A NOVEL APPROACH FOR PRIVACY PRESERVING PUBLIC AUDITING TO CLOUD DATA STORAGE WITH REGENERATING –CODE BASED APPROACH

GODAVARTHI DEEPTHI [#1] , K.VENKATA RAO [#2]

[#1] M.Tech Scholar,Department of Computer Science and Systems Engineering, College of Engineering,Andhra University,Visakhapatnam,AP,India.

[#2] Professor,Department of Computer Science and Systems Engineering, College of Engineering,Andhra University,Visakhapatnam,AP,India

ABSTRACT

Cloud Computing is one of the practices of using a network of remote servers hosted on internet to store, access, and retrieve data from remote machines not from local machines. As the cloud is used mainly for storing the data on remote servers it has various services like PaaS, IaaS, SaaS, DaaS and so on. Here each and every service has its own advantages and limitations due to its hardware and software usage. In this cloud the users who are placing the data in the cloud are known as Data Owners and the user who is accessing that file is known as Data users. As the data is been placed on remote system not on our local machines data integrity plays a very important role for both data owners and data users, these two users need to have audit for the cloud data without retrieving the entire data. The data audit should be commonly applied for both public and private cloud users. Till now there are many audit mechanisms in the cloud servers which were not at all achieved total data integrity as they have failed in some instances. Now a day's one of the new technique like regenerating codes have gained more popularity due to their low repair bandwidth which has fault tolerance as its inbuilt property. In the existing remote methods for regenerating the coded data they use to stay online all the time and they used to handle auditing. But now there is no need to stay in the online while conducting the data auditing the data stored by owners and users. In this paper, we have designed a novel public auditing scheme for the regenerating-code-based cloud storage. In order to solve or overcome the problem that occur with the failed authenticators at the time of data owner absence, we have introduced a new layer in our proposed model like a proxy, which is a privilege to regenerate the authenticators, into the traditional public auditing system model. Along with this we have proposed a public verifiable authenticator, which is formed by combining a couple of keys and in turn regenerated using a partial keys.As an extension we have also implemented the proposed concept on a live cloud Server like DROPBOX by creating a public account for data storage and data auditing. By conducting various experiments on this proposed new approach we finally came to a conclusion that this proposed cloud approach have more effectiveness and when auditing shared data integrity.

Key Words: Encryption, Cloud Server, Regenerating Codes, Data Audit, Proxy Server, Authenticators.

I. INTRODUCTION

Now a day's there was a lot of user's attention towards the cloud data storage as well as retrieving of data from the cloud server. As the data is been increasing day by day almost all the companies are unable to store their valuable data on their own individual devices, so in this situation they opt for a new data storage area known as Cloud Data Storage [1], [2]. Generally cloud service providers allow the users to access their services for a low economical and ascendable marginal cost compared with primitive data storage services. Generally the data which is stored in the cloud server is mainly used for sharing within the users of same group or between the users of different group with a valid authentication. Some of the best cloud data storage services are as follows: Google Drive, DriveHq Server, DropBox and ICloud. As these all are the best among various types of cloud service providers in which the data can be stored either in public cloud or private cloud, sometimes can be stored in both combine known as Hybrid Cloud.

Figure.1. Represents various types of Cloud Service Providers and Its Applications

From the figure 1, we can clearly find out that there are various cloud service providers that are available in the real time environment that are used for storing various applications like word documents,pdf,excel and many more files. If u look at the above figure u can find out the various cloud service providers like Zip Cloud, Just Cloud, BOX, Google Drive, DROP BOX and a lot more. Of all these we are using DROP BOX as the storage medium for storing the uploaded files in this proposed application.

Generally the data which is stored in cloud servers will be stored in the form of plain text without having any security bounds, so that there is a chance of getting that data damage due to the sudden failure or problem that occurs in hardware, OS or any other software failures. This failure leads to the problem of damage of original content which is present in the cloud server. So in this paper we mainly concentrate on providing data integrity for the cloud data, therefore, the integrity of cloud information ought to be verified thoroughly before any information utilization, like search or computation over cloud information [6].From the below figure 2, we can clearly find out the data integrity in the cloud as follows: Initially the data owner will encrypt the data which is to be stored on the cloud server and then he issues a random keys for the users who reside in their cloud area. Now the data users try to access the files either with valid keys and some users try to access the data illegally by applying fuzzy search request. From figure 1, we can clearly tell that the users who have a valid authentication and access privilege can only access the data files which is stored in the cloud, but the data users who try to access the cloud data with fuzzy search request cant able to retrieve the data request [3], [4], [5].

