Volume 4, Issue 2, February-2017, pp. 57-62 ISSN (O): 2349-7084

International Journal of Computer Engineering In Research Trends

Available online at: www.ijcert.org

Secure Data Deduplication over Distributed Cloud Server Framework with Effective User Revocation and Load Balancing.

[1]D.Jayanarayana Reddy, [2]M.Janardhan , [3]U.Veeresh

[1,2]Assistant Professor, Department of CSE, GPCET, Kurnool.
[3]Associate Professor, Department of CSE, GPCET, Kurnool.

Abstract: Nowadays Cloud Computing is an emerging Technology which leads various primitive services like SaaS, IaaS, and PaaS. Data deduplication mechanism is widely used to improve the bandwidth and storage space by removing duplicate copies of data from distributed cloud server.In Multi-owner manner data is stored and shared on distributed cloud server architecture, we have noticed some of the challenging issues, i.e., Users Privacy, Data Integrity, Load Balancing and Dynamic Ownership changes in attributes I,e User revocation Issues. To address the above challenges, we suggested a novel framework for Secure Data Deduplication over Distributed Cloud Server Framework with Effective User Revocation and Load Balancing Management. In our proposed framework, Block level hashing is pragmatic for every outsourced data and distributed into chunks and stored on distributed cloud servers, PoW protocol trappings Secured data deduplication and also provide an optimized solution for user revocation and load balancing issues , our projected approach is effective as the previous schemes while the added computational in the clouds is negligible.

Keywords: Cloud Computing, Secure DataDeduplication, Load balancing, distributed cloud server, PoW Protocol.

---

1. Introduction

In cloud storage services, Data deduplication mechanism is widely used to improve the bandwidth and storage space by removing duplicate copies of data from distributed cloud server.when multiple users outsource the same data to the cloud storage, but it raises issues relating to security and ownership. Proof of ownership (PoW) systems allow any possessor of the same data to prove to the cloud storage server that he owns the data in a robust way. However, many users are likely to encrypt their data before outsourcing them to the cloud storage to preserve privacy, but this hampers deduplication because of the randomization property of encryption. Recently, several deduplication schemes have been proposed to solve this problem by allowing each owner to share the same encryption key for the same data. However, most of the schemes suffer from security flaws, since they do not consider the dynamic changes in the ownership of outsourced data that frequently occur in an efficient cloud storage service.

In this paper, we propose a novel server-side deduplication scheme for encrypted data. It allows the cloud server to control access to outsourced data even when the ownership changes dynamically by exploiting randomised convergent encryption and secure ownership group key distribution it prevents data leakage not only to revoke users even though they previously owned that data but also to an honest-but-curious cloud storage server. Also, the proposed scheme guarantees data integrity

against any tag inconsistency attack. Thus, security is improved in the proposed scheme. The efficiency analysis results demonstrate that the proposed scheme is almost as efficient as the previous regimes, while the additional computational overhead is negligible.

2. The Contribution of the work

We propose a deduplication scheme over encrypted data. The proposed scheme ensures that only authorised access to the shared data is possible, which is considered to be the most significant challenge for efficient and secure cloud storage services [1] in the environment where ownership changes dynamically. It achieved by exploiting a group key management mechanism in each ownership group. As compared to the previous deduplication schemes over encrypted data, the proposed scheme has the following advantages regarding security and efficiency. First, dynamic ownership management guarantees the backwards and forward secrecy of deduplicated data upon any ownership change.

As divergent to the previous schemes, the data encryption key is efficient and selectively distributed to effective owners upon any ownership change of the data through a stateless group key distribution mechanism using a binary tree. The ownership and key management for each user can conduct by the semi-trusted cloud server deployed in the system. Thus, the proposed scheme delegates the most laborious tasks of ownership, management to the cloud server without leaking any confidential information to it, rather than to the others. Second, the proposed scheme ensures security in the setting of PoW by introducing a re-encryption mechanism that uses an additional group key for the dynamic ownership group. Thus, although the encryption key (that is the hash value of the file) is revealed in the setting of PoW, the privacy of the outsourced data is still preserved against outside adversaries, while deduplication over encrypted data is still enabled and data integrity against poison attacks is guaranteed.

