ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

A Survey on Privacy Preserving Methods for Regenerating-code-based Cloud Storage

Samruddhi Pode

M.E Computer Engineering, RMD SSOE, Warje, Pune, India

ABSTRACT: To identify the outsourced data whether they are corrupted or not, it is critical to add fault tolerance to the cloud storage. It is also difficult to add data integrity checking and failure reparation to the storage. Recently, the concept of using regenerating codes in the cloud system has gained popularity because of their lower repair bandwidth while providing fault tolerance. There are some existing methods for regenerating-coded data which provide the private auditing, requiring data owners to always stay online and handle auditing, and also repairing, which sometimes impossible. To address the problem of failed authenticators in the absence of data owners, some other existing methodologies introduce a proxy, which have rights to regenerate the authenticators. In this paper, we studied and review the whole idea of preserving the privacy of outsourced data in regenerating-code-based cloud storage.

KEYWORDS: Cloud Storage, regenerating codes, privacy preserving, authenticator regeneration, proxy.

I. INTRODUCTION

Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in third-party data centres. Cloud storage is a model of data storage where the digital data is stored in logical pools. Cloud storing become popular among different sectors of business as well as daily life since it offers a flexible on-demand data outsourcing service. It also provide appealing benefits like relief of the burden for storage management, universal data access with location independence, and avoidance of capital expenditure on hardware and software. It allows users to upload files that could then be accessed over the internet from a different computer, tablet, smart phone or other networked device, by the same user or possibly by other users, after a password or other authentication is provided [1].However, this concept of data hosting service also brings new security threats toward users' data, thus making individuals or organization still feel hesitant. Hence we have to more focus on the different methods by using which one can preserve the privacy of their own data on a cloud.

There are many mechanisms, which are dealing with the integrity of outsourced data, proposed under different system and security models up to now. Hovav Shacham and Brent Watersbuilt [2] suggested a proof-of- retrievability system, where a data storage center convinces a verifier that he is actually storing all of a client's data. They built BLS signatures and secure in the random oracle model, has the shortest query and response of any proof-of-retrievability with public veriability. Their second scheme, which builds elegantly on pseudorandom functions (PRFs) and is secure in the standard model, has the shortest response of any proof-of-retrievability scheme with private variability. Both schemes rely on homomorphic properties to aggregate a proof into one small authenticator value. But these schemes are only for single-server scenario. Considering that files are usually striped and redundantly stored across multi-servers or multi-clouds, O. Rahamathunisa Begam, T. Manjula, T. Bharath Manohar, B. Susrutha [3] explore the idea of Provable data possession (PDP) which is a technique for ensuring the integrity of data in storage outsourcing for multi-clouds with different redundancy schemes, such as replication, erasure codes, and, more recently, regenerating codes.

In this paper, we focus on the integrity verification problem in regenerating-code-based cloud storage. When a cloud suffers from a permanent failure and loses all its data, then we need to repair the lost data from other surviving clouds to preserve data redundancy. Hence, Henry C. H. Chen and Patrick P. C. Lee [4] presented a proxy-based system for multiple-cloud storage called NCCloud (Network Code Cloud). It aims to achieve cost-effective repair for a permanent single-cloud failure. They proposed the implementable design for the functional minimum-storage regenerating code (F-MSR), double fault tolerance is maintained and has the same storage cost as in traditional erasure coding schemes. However, these are designed for private audit, only the data owner is allowed to verify the integrity and repair the faulty servers. The auditing schemes in [5] consider the large size of the outsourced data as well as user's

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

constrained resource capability. But the tasks of auditing and reparation in the cloud can be formidable and expensive for the users. Remote Data Checking scheme in [6] prove the problem that users need to always stay online, which may imply its adoption in practice, especially for long-term archival storage.