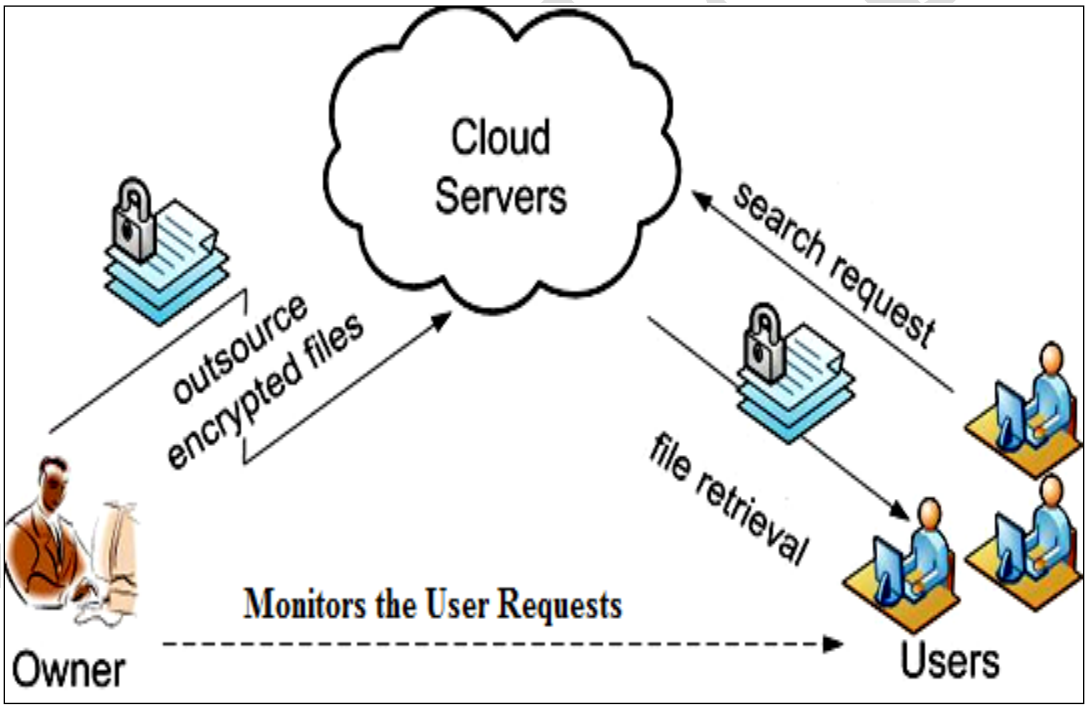

Figure.2. Represents the Process of Providing Data Integrity in the Cloud Server

From the above figure 2, we can able to find a clear view about the representation of data integrity for the data which is stored into the cloud. Here in the above figure, if we see the architecture there are totally three roles like: Owner, Users and Cloud Server. Where the owner is the person who will upload the data into the cloud in an encrypted manner. He will outsource all his valuable and sensitive information files into the cloud server by applying encryption for the data. Once the files are uploaded the user is a person who will try to download the data from the cloud server. If the user is a valid user and he has proper authentication then only he can able to download the file in a decrypted manner.

II. BACKGROUND WORK

In this section we will mainly discuss about the various types of cloud servers and also the traditional approach of data audit. Now let us look about that in detail in this below section:

A) TRADITIONAL DATA HOSTING SERVICES

Generally in some respects cloud servers work in the same way as physical servers but the functions they provide can be somewhat different. When opting for cloud hosting, clients are renting virtual server space rather than renting or purchasing physical servers. They are often paid for by the hour depending on the capacity required at any particular time.

In the traditional hosting of data, there are mainly two processes like:

1. Shared Hosting Approach (&)
2. Dedicated Hosting Approach.

If we look at those two types of hosting in detail: shared hosting is the cheaper option whereby servers are shared between the hosting provider's clients. One client's website will be hosted on the same server as websites belonging to other clients. This has several disadvantages including the fact that the setup is inflexible and cannot cope with a large amount of traffic. Dedicated hosting is a much more advanced form of hosting, whereby clients purchase whole physical servers. This means that the entire server is dedicated to them with no other clients sharing it. In some instances the client may utilize multiple servers which are all dedicated to their use. Dedicated servers allow for full control over hosting. The downside is that the required capacity needs to be predicted, with enough resource and processing power to cope with expected traffic levels. If this is underestimated then it can lead to a lack of necessary resource during busy periods, while overestimating it will mean paying for unnecessary capacity.