3. Literature survey

Several deduplication schemes have been proposed by the research community [2–4] showing how deduplication allows very appealing reductions in the usage of storage resources [5, 6]. Most works do not consider security as a concern for deduplicating systems; recently, however, Harnik et al. [7] Have presented some attacks that can lead to data leakage in storage systems in which client-side deduplication is in place. To thwart such attacks, the concept of proof of ownership has been introduced [8, 9]. None of these works, however, can provide real end-user config- mentality in the presence of a malicious or honest-but-curious cloud provider.

Convergent encryption is a cryptographic primitive introduced by Douceur et al. [10, 11], attempting to combine data confidentiality with the possibility of data deduplication. The Convergent encryption of a message consists of encrypting the plaintext using a deterministic (symmetric) encryption scheme with a key which is deterministically derived solely from the plaintext. Clearly, when two users independently attempt to encrypt the same file, they will generate the same ciphertext which can be easily deduplicated. Unfortunately, convergent encryption does not provide semantic security as it is vulnerable to content-guessing attacks. Later, Bellare et al. [12] formalised convergent encryption under the name message-locked encryption. As expected, the security analysis presented in [12] highlights that message-locked encryption offers confidentiality for unpredictable messages only, clearly failing to achieve semantic security.

Xu et al. [13] present a PoW scheme allowing client-side deduplication in a bounded leakage setting. They provide a security proof in a random oracle model for their solution , but do not address the problem of flat min-entropy files.

Recently, Bellare et al. presented DupLESS [14], a server-aided encryption for deduplicated storage. Similarly to ours, their solution uses a modified convergent encryption scheme with the aid of a secure component for the major generation. While DupLESS offers the possibility to use server-side deduplication securely, our scheme targets secure client-side deduplication.

4. Presented system

As per presented system data deduplication practices can be categorised into two different approaches: deduplication over open data and deduplication over

encrypted data. In the previous approach, most of the existing schemes have been proposed in order to perform a PoW procedure in an efficient and robust manner, since the hash of the file, which is treated as a "proof" for the entire file, is susceptible to being leaked to outside adversaries because of its comparatively small size. Whereas, in the second approach, Users data privacy is the primary security requirement to protect against not only outside adversaries but also inside the cloud server. Thus, most of the schemes have been proposed to provide data encryption, while still benefiting from a deduplication technique, by enabling data owners to share the encryption keys in the presence of the inside and outside adversaries. Since encrypted data are given to a user, data access control can be additionally implemented by selective key distribution after the PoW process. However, not much work has yet been done to address dynamic ownership management and its related security problem. I.e., Data Integrity, Load Balancing and Dynamic Ownership changes in attributes I,e User revocation Issues.still to be addressed.

Table 1. Comparing Various Deduplication Approaches

Approach	**Cost**	**Throughput**	**Used Bandwidth**	**Deduplication Ratio**	**Required Storage**
File Level Deduplication	Low	High	Low	Low	Medium
Block level Deduplication	High	Low	Low	High	Less
Source based Deduplication	Relatively Low	Medium	Low	Medium	Medium
Target based Deduplication	High	Medium	High	Medium	Medium

5 .Proposed System

We proposed a novel framework for Secure Data Deduplication over Distributed Cloud Server Framework with Effective Load Balancing Management. In our approach, Block level hashing is applied for every outsourced data and divided into chunks and stored on Distributed cloud servers, PoW protocol trappings Secured data deduplication and also provide an optimised solution for load balancing issues. In Cloud Storage services, the data deduplication mechanism is widely used to improve the storage space and reduce the bandwidth by removing duplicate copies of data from distributed cloud server. Our proposed approach is well-organized as the previous schemes while the additional computational overhead is negligible.

5.1 System model:

In the shown figure 1. It Describes General Scenario of data outsourcing over Cloud Computing via auditor .