When we outsource storage as a service there is possibility of data loss. To provide fault tolerant in cloud storage it is necessary to strip data in multiple cloud server and need to reconstruct the lost data from other existing cloud server to preserve data redundancy is known as repair operation [7]. The repair is performed using the regenerating codes. Main objective is to minimize the repair traffic, since the users will be charge for the outbound data from cloud. Henry C.H. Chen and Patrick P.C. Lee [8] suggested the idea for the protection of outsourced data in cloud storage against corruptions. They studied the problem of remotely checking the integrity of regenerating-coded data against corruptions under a real-life cloud storage setting. Authors designed and implemented a practical data integrity protection (DIP) scheme for a specific regenerating code, while preserving its intrinsic properties of fault tolerance and repair-traffic saves. They demonstrate that remote integrity checking can be feasibly and simply integrated into regenerating codes in practical deployment.

For ensuring the data integrity and to save the users' computation resources as well as online burden, we can propose a public auditing scheme for the regenerating-code-based cloud storage. This public auditing implements the integrity checking and regeneration of failed data blocks and authenticators by a third-party auditor and a semi-trusted proxy separately on behalf of the data owner.

II. RELATED WORK

There are several systems proposed for multiple-cloud storage.

HAIL A High-Availability and Integrity Layer for Cloud Storage places the task of file-integrity checking in the hands of the client or some other trusted, external service and avoids communication among servers. Unlike previous work, which verifies integrity at the level of individual file blocks, HAIL provides assurance at the granularity of a full file [1].

HAIL offers the benefits like:

- *Strong file-intactness assurance*: It enables a set of servers to prove to a client through a challenge-response protocol that a stored file *F* is fully intact such that the client can recover *F* with overwhelming probability.
- *Strong adversarial model:* HAIL protects against an adversary that is *active*. Adversaries can corrupt servers and alter file blocks and mobile, i.e., can corrupt every server over time.
- *Direct client-server communication:* HAIL involves one-to-one communication between a client and servers. Servers need not intercommunicate or even be aware of other servers' existence.

HAIL provides assurance at the granularity of a full file.

Multiple cloud storage

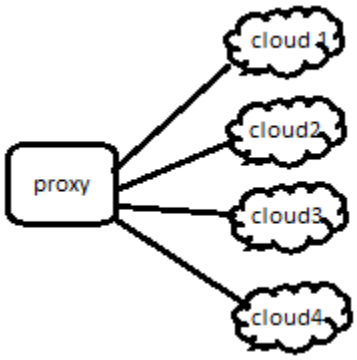
In a storage system the data are stored on a large number of commodity disks and there is a possibility of disk failure. Therefore, it is necessary to store redundant data in storage system in such a way that when a data is lost in certain disk, still the data can be accessed from other disks. For example, the N -way replication, where N replicas are stored in N different disks, can tolerate at most N - 1 disk failures. The simplest form of data redundancy is replication, with one or more copies of an original data and available if that original gets lost.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

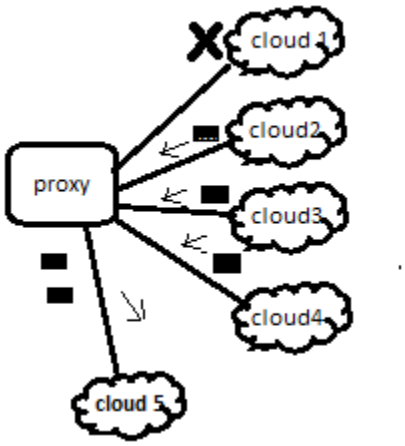
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015



(a) Normal operation



(b) Repair operation

Fig.1: The multiple-cloud storage design that uses proxy server

Provable data possession (PDP) This technique is used to check the availability and integrity of outsourced data in cloud storages. Researchers have proposed two basic approaches called Provable Data Possession and Proofs of Retrievability [2]. Authors proposed a lightweight PDP scheme which is based on cryptographic hash function and symmetric key encryption. But the servers can deceive the owners by using previous metadata or responses due to the lack of randomness in the challenges. Also the limitation of such system is that numbers of updates and challenges are limited and fixed in advance and users cannot perform block insertions anywhere.

NCCloud It is implemented as a proxy that connects user applications and multiple clouds. NCCloud is built on top of network-coding-based storage schemes called regenerating codes. It is built on three layers. The file system layer presents NCCloud as a mounted drive, which can thus be easily interfaced with general user applications. The coding layer deals with the encoding and decoding functions. The storage layer deals with read/write requests with different clouds. NCCloud not only achieves fault tolerance of storage, but also allows cost-effective repair when a cloud permanently fails.

