Research Paper | Computer Science | E-ISSN : 2454-9916 | Volume : 1 | Issue : 5 | Dec 2015

SECURE DISTRIBUTED DEDUPLICATION SYSTEM TO IMPROVE DATA INTEGRITY IN CLOUD USING TPA

MANJUSHA BEHARA [1] | CHITRA BAZAZ [1] | RADHIKA GUPTA [1] | SRINIVAS D. [2]

[1] STUDENT, COMPUTER, G.S.M.C.O.E, PUNE, INDIA - 411045

[2] PRFESSOR, COMPUTER, G.S.M.C.O.E, PUNE, INDIA - 411045

ABSTRACT

Information deduplication is a system for taking out duplicates of information, for better utilization of storage room and data transfer capacity. There is only one copy of each document in cloud, possibility there may be N number of user for the same document. The data which is outsourced by user to cloud must b delicate information and it should be protected by leaking. In this paper we introduce TPA with secure distributed system for information integrity and tag consistency. The TPA is a public verifier that verifies that the data stored by the user is unchanged or corrupted in the cloud.

KEYWORDS: Deduplication, Distributed storage system, Integrity, TPA.

Introduction:

There are two types of deduplication:

i. File level deduplication, which reduce redundancies between diverse documents and eliminate the duplicate copy.

ii. Block level deduplication, the record is fragmented into blocks of fixed or variable size and then deduplication is performed to check the similar content in the files.

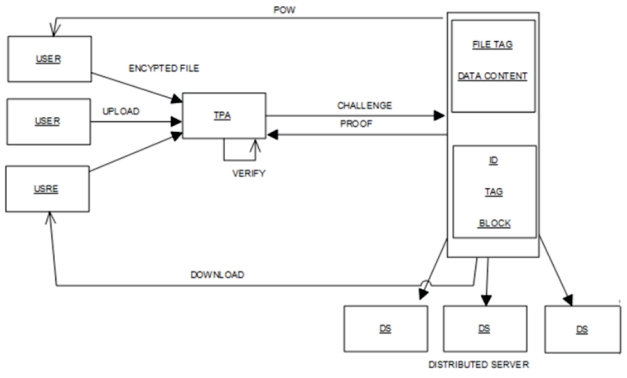
Ramp secret sharing method is use to divide the file into N no. of shares and distribute across the servers. Share and Recover algorithm is used to share and recover the data from distributed server. By using this method data can be recovered in case if the data is lost or corrupted in cloud without letting user know about it.

TPA is use to maintain the data integrity of outsourced data. It is a public verifier which acts as an intermediate between user and cloud.

It works in three steps:

i. Challenge
ii. Proof
iii. Verification

Materials and Methods:



Architecture

The system model involves three parties: the cloud server, a group of users, a public verifiers.

A public verifier such as TPA provides expert data auditing services to publicly verify the integrity of shared data stored in the cloud server.

When a public verifier wishes to check the integrity of shared data, it first sends, an auditing challenge, the cloud server responds to the public verifier with an auditing proof of the possession of shared data.

Essentially the process of public auditing is a challenge and response protocol between a public verifier and the cloud server.

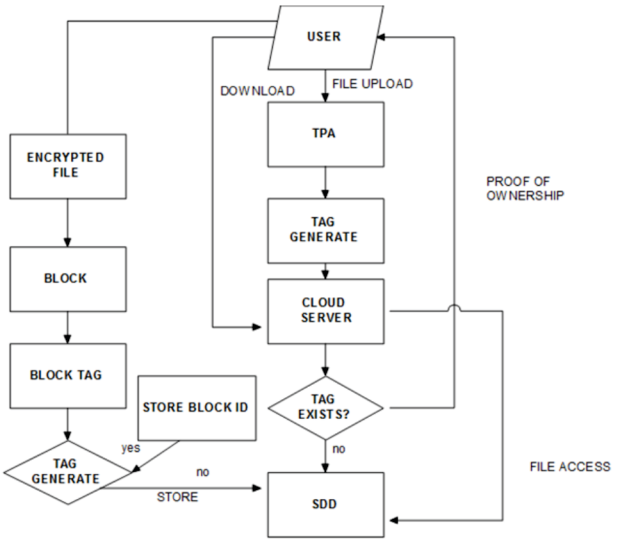
The TPA can be divided as

i. Proof of Data Possession(PDP)
ii. Proofs of Retrievability (PoR).

PDP scheme, are related protocols that only detect a large amount of corruption in outsourced data. [1, 2]

While PoR scheme [3], is a challenge-response protocol that enables a cloud provider to demonstrate to a client that a file is retrievable, i.e., recoverable without any loss or corruption. Their scheme use spot-checking and error correcting codes to ensure both "possession" and "retrievability" of remote data files.

The TPA sends a challenge to the cloud and in response the cloud sends a proof to TPA, After receiving the proof the last step is verification where it verifies that the data stored by the user is not corrupted and changed.



Flow chart:

Modules used:

i. File deduplication
ii. Block deduplication
iii. TPA
iv. Distributed storage server

Results:

The distributed deduplication systems is use to improve the reliability of data. Our model support file-level and block-level data deduplication. Deduplication systems uses the Ramp secret sharing scheme and demonstrated that it incurs small encoding/decoding overhead compared to the network transmission overhead in regular upload/download operations.

The public auditing scheme provide ease to the user's fear of their outsourced data leakage.

TPA can perform multiple auditing tasks in a batch manner for better efficiency.

Schemes used are secure and highly efficient.

The adversary cannot deduce any information of the file stored through the auditing interaction between CS and TPA.

Discussion:

The earlier systems performed the tasking of verifying the data by downloading the entire file from the cloud which was costly and time consuming. Also , public verifiers were themselves responsible for data leakage, therefore to overcome this problem ,TPA is used.TPA do not have any knowledge about the data contents stored on cloud server during the efficient auditing process .TPA can concurrently handle multiple audit session from different user for their outsourced data.

The data security and privacy has always been an issue so in future this can be enhanced. In order to make the system reliable concept of multiple servers has been introduced, however the same file is present at multiple location due to which unnecessary storage space is used, this problem can be overcome in future.

Acknowledgments: (optional)

We are profoundly grateful to Prof. Srinivas D, Project Coordinators for their expert guidance and continuous encouragement throughout to see that this project rightsits target since its commencement to its completion. We are also grateful to prof.SrinivasD. for his support and guidance that have helped us to expand our horizons of thought and expression.We would like to express our deepest appreciation towards Prof. F.B.Sayyad, Principal, G.S.Moze.Collge of Engineering, Pune and Prof. J. Ratnaraj, Head of theDepartment, Computer Engineering Department whose invaluable guidance supported us in completing this project. At last we must express our sincere heartfelt gratitude to all staff members of Computer Engineering Department who helped us directly or indirectly during this course of work.

REFERENCES AND FOOTNOTES:

[1] G. Ateniese, R. Burns, R. Curtmola, et al. Provable data possession at untrusted stores. Cryptology ePrint Archive, Report 2007/202, 2007. Online: http://eprint.iacr.org/. Version of 7 Dec. 2007; visited 10 Feb. 2008.

[2] G. Ateniese, R.D. Pietro, L.V. Mancini, et al. Scalable and efficient provable data possession [C].Proceedings of the 4th international conference on security and privacy in Communication networks. Istanbul, Turkey: ACM, 2008: 90-99.

[3] A. Juels, B. Kaliski. Pors: proofs of retrievability for large files[C]. Proceedings of CCS 2007.Alexandria, VA, USA, 2007. 584-597.

[4] A Hybrid Cloud Approach for Secure Authorized Deduplication.