Phase

4.1 Privacy Safeguard Community Auditing

Homomorphic authenticators are unforgettable verification metadata generated from individual data blocks, which can be securely aggregated in such a way to assure an auditor that a linear combination of data blocks is correctly computed by verifying only the aggregated authenticator. Overview to achieve privacy-preserving public auditing, we propose to uniquely integrate the homomorphic authenticator with random mask technique. In our protocol, the linear combination of sampled blocks in the server's response is masked with randomness generated by a pseudo random function (PRF).

The proposed scheme is as follows:

a. Setup Phase
b. Audit Phase

4.2 Batch Auditing

With the establishment of privacy-preserving public auditing in Cloud Computing, TPA may concurrently handle multiple auditing delegations upon different users' requests. The individual auditing of these tasks for TPA can be tedious and very inefficient. Batch auditing not only allows TPA to perform the multiple auditing tasks simultaneously, but also greatly reduces the computation cost on the TPA side.

4.3 Data Dynamics

Hence, supporting data dynamics for privacy-preserving public risk auditing is also of paramount importance. Now we show how our main scheme can be adapted to build upon the existing work to support data dynamics, including block level operations of modification, deletion and insertion. We can adopt this technique in our design to achieve privacy-preserving public risk auditing with support of data dynamics.

4.4 Simply Archives

This problem tries to obtain and verify a proof that the data that is stored by a user at remote data storage in the cloud (called cloud storage archives or simply archives) is not modified by the archive and thereby the integrity of the data is assured. Cloud archive is not cheating the owner, if cheating, in this context, means that the storage archive might delete some of the data or may modify some of the data. While developing proofs for data possession at untrusted cloud storage servers we are often limited by the resources at the cloud server as well as at the client.

4.5 Sentinels

In this scheme, unlike in the key-hash approach scheme, only a single key can be used irrespective of the size of the file or the number of files whose retrievability it wants to verify. Also the archive needs to access only a small portion of the file F unlike in the key-has scheme which required the archive to process the entire file F for each protocol verification. If the prover has modified or deleted a substantial portion of F, then with high probability it will also have suppressed a number of sentinels.

4.6 Verification Phase

The verifier before storing the file at the archive pre-processes the file and appends some Meta data to the file and stores at the archive. At the time of verification the verifier uses this Meta data to verify the integrity of the data. It is important to note that our proof of data integrity protocol just checks the integrity of data i.e. if the data has been illegally modified or deleted. It does not prevent the archive from modifying the data.

V. CONCLUSION

The proposed project provides data integrity to the cloud; the TPA performs direct and downloaded verification to verify the data integrity in cloud. This project also proposed a framework for TPA to perform batch auditing to audit multiple requests from different data owners. This project enhanced the feature of auditing high risk based audit request in batch auditing and proposed a regeneration scheme called Functional Minimum Storage Regeneration code to regenerate or recover or to repair the corrupted data on the servers by the use of proxy server. Hence the system proves the data integrity in extreme level as well as preserving the privacy of user's data. This proposed scheme ensures security by intimating the details to the data owner about the status of the remote file whether it is modified or not. Batch auditing is performed to handle multi-user requests. The risky files that are requesting for audit are audited in batch auditing prior to other files to ensure the efficiency and performance. These proposed schemes will shed light on economy of scale for Cloud Computing.

VI. FUTURE WORK

For future work to provide data integrity we will try to improve the encryption technique provided during the file uploading process. We will seek more efficient regeneration schemes to recover the corrupted data from the cloud and will also try to implement efficient methods to greatly reduce the repair traffic during the recovery of resources.

VII. ACKNOWLEDGMENT

We would like to sincerely thank Assistant Prof. D.Elavarasi for his advice and guidance at the start of this article. His guidance has also been essential during some steps of this article and his quick invaluable insights have always been very helpful. His hard working and passion for research also has set an example that we would like to follow. We really appreciate his interest and enthusiasm during this article. Finally we thank the Editor in chief, the Associate Editor and anonymous Referees for their comments.

REFERENCES

[1] Farzad Sabahi Student Member, IEEE, "Cloud Computing Security Threats and Responses", Mar 2006.

[2] Qian Wang, Student Member, IEEE, Cong Wang, Student Member, IEEE, KuiRen, Member, IEEE,Wenjing Lou, Senior Member, IEEE, and Jin Li "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing",Dec 2008.

[3] Yuchong HuStudent Member, IEEE, Lee, P.P.C. Student Member, IEEE; Shum, K.W,"Analysis and construction of functional regenerating codes with uncoded repair for distributed storage systems", Jan 2009.

[4] H. Abu-Libdeh, L. Prince House, and H. Weather spoon. RACS: A Case for Cloud Storage Diversity. In Proc. of ACM SoCC, 2010.

[5] Patel.HStudent Member, IEEE,"A Review of Approaches to Achieve Data Storage Correctness in Cloud Computing Using Trusted Third Party Auditor", Mar 2010.

[6] Henry C. H. Chen Student Member, IEEE, Yuchong Hu, Patrick P. C. Lee Student Member, IEEE, and Yang Tang Student Member, IEEE,"NCCloud: A Network-Coding-Based Storage System in a Cloud-of-Clouds",Jun 2011.

[7] Cong Wang, Student Member, IEEE, Sherman S.-M. Chow, Qian Wang, Student Member, IEEE, KuiRen, Member, IEEE, and Wenjing Lou, Member, IEEE"Privacy-Preserving Public Auditing for Secure Cloud Storage", Jun 2012.

[8] G. Ateniese, R. Di Pietro, L. V. Mancini, and G. Tsudik, "Scalable and Efficient Provable Data Possession" In Proc of Secure Comm, 2008.

[9] R. Curtmola, O. Khan, R. Burns, and G. Ateniese. MR-PDP: Multiple-Replica Provable Data Possession. In Proc. of IEEE ICDCS, 2008.

BIOGRAPHY

	**S.Nancy Priya** is currently a PG scholar in Computer Science Engineering from the Department of Computer Science and Engineering at Mount Zion College of Engineering and Technology, Pudukkottai. She received his Bachelor Degree in Computer Science Engineering from Vel Tech Engineering College, Chennai and Tamilnadu. Her Research areas include Cloud computing, grid computing and distributed system.
	**D.Elavarasi** is currently working as an Assistant Prof. from the Department of Computer Science and Engineering at Mount Zion College of Engineering and Technology, Pudukkottai. She received his Bachelor Degree from Bharath Niketan Engineering College, Andipatti and Theni District. She received his master degree from Anna University, Thiruchirappalli and Tamilnadu. Her main research interests lie in the area of cloud computing and wireless sensor networks.