Cloud storage can provide the benefits of greater accessibility and reliability; relief of the burden for storage management, universal data access with location independence, and avoidance of capital expenditure on hardware, software, and personal maintenances, etc.,

B) TYPES OF CLOUD STORAGE

Now let us look about the some of the various types of cloud data storage in detail :

i) PERSONAL CLOUD STORAGE

It is also known as mobile cloud storage, personal cloud storage is a subset of public cloud storage that applies to storing an individual's data in the cloud and providing the individual with access to the data from anywhere. It also provides data syncing and sharing capabilities across multiple devices. Apple's iCloud is an example of personal cloud storage. This personal

cloud storage is mainly dealt in various mobile operators which provide easy access for data storage inside the cloud.

II) PUBLIC CLOUD STORAGE

Public cloud storage is where the enterprise and storage service provider are separate and there aren't any cloud resources stored in the enterprise's data center. The cloud storage provider fully manages the enterprise's public cloud storage. Normally in the current days all the cloud storage providers provide public cloud as storage medium with min 2GB and Max 10 GB for data storage with free data access and usage. If that space exceeds then we need to pay the excess storage cost more than 10 GB, which acts as a private cloud.

III) PRIVATE CLOUD STORAGE

It is a form of cloud storage where the enterprise and cloud storage provider are integrated in the enterprise's data center. In private cloud storage, the storage provider has infrastructure in the enterprise's data center that is typically managed by the storage provider. Private cloud storage helps resolve the potential for security and performance concerns while still offering the advantages of cloud storage.

IV) HYBRID CLOUD STORAGE

Hybrid cloud storage is a combination of public and private cloud storage where some critical data resides in the enterprise's private cloud while other data is stored and accessible from a public cloud storage provider. A Hybrid Cloud Storage is the combination of public and private cloud storage, till less than 10 GB it is treated as public cloud and more than 10 GB treats as a private cloud storage. Hence an account which contains both these combine is known as hybrid cloud data storage.

C) TRADITIONAL APPROACH OF DATA AUDIT

The traditional approach for checking data correctness is to retrieve the whole data from the cloud, and then verify data integrity by checking the correctness of signatures (e.g., RSA) or hash values (e.g., MD5) of the entire data. Certainly, this conventional approach is able to successfully check the correctness of cloud data. However, the efficiency of using this traditional approach on cloud data is in doubt [8]. The main reason is that the size of cloud data is large in general. Downloading the entire cloud data to verify data integrity will cost or even waste user's amounts of computation and communication resources, especially when data have been corrupted in the cloud. Besides, many uses of cloud data (e.g., data mining and machine learning) do not necessarily need users to download the entire cloud data to local devices. It is because cloud providers, such as Amazon, can offer users computation services directly on large-scale data that already existed in the cloud.

The main motivation for conducting audit on cloud data is due to the fact like, many mechanisms [9] –[12] have been proposed to allow not only a data owner itself but also a public verifier to efficiently perform integrity checking without downloading the entire data from the cloud, which is referred to as public auditing [5]. During this time the data is divided into a small blocks where each and every block is independently signed by the data owner. A random combination of all the blocks instead of the whole data is retrieved during integrity checking.

Generally a public verifier could be a data user (e.g., researcher) who would like to utilize the owner's data via the cloud or a trusted third-party auditor (TTPA) who can provide expert integrity checking services. Moving a step forward, Wang et al. designed an advanced auditing mechanism [5] (named as WWRL in this paper), so that during public auditing on cloud data, the content of private data belonging to a personal user is not disclosed to any public verifiers. Unfortunately, current public auditing solutions mentioned above only focus on personal data in the cloud [1].

III. PROPOSED RE-GENERATING CODES MODEL FOR PUBLIC AUDITING OF CLOUD DATA STORAGE

In this section we will find out the proposed system that was used in the current paper for public auditing for cloud data storage .This is mainly done by using re-generating codes model that was used in order to re-generate the lost or deleted data by the un-authorized users.This re-generation code is mainly operated and monitored by proxy module which is developed as a extension for the primitive models.