Public audit The public auditing scheme is proposed which provides a complete outsourcing solution of data; not only the data itself, but also its integrity checking [5]. Researchers utilize the homomorphic linear authenticator and random

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

masking to guarantee that the TPA would not learn any knowledge about the data content stored on the cloud server during the efficient auditing process, which not only eliminates the burden of cloud user from the tedious and possibly expensive auditing task, but also alleviates the users' fear of their outsourced data leakage.

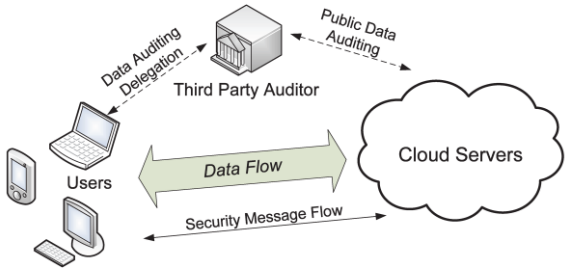


Fig.2. Architecture of cloud storage

Remote Data Checking scheme Remote data checking (RDC) schemes is presented for distributed storage systems based on network coding. When the client detects the failure of a server, it needs to take measures to ensure the data recovery condition is maintained. Recently, RDC schemes were proposed for replication-based and erasure coding-based distributed storage systems. To the best of our knowledge, RDC was not considered for network coding-based distributed storage systems.

Regenerating-code-based The regenerating codes are the code that achieves the min-cut in the information flow graph. There are two types of regenerating code based one minimum storage requirement space at storage nodes and the minimum bandwidth consumption during repair. Regenerating codes are based on the concept of network coding and tradeoff the repair traffic is reduced among storage nodes. The storage cost and repair traffic achieve optimal, and the optimal points are two, one optimal point to the minimum storage regenerating (MSR) codes, focus on minimize the repair bandwidth the condition that each node stores the minimum amount of data as in Reed-Solomon codes. Another optimal point is the minimum bandwidth regenerating (MBR) codes, which minimize the repair bandwidth further allowing each node to store more data.

III. RESEARCH ELABORATION

Cloud computing relies on sharing of resources to achieve coherence and economies of scale similar to a utility (like the electricity grid) over a network. At the foundation of cloud computing is the broader concept of converged infrastructure and shared services. The cloud also focuses on maximizing the effectiveness of the shared resources. Cloud resources are usually not only shared by multiple users but as dynamically re-allocated per demand. This can work for allocating resources to users. For example, a cloud computer facility, which serves European users during European business hours with a specific application (e.g. email) while the same resources are getting reallocated and serve North American users during North America's business hours with another application (e.g. web server). This approach should maximize the use of computing powers thus reducing environmental damage as well since less power, air conditioning, rack space, etc. is required for a variety of functions.

Modern cloud computing systems operate in a new and dynamic world, characterized by continual changes in the environment and in the system and performance requirements that must be satisfied. Continuous changes occur without warning and in an unpredictable manner, which are outside the control of the cloud provider. Therefore, advanced solutions need to be developed that manage the cloud system in a dynamically adaptive fashion, while continuously providing service and performance guarantees. In particular, recent studies have shown that the main challenges faced by cloud providers are to: (i) reduce costs, (ii) improve levels of performance, and (iii) enhance availability and dependability.

In existing system, the system proposed a formal definition of the PDP model for ensuring possession of files on untrusted storage. And also it introduced the concept of RSA-based homomorphic tags and suggested randomly sampling a few blocks of the file. In that work, the system proposed a dynamic version of the prior PDP scheme based

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

on MAC. This scheme allows very basic block operations with limited functionality but block insertions. And it used hash tree to improve the efficiency of dynamic PDP. However, the existing system cannot release the data owner from online burden. There are many mechanisms dealing with the integrity of outsourced data without a local copy have been proposed under different system and security models up to now. The most significant work among these studies are the PDP (provable data possession) model and POR (proof of retrievability) model, which were originally proposed for the single-server scenario. Considering that files are usually striped and redundantly stored across multi-servers or multi-clouds, explore integrity verification schemes suitable for such multi-servers or multi-clouds setting with different redundancy schemes, such as replication, erasure codes, and, more recently, regenerating codes.