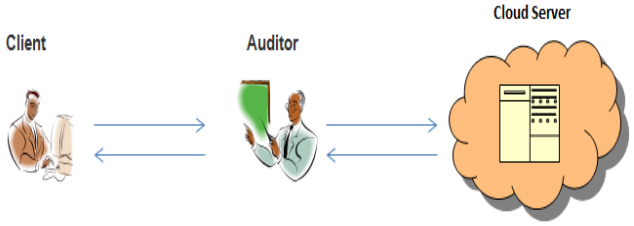
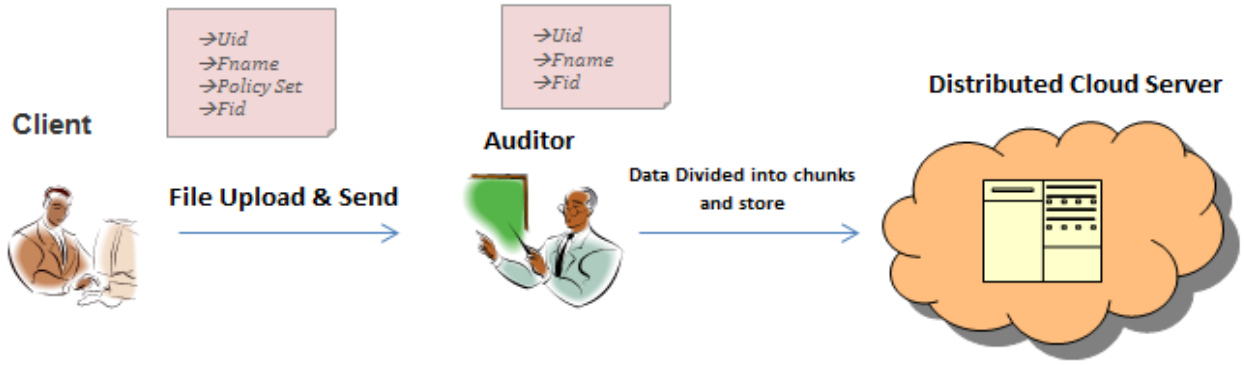


Fig 1.General Scenario of data outsourcing over Cloud Computing via auditor

The client always outsources his/her data to a cloud server for storing and sharing through auditor for better data integrity over the cloud framework.Data outsourced in an encrypted manner for the sake of privacy.

In fig 2. Describes the Secure data outsourcing Scenario over the distributed Cloud framework The client always sends data to the cloud server for storing and sharing through auditor, behalf of works cloud server for providing data integrity.



File 2. Secure data outsourcing Scenario over the distributed Cloud framework.

Client Module: In this module, the client can upload an encrypted file and send to the auditor, uploaded file contains the User ID(Uid), File name(name), Access Policy and File Id.

Auditor: In this module, auditor verifies Uid, Fame, and Fid for uploaded file and distribute into chunks and store in cloud servers.

Distributed cloud server: Distributed encrypted data has been stored in different servers in cloud framework.

Data duplication checking on Auditor level

When the user wants to upload same data to the cloud server via auditor, Duplicate Occurs at File Level file will not be uploaded as shown in fig 3., the immediate PoW protocol will be activated for data deduplication using a convergent key encryption mechanism which allows only authorised user to data deduplication in a secured manner as shown in fig 4.

Auditor maintains a log table for every incoming file to upload in cloud server. Log table maintains and file details and user details , if same user wants to send same file, then it verified with log table ,if matches data will not send to the cloud server.

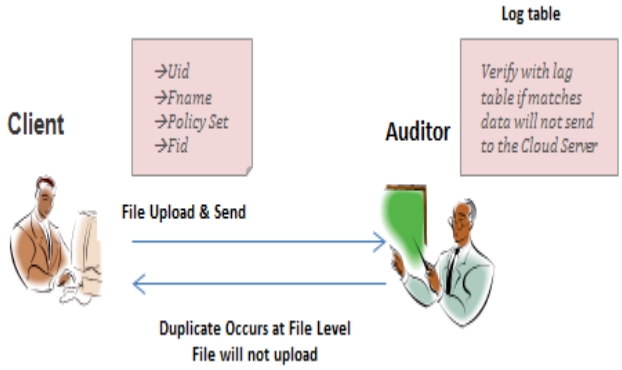


Fig 3. Data duplication checking on Auditor level

In fig 4 .When same data is passed on through auditor to store in cloud server ,now the data will identified as duplicate due to Block level checking .in this case distributed cloud server requests for PoW protocol for Proof of Ownership verification for secure data deduplication, when a client verifies his ownership then secure data deduplication will be performed