MAIN MOTIVATION

As we already seen the primitive model that was used in real time environment as seen in figure 2, we have motivated with that primitive model and extended the current model with some more extension like proxy layer which is used for re-generating the code if it is deleted by any un-authorized user during the cloud data access.

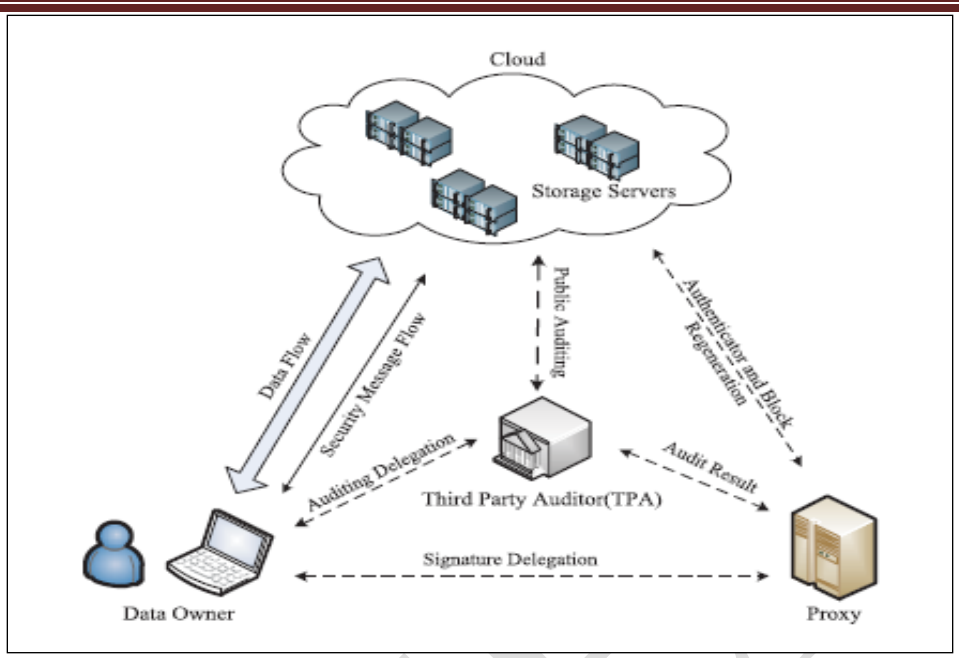

Figure.3. Represents the Process of Proposed Novel Auditing Model for Data Storage

We have proposed a novel auditing model for Regenerating-Code-based cloud storage as shown in Figure. [3], which consist mainly of four blocks like :

1. Data Owner
2. The Cloud Server
3. Proxy Server (&)
4. Third Party Auditor

Now let us look about the proposed model in detail like: Data Owner is the person who tries to upload his valuable data into the cloud server, where he has a large amount of data to be stored in the cloud; the cloud server is one which will provide access to store the valuable data inside its storage area. Also the cloud server is one who provide various cloud services and have a significant computational resources; the third party auditor (TPA) conducts public audits on the coded data in the cloud, its audit results are unbiased for both data owner and cloud servers; and proxy agent, is one who is extended by the current paper apart from the previous audit model with a semi- trusted and acts on behalf of the data owner to regenerate authenticators and data blocks on the failed servers during the repair procedure. Here the data owner need not wait all the time in online as he can logout from his session once after uploading his valuable file into the cloud server. Here the main role of the application is performed by the proxy, as it is supposed to do much more powerful than the data owner but less than the cloud servers in terms of computation and memory capacities which would be always in online.

In order to save the online burden, there is a continuous periodic auditing and accidental repairing that is used to save resources. The data owners resort to the TPA for integrity verification and delegate the reparation to the proxy. As compare to the traditional public auditing system model, our system model involves an additional proxy agent. In order to reveal

the rationality of our design, we consider a scenario: A company employs a commercial regenerating code-based public cloud and provides long-term archival storage service for its staffs, the staffs are equipped with low end computation devices (e.g., Laptop PC, Tablet PC, etc.) and will be frequently off-line. For public data auditing, the company relies on a trusted third party organization to check the data integrity; Similarly, to release the staffs from heavy online burden for data and authenticator regeneration, the company supply a powerful workstation (or cluster) as the proxy and provide proxy reparation service for the staffs' data.

