International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 04 | April - 2022 Impact Factor: 7.185 ISSN: 2582-3930

HYBRID CLOUD APPROACH WITH SECURITY AND DATA DEDUPLICATION

R Amrutha[1] and Dr. Murugan R[2]

[1]Research Scholar, School of Computer Science and Information Technology, JAIN (Deemed to be University), Bangalore, India

[2]Associate Professor, School of Computer Science and Information Technology, JAIN (Deemed to be University), Bangalore, India

ABSTRACT

Data deduplication is one of important data compression techniques for eliminating duplicate copies of repeating data and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this work makes the first attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. To protect the confidentiality of sensitive data the convergent encryption technique has been used that is, encrypting the data before outsourcing it. In order to support secure and authorized deduplication, differential privileges of the users are considered. For better security this system uses a hybrid cloud architecture, which also supports the authorized duplicate check.

1. INTRODUCTION

Cloud computing provides seemingly unlimited "virtualized" resources to users as services across the whole Internet, while hiding platform and implementation details. Today's cloud service providers offer both highly available storage and massively parallel computing resources at relatively low costs. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and shared by users with specified privileges, which define the access rights of the stored data. One critical challenge of cloud storage services is the management of the ever-increasing volume of data. To make data management scalable in cloud computing, deduplication has been a well-known technique and has attracted more and more attention recently.

Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication can take place at either the file level or the block level. For file level deduplication, it eliminates duplicate copies of the

same file. Deduplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files.

2. LITERATURE SURVEY

Secure deduplication is a technique for eliminating duplicate copies of storage data and provides security to them. To reduce storage space and upload bandwidth in cloud storage deduplication has been a well-known technique. For that the purpose convergent encryption has been extensively adopt for secure deduplication, critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. The basic idea in this project is that we can eliminate duplicate copies of storage data and limit the damage of stolen data if we decrease the value of that stolen information to the attacker. This project makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure deduplication. We first introduce a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them. However, such a baseline key management scheme generates an enormous number of keys with the increasing number of users and requires users to dedicatedly protect the master keys. To this end, we propose Dekey, User Behaviour Profiling and Decoys technology. Dekey new construction in which users do not need to manage any keys on their own but instead securely distribute the convergent key shares across multiple servers for insider attacker. As a proof of concept, we implement Dekey using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments. User profiling and decoys, then, serve two purposes: First one is validating whether data access is authorized when abnormal information access is detected, and second one is that confusing the attacker with bogus information. We posit that the combination of these security features will provide unprecedented levels of security for the deduplication in insider and outsider attacker. [1]

Ever increasing volume of back up data in cloud storage may be a vital challenge back up windows are shrinking due to growth of information. We use the concept of de-duplicate. Deduplication means duplicate data is eliminated a pointer is created to reference a data that is backed up. Deduplication can take place at file level, in this it detects redundant data within and across files or at the block level, in this it removes redundant copies of identical files. [2]

Information deduplication is one of critical information packing strategies for wiping out copy duplicates of rehashing information and has been broadly utilized as a part of Cloud storage to diminish the measure of storage room and spare data transfer capacity. To secure the secrecy of delicate information while supporting deduplication, the merged encryption system has been proposed to encode the information before outsourcing. To better secure information security, this project makes the first endeavor to formally address the issue of approved information deduplication. Not the same as customary deduplication frameworks, the differential benefits of clients are further considered in copy check other than the information itself. We additionally show a few new deduplication developments supporting approved copy weigh in a half and half cloud building design. Security dissection exhibits that our plan is secure regarding the definitions defined in the proposed security model. [3]

3. EXISTING SYSTEM

- To prevent unauthorized access, a secure proof of ownership protocol is also needed to provide the proof that the user indeed owns the same file when a duplicate is found. After the proof, subsequent users with the same file will be provided a pointer from the server without needing to upload the same file. A user can download the encrypted file with the pointer from the server, which can only be decrypted by the corresponding data owners with their convergent keys. Thus, convergent encryption allows the cloud to perform deduplication on the ciphertexts and the proof of ownership prevents the unauthorized user to access the file.

In an authorized deduplication system, each user is issued a set of privileges during system initialization. Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files. Before submitting his duplicate check request for some file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud.