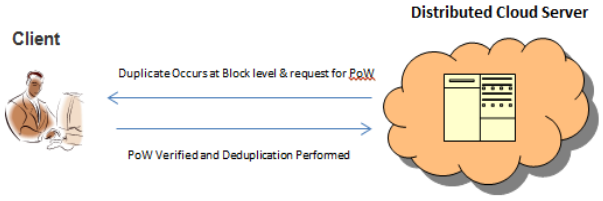


Fig 4. Secure data deduplication

Fig 4. Shows the Multi-owner manner data is stored and shared on distributed cloud server architecture

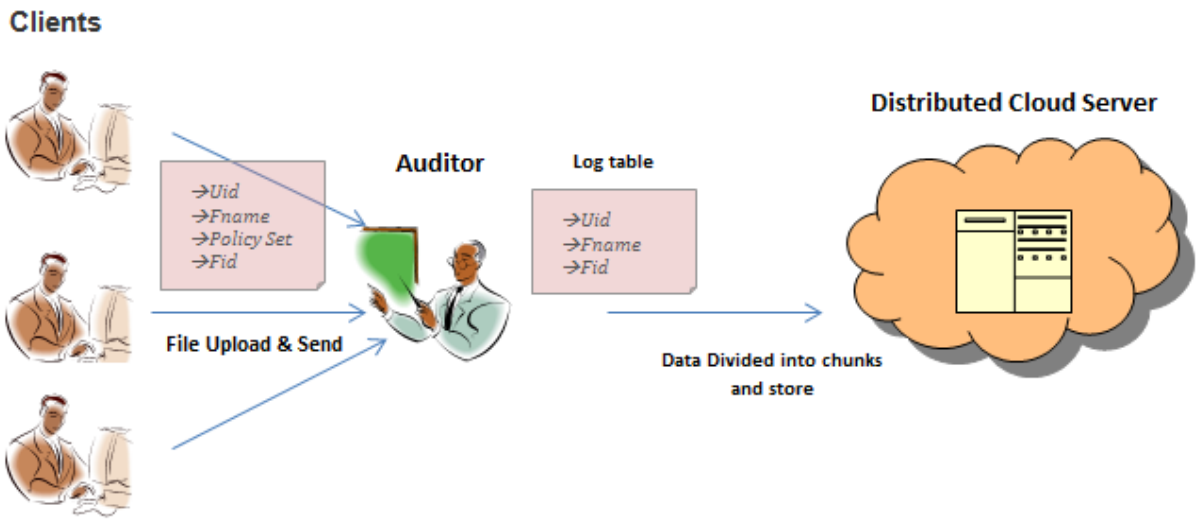


Fig 5.Multi-owner manner data is stored and shared on distributed cloud server architecture

Deduplication is utmost in effect when multiple users outsource the same data to the cloud storage, but it raises issues relating to safety and ownership. Evidence of the property schemes allows every owner of the similar data to verify to the cloud storage server that he possesses the data in a dynamic way. However, many users are likely to encrypt their data before outsourcing them to the cloud storage to preserve privacy, we recommend an innovative server-side deduplication scheme for encrypted data. It permits the cloud server to control access to outsourced data even when the ownership changes dynamically by exploiting randomized convergent encryption and secure ownership group key distribution it prevents data leakage not only to revoke users even though they previously owned that data, but also to an honest-but-curious cloud storage server. In adding, the recommended scheme guarantees data integrity against any tag inconsistency attack. Thus, security is improved in the proposed scheme. The efficiency analysis results demonstrate that the proposed scheme is almost as efficient as the previous systems, while the additional computational overhead is negligible.