Our proposed novel auditing scheme consists 3 Phases, they are as follows:

1. Setup Phase,
2. Audit Phase,
3. Repair Phase

Now let us look about those phases in details as follows:

1. **SETUP PHASE:** Here the data owner will initially try to initialize the setup phase and try to initialize our proposed auditing scheme. Where the setup phase only needs to initialize all the other attributes that are available in our current paper.

2. **AUDIT PHASE:**

In this phase the cloud servers and TPA interact with one another to take a random sample on the blocks and check the data intactness in this procedure. This phase is known as audit phase where the data owner will upload the data and it will be reached to the auditor level to verify if the content is reached accurately or it is modified by any one. If the auditor gives the approval for the data to be upload with no modification then only it will be uploaded into the cloud server. At this stage only the audit phase plays a vital role in the application for finding the data genuineness or data modified status.

3. **REPAIR PHASE:**

In the absence of the data owner, the proxy interacts with the cloud servers during this procedure to repair the wrong server detected by the auditing process. This phase is mainly included as an extension for the primitive cloud auditing model in order to provide the repair capability for the data which is modified or deleted by the un-authorized users at any end and this has greatly reduced the effort of data owner not to be in online all the time.

IV. PUBLIC AUDITING SCHEME FOR THE REGENERATING-CODE-BASED CLOUD STORAGE

In this section we will find the public auditing scheme for the re-generating code based cloud storage, which greatly reduces the effort of data owner not to stay long time in online throughout the process.

The integrity checking and regeneration (of failed data blocks and authenticators) are implemented by a third-party auditor and a semi-trusted proxy separately on behalf of the data owner.

REGENERATING CODES PRILIMINARIES

The following are the preliminaries that are used for the current re-generating codes process. They are as follows:

a) Reduces Repair Bandwidth -Bandwidth describes the maximum data transfer rate of a network or Internet connection. It measures how much data can be sent over a specific connection in a given amount of time.

b) Cloud storage to be a collection of *n* **storage servers**, **data file** *F* is encoded and stored redundantly across these servers. When data corruption at a server is detected, the client will contact _ **healthy servers** and download β_ bits from each server, thus regenerating the corrupted blocks without recovering the entire original file.

c) The corrupted blocks can be exactly regenerated, there are two versions of repair strategy:

Exact repair and functional repair

- **Exact repair** strategy requires the repaired server to store an exact replica of the corrupted blocks.
- **Functional repair** indicates that the newly generated blocks are different from the corrupted ones with high probability.