3.1 Disadvantages of the existing system

- Existing deduplication systems cannot support differential authorization duplicate check, which is important in many applications.
- Traditional deduplication systems based on convergent encryption, although providing confidentiality to some extent, do not support the duplicate check with differential privileges.
- No differential privileges have been considered in the deduplication based on convergent encryption technique.

4. PROPOSED SYSTEM

Aiming at efficiently solving the problem of deduplication with differential privileges in cloud computing, a hybrid cloud architecture is considered consisting of a public cloud and a private cloud. Private cloud is involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges. Such an architecture is practical and has attracted much attention from researchers. The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud. A new deduplication system supporting differential duplicate check is proposed under this hybrid cloud architecture where the S-CSP resides in the public cloud. The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.

4.1 Advantage of Proposed System

- Enhanced system security
- An advanced scheme to support stronger security by encrypting the file with differential privilege keys.
- The users without corresponding privileges cannot perform the duplicate check.
- Unauthorized users cannot decrypt the cipher text even collude with the S-CSP.

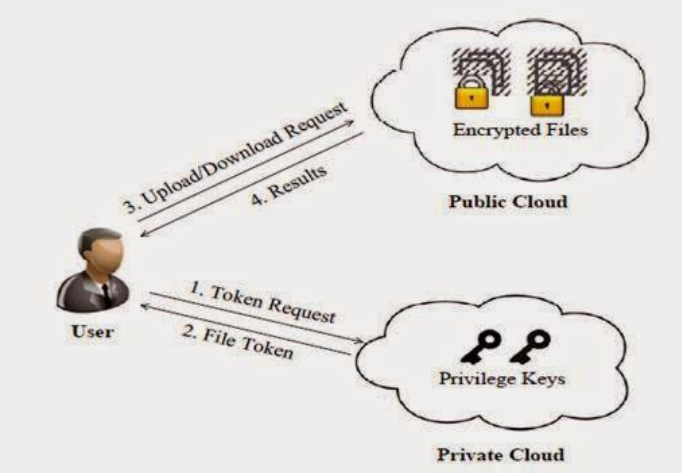

Fig 4: System Architecture

Cloud computing has become an effective solution for various services on the internet. Additionally, a cloud environment acts as a default storage location for application users. Large cloud storage service providers receive terabytes of data per second with an enormous amount of duplicated content. The duplicate copies can be eliminated using the deduplication technique. The proposed research work detects redundant audio content of the existing files in a cloud environment. Additionally, this study investigates the cloud computing environment which consists of numerous audio files (waveform audio file format). The proposed work detects redundant content and identifies only a part of the existing audio file, which refines the duplicated content over the space. This can be accomplished using the refined super subset identification algorithm, which processes a waveform audio file format content as numerical data and efficiently detects the repeated contents in an elastic cloud computing environment. The results demonstrate the accuracy of detecting duplicated files present in various files. The visual representation of the results proves the accuracy and exhibits that the quality of the audio content was not compromised. Finally, the method is effectively validated in a real-time environment.

5. ACKNOWLEDGEMENT

I should convey my real tendency and obligation to Dr M N Nachappa and Dr. Murugan R undertaking facilitators for their effective steerage and consistent inspirations all through my assessment work. Their ideal bearing, absolute co-action and second discernment have made my work gainful.

7. REFERENCES

[1]. “Secure Deduplication And Data Security With Efficient And Reliable Cekm” N.O.Agrawal, Prof Mr. S.S.Kulkarni, Published In 2014.

[2]. “Survey on Authorized Data Deduplication System using Cryptographic and Access Control Techniques” Santoshi S Patil, Samprati T, Asst. Prof. Swetha K S,published in 2014.

[3]. “A Hybrid Cloud Move Toward For Certified Deduplication” E.Mounika, P.Manvitha, U.Shalini, Mrs. K. Lakshmi, published in 2014.

[4]. S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In *Workshop on Cryptography and Security in Clouds (WCSC 2011)*, 2011.

[5]. S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, *ACM Conference on Computer and Communications Security*, pages 491–500. ACM, 2011.

[6]. W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S. Ossowski and P. Lecca, editors, *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pages 441–446. ACM, 2012