5.2 Load Balancing

Load balancing is a method to distribute the workload consistently across two or more computers, network links, CPUs, hard drives, or other properties, to get optimal resource utilization, maximize throughput, minimize response time, and avoid overload of any one of the properties. Using multiple components with load balancing in its place of a single component may increase reliability through redundancy.

The users submit their diverse applications to the Cloud Service Provider through a communication channel. The requests from the users are queued up under the Cloud Service Provider?S Data Center. The sub-servers are then checked up for the minimum load with the CPU performance level of the currently executing task. The Cloud Service Provider then allocates the requested job to the sub-servers that have a minimum load to process the job in a First In First Out (FIFO) manner. Thus the User requested task will be assigned to the available sub server which contains minimum load and it is concerned to process the User requested a job.

6. Conclusion:

In this paper, we examined various challenging issues related to Data Deduplication Over Distributed Cloud Server Framework. I.e., Users Privacy, Data Integrity, Load Balancing and Dynamic Ownership changes in attributes I,e User revocation Issues to address the above challenges, we recommended a novel framework for Secure Data Deduplication Over Distributed Cloud Server Framework With Effective User Revocation and Load Balancing Management which offer secure data storage, to maintain integrity of the data, to increase the user level of authentication and to improve the performance efficiently by 70-80% of balancing the load.

References:

[1] M. Mulazzani, S. Schrittwieser, M. Leithner, and M. Huber, "Dark clouds on the horizon: using cloud storage as an attack vector and online slack space," Proc. USENIX Conference on Security, 2011.

[2] Meister, D., Brinkmann, A.: Multi-level comparison of data deduplication in a backup scenario. In: SYSTOR '09, New York, NY, USA, ACM (2009) 8:1–8:12

[3] Mandagere, N., Zhou, P., Smith, M.A., Uttamchandani, S.: Demystifying data deduplication. In: Middleware '08, New York, NY, USA, ACM (2008) 12–17

[4] Aronovich, L., Asher, R., Bachmat, E., Bitner, H., Hirsch, M., Klein, S.T.: The design of a similarity based deduplication system. In: SYSTOR '09. (2009) 6:1–6:14

[5] Dutch, M., Freeman, L.: Understanding data de-duplication ratios. SNIA forum (2008) http://www.snia.org/sites/default/files/Understanding_Data_ Deduplication_Ratios-20080718.pdf.

[6] Harnik, D., Margalit, O., Naor, D., Sotnikov, D., Vernik, G.: Estimation of deduplication ratios in large data sets. In: IEEE MSST '12. (april 2012) 1 –11

[7] Harnik, D., Pinkas, B., Shulman-Peleg, A.: Side channels in cloud services: Deduplication in cloud storage. Security Privacy, IEEE 8(6) (nov.-dec. 2010) 40 –47

[8] Halevi, S., Harnik, D., Pinkas, B., Shulman-Peleg, A.: Proofs of ownership in remote storage systems. In: CCS '11, New York, NY, USA, ACM (2011) 491–500

[9] Di Pietro, R., Sorniotti, A.: Boosting efficiency and security in proof of ownership for deduplication. In: ASIACCS '12, New York, NY, USA, ACM (2012) 81–82

[10] Douceur, J.R., Adya, A., Bolosky, W.J., Simon, D., Theimer, M.: Reclaiming space from duplicate files in a serverless distributed file system. In: ICDCS '02, Washington, DC, USA, IEEE Computer Society (2002) 617–632

[11] Storer, M.W., Greenan, K., Long, D.D., Miller, E.L.: Secure data deduplication. In: StorageSS '08, New York, NY, USA, ACM (2008) 1–10

[12] Bellare, M., Keelveedhi, S., Ristenpart, T.: Message-locked encryption and secure deduplication. In: Advances in Cryptology–EUROCRYPT 2013. Springer 296–312

[13] Xu, J., Chang, E.C., Zhou, J.: Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. In: 8th ACM SIGSAC symposium. 195–206 14. Bellare, M., Keelveedhi, S., Ristenpart, T.: DupLESS: server-aided encryption for deduplicated storage. In: 22nd USENIX conference on Security. (2013) 179–194