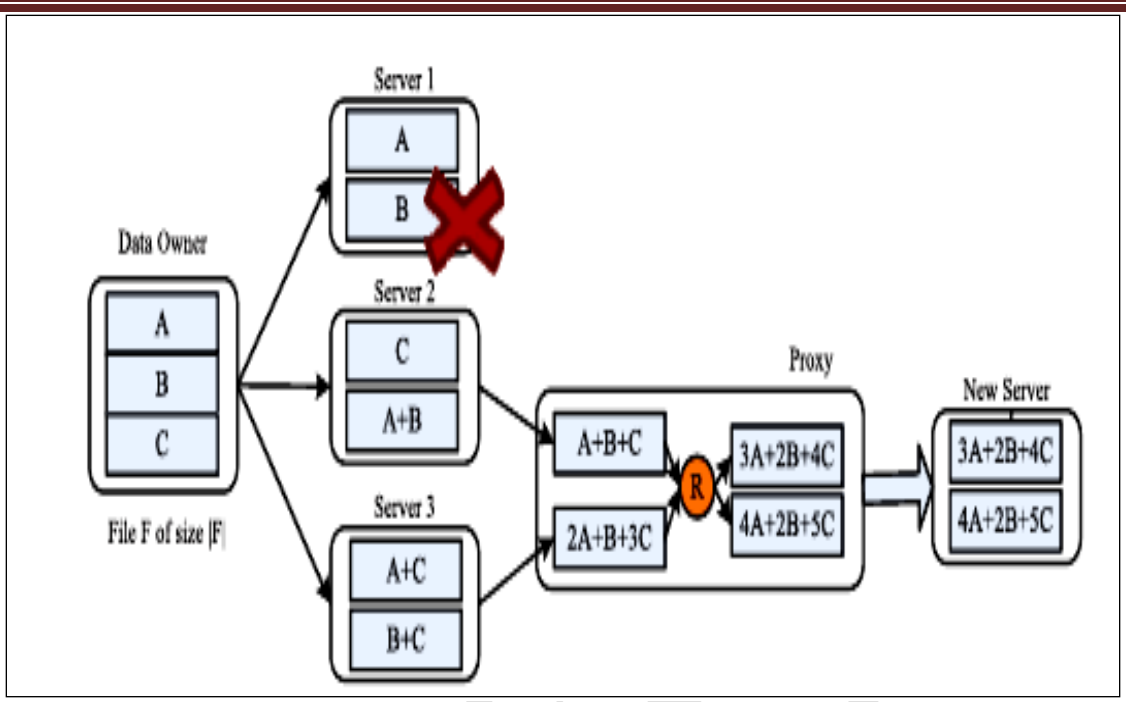

Figure.4. Represents an example of functional repair regenerating code with parameters ($n = 3, k = 2, l = 2, \alpha = 2, \beta = 1$).

The data owner computes six coded blocks as random linear combinations of the native three blocks, and distributes them across three servers. When Server 1 gets corrupted, the proxy contacts the remaining two servers and retrieves one block (obtained also by linear combination) from each, then it linearly combines them to generate two new coded blocks. Finally, the new coded blocks are sent to a new healthy server. The resulting storage system turns out to satisfy the *(*3, 2*)* MDS property.

V. IMPLEMENTATION PHASE

Implementation is the stage where the theoretical design is converted into programmatically manner.In this stage we will divide the application into a number of modules and then coded for deployment. We have implemented the proposed concept on Java programming language with JEE as the chosen language in order to show the performance this proposed re-generating codes model for public auditing of cloud data storage. The front end of the application takes JSP, HTML and Java Beans and as a Back-End Data base we took My SQL data base along with a Real Cloud Service provider called as DROPBOX Cloud Service provider. This cloud service provider will provide a space up to 10 GB for storing the files which is used by the application. The application is divided mainly into following 4 modules. They are as follows:

1. Data Owner Module
2. Third Party Auditor Module

3. Cloud Admin Module
4. Data Users Module

1. Data Owner Module

Data Owner uploads files to group users via Cloud Storage like Drop box. If Owner Upload a file in background that the file will be encrypted using Blowfish Cryptography and validation the Time based one time password and enter in to cloud storage.

Encryption/Decryption Algorithm:

Blowfish is a symmetric block cipher algorithm for encryption/decryption. Blowfish is accepted as a fast and strong encryption algorithm because it has not been cracked. Blowfish is fixed 64 bit block cipher and a takes key length from 32-448bits. Total 16 processing rounds of data encryption is performed in Blowfish. This algorithm is divided into two parts- (i) Key expansion and (ii) Data encryption. In key expansion process 448 bits key is converted into 4168 bytes. The advantages of blowfish algorithm are that it is secure and easy to implement and best for hardware implementation, but the disadvantage is it require more space for the ciphertext because of difference between key size and block size.

Time Based One Time Password

Time-based One-time Password Algorithm (TOTP) is an algorithm that computes a one-time password from a shared secret key and the current time. It has been adopted as Internet standard RFC 6238,[1] is the cornerstone of Initiative For Open Authentication (OATH) and is used in a number of authentication systems.TOTP is an example of a hash-based message authentication code (HMAC). It combines a secret key with the current timestamp using a cryptographic hash function to generate a one-time password. The timestamp typically increases in 30-second intervals, so passwords generated close together in time from the same secret key will be equal.

2. Third Party Auditor Module

In this module ,the TPA Third Party Auditor verify/audits all the uploaded owner files in cloud storage and sends into the cloud admin for final approval. If admin approve the file then only he can able to send the files to a group of users those are available in that cloud server.This TPA acts like a third party to audit the data which is stored to and from the cloud server.

3. Cloud Admin Module

In this module the cloud admin is the one which will verify the auditor approved files after granting the final approval by the TPA.This will be used to store the data inside its memory location after the TPA Verifies the data from its side. Once the cloud owner uploads the data it

will be passed to the TPA for verification,then finally it is uploaded into the cloud server if the file was verified successfully by the TPA.