To fully ensure the data integrity and save the users' computation resources as well as online burden, researchers proposed a public auditing scheme for the regenerating-code-based cloud storage, in which 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. Instead of directly adapting the existing public auditing scheme to the multi-server setting, we design a novel authenticator, which is more appropriate for regenerating codes.

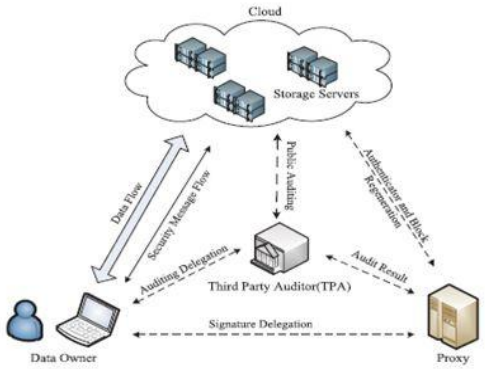


Fig.3. Regeneration-code-based cloud in public audit

IV. DISCUSSION

Cloud storage is a very popular technology for offering a flexible on-demand data outsourcing service. The technology have benefits like 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. It is observed that data owners lose ultimate control over the fate of their outsourced data; thus, the correctness, availability and integrity of the data are being put at risk. On one hand, the cloud service is usually faced with a broad range of internal/external adversaries, who would maliciously delete or corrupt users' data; on the other hand, the cloud service providers may act dishonestly, attempting to hide data corrupted and claiming that the files are still correctly stored in the cloud for reputation or monetary reasons. Thus it makes great sense for users to implement an efficient protocol to perform periodical verifications of their outsourced data to ensure that the cloud indeed maintains their data correctly.

V. CONCLUSION

In this paper, the various regenerating codes for cloud storage systems overview is given. The exact provable data possession is systematic system because it stores the data as it is such that original file can recovered with no decoding operations. The NCCLOUD is used with no arithmetic operation is performed on both the newcomer and the providers during repair operation. The Remote Data Checking scheme which regenerates the data across a multi-server and two-phase checking are performed on the new code chunks generated in the repair operation. These systems reduce the bandwidth consumption of the repair operations.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

REFERENCES

1. K. D. Bowers, A. Juels, and A. Oprea, "HAIL: A high-availability and integrity layer for cloud storage," in Proc. 16th ACM Conf. Comput. Commun. Secure. , 2009, pp. 187–198.
2. H. Shacham and B. Waters, "Compact proofs of retrievability," in Advances in Cryptology. Berlin, Germany: Springer-Verlag, 2008, pp. 90–107.
3. O. Rahamathunisa Begam, T. Manjula, T. Bharath Manohar, B. Susrutha, "Cooperative provable data possession for integrity verification in multicloud storage," IEEE Trans.Parallel Distrib. Syst., vol. 23, no. 12, pp. 2231–2244, Dec. 2012.
4. 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.
5. C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in Proc. IEEE INFOCOM, Mar. 2010, pp. 1–9.
6. B. Chen, R. Curtmola, G. Ateniese, and R. Burns, "Remote data checking for network coding-based distributed storage systems," in Proc.ACM Workshop Cloud Comput. Secur. Workshop, 2010, pp. 31–42.
7. Sophia S and Dr. Sharvani G S, "A Survey on Regenerating Codes for Distributed Cloud Storage", IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 9, pp. 1717–1726, Sep. 2013.
8. Henry C.H. Chen and Patrick P.C. Lee, "Enabling Data Integrity Protection in Regenerating-Coding-Based Cloud Storage: Theory and Implementation", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 2, FEBRUARY 2014

BIOGRAPHY

Samruddhi Devidas Pode is a Master of Engineering student in the Computer Engineering Department, RMD Sinhgad School of Engineering, Savitribai Phule Pune University. She received Bachelor of Engineering degree in 2014 from BDCOE, Wardha. Her research interests are Cloud computing, Parallel Computing, Network security.