4. Data Users Module

In this module the users initially register, and after registration he will try to login so that to join any group. Each and Every Group Signature is Unique within the same group. User sign in the group login after download all files and not like this group he will revoke the group membership easily. Each and Every Users activity are tracked by Session Tracking Methodology. If the users try to hack the another users group signature and sign in the group login that time the user system configuration & the Users Details are intimated to cloud admin. Admin block the hack users after the user common username & password are going to invalid mode.

Hackers are stolen my files via routers & hacking approved users password to get our secret key (cross-site). In that time we are using randomized key to prevent the router hackers suppose hackers try to hack the file packets he can't predict the cryptography to decrypt the original content of our files then the password hackers are prevent like this: based on Session Tracking to track the users activity in the users page one more functionality is there that is request for secret key. This page user will get the secret key he must entered his registered password suppose hacker will try to enter approved user password and get the secret key in that time our project will mark the users will be the hacker and sends that system configuration details to cloud admin suppose Admin will block the user in that time users registered username & password change to invalid mode.

VI. CONCLUSION

In this paper, we for the first time implemented a secure re-generating codes model for public auditing of cloud data storage, where the data owners are privileged to delegate TPA for their data validity checking. To protect the original data privacy against the TPA, we randomize the coefficients in the beginning rather than applying the blind technique during the auditing process. Considering that the data owner cannot always stay online in practice, in order to keep the storage available and verifiable after a malicious corruption, we introduce a semi-trusted proxy into the system model and provide a privilege for the proxy to handle the reparation of the coded blocks and authenticators. To better appropriate for the regenerating-code-scenario, we design our authenticator based on the BLS signature method. As an extension we have implemented the proposed algorithm on live cloud server like DROPBOX for showing the performance of the proposed application in terms of data integrity.

VII. REFERENCES

[1] B. Wang, B. Li, and H. Li, "Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud," Proc. IEEE Fifth Int'l Conf.Cloud Computing, pp. 295-302, 2012.

[2] "Top Threats to Cloud Computing v1.0" (PDF). Cloud Security Alliance. Retrieved 2014-10-20.

[3] D. Boneh, B. Lynn, and H. Shacham, "Short Signatures from the Weil Pairing," Proc. Seventh Int'l Conf. Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT'01), pp. 514-532, 2001.

[4] R.L. Rivest, A. Shamir, and Y. Tauman, "How to Leak a Secret," Proc. Seventh Int'l Conf. Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT'01),

[5] "Swamp Computing a.k.a. Cloud Computing". Web Security Journal. 2009-12-28. Retrieved 2010-01-25.

[6] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A. Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A View of Cloud Computing," Comm. ACM, vol. 53, no. 4, pp. 50-58, Apr. 2010.

[7] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable Data Possession at Untrusted Stores," Proc. 14th ACM Conf. Computer and Comm. Security (CCS '07), pp. 598-610, 2007.

[8] K. Yang and X. Jia, "An efficient and secure dynamic auditing protocol for data storage in cloud computing," IEEE Trans. Parallel Distrib. Syst.,vol. 24, no. 9, pp. 1717–1726, Sep. 2013.

[9] Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, "Cooperative provable data possession for integrity verification in multicloud storage," IEEE Trans.Parallel Distrib. Syst., vol. 23, no. 12, pp. 2231–2244, Dec. 2012.

[10] A. G. Dimakis, K. Ramchandran, Y. Wu, and C. Suh, "A survey on network codes for distributed storage," Proc. IEEE, vol. 99, no. 3,pp. 476–489, Mar. 2011.

[11] H. Shacham and B. Waters, "Compact proofs of retrievability," in Advances in Cryptology. Berlin, Germany: Springer-Verlag, 2008,pp. 90–107.

[12] Y. Hu, H. C. H. Chen, P. P. C. Lee, and Y. Tang, "NCCloud: Applying network coding for the storage repair in a cloud-of-clouds," in Proc.USENIX FAST, 2012, p. 21.

[13] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security

VIII .ABOUT THE AUTHORS

GODAVARTHI DEEPTHI is currently pursuing her 2 Years M.Tech in Computer Science and Systems Engineering,College of Engineering,Andhra University,Visakhapatnam,AP,India. Her area of interests includes Networks.

K.VENKATA RAO is currently working as a Professor, in Computer Science and Systems Engineering, College of Engineering, Andhra University, Visakhapatnam, AP, India. He has more than 15 years of experience in teaching field. His research interest includes Image Processing.