A Secured Authorized Deduplication of Data in a Hybrid Cloud

Harshit Sharma[1], Sanket Khamkar[2], Kunal Kinholkar[3], Akshay Tambat[4]

[1,2,3,4] *BE Scholar, Department of Computer Engineering, D Y Patil College of Engineering, Akurdi, Pune, MH, India*

ABSTRACT

In personal computing devices that rely on a cloud storage environment for data backup, an imminent challenge facing source deduplication for cloud backup services is the low deduplication efficiency due to a combination of the resource intensive nature and the limited system resources. 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 paper makes an 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. We also present several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct tested experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

Keywords: - *Data deduplication, Convergent encryption, Confidentiality, Hybrid cloud, Authorized Duplicate check.*

1. Introduction

Cloud computing provides apparently unlimited virtualized resources to users as services across the entire web, whereas activity platform and implementation details. Today's cloud service suppliers provide each extremely offered storage and massively parallel computing resources at comparatively low prices. As cloud computing becomes prevailing, associate degree increasing quantity of knowledge is being hold on within the cloud and shared by users with specified privileges, that outline the access rights of the hold on information. One important challenge of cloud storage services is that the management of the ever-increasing volume of knowledge to form information management scalable in cloud computing, deduplication has been a well-known technique and has attracted a lot of attention recently. Information deduplication may be a specialized information compression technique for eliminating duplicate copies of continuance information in storage. The technique is employed to boost storage utilization and may be applied to network information transfers to scale back the amount of bytes that has to be sent. Rather than keeping multiple information copies with a similar content, deduplication eliminates redundant information by keeping just one physical copy and referring different redundant information thereto copy. Deduplication will occur at either the file level or the block level. For file level deduplication, it eliminates duplicate copies of a similar file. Deduplication also can occur at the block level that eliminates duplicate blocks of knowledge that occur in non-identical files.

Cloud computing is associate degree rising service model that has computation and storage resources on the web. One enticing practicality that cloud computing offers is cloud storage. People and enterprises area unit usually needed to remotely archive their information to avoid any data loss just in case there are a unit any hardware/software failures or unforeseen disasters. Rather than getting, the required storage media to stay information backups, people and enterprises will merely source their information backup services to the cloud service suppliers, which offer the mandatory storage resources to host the info backups whereas cloud storage is

enticing, a way to offer security guarantees for outsourced information becomes a rising concern. One major security challenge is to supply the property of assured deletion, i.e., information files area unit for good inaccessible upon requests of deletion. Keeping information backups for good is undesirable, as sensitive data could also be exposed within the future attributable to information breach or inaccurate management of cloud operators.

1.1 Background

Current data de-duplication frameworks, the personal cloud area unit enclosed as a mediator to allow data proprietor/clients to perform soundly copy talk over with differential edges. Such style is practical and has force in a lot of thought from specialists. The data proprietors simply source their information reposting by exploitation open cloud whereas the data operation overseen privately cloud. We have a tendency to show a propelled commit to bolster additional grounded security by scrambling the document with differential profit keys. On these lines, the purchasers while not comparison edges cannot perform the copy check. Moreover, such unapproved purchasers cannot decipher the figure message even connive with the S-CSP.

1.2 Cloud Computing

Cloud computing refers to applications and services that run on a distributed network victimization virtualized resources and accessed by common net protocols and networking standards. It is distinguished by the notion that resources area unit virtual and limitless, which details of the physical systems on that software package runs area unit abstracted from the user. Cloud computing takes the technology, services, and applications that area unit kind of like those on the net and turns them into a self-service utility.

1.2 Deployment Models

A preparation model defines the aim of the cloud and the nature of however the cloud is found. The National Institute of Standards and Technology definition for the four preparation models is as follows:
Public cloud: The general public cloud infrastructure is offered for public use or else for an outsized business cluster and is owned by a company commercialism cloud services.
Private cloud: The personal cloud infrastructure is operated for the exclusive use of a company. The cloud could also be managed by that organization or a third party. Personal clouds could also be either on- or off-premises.
Hybrid cloud: A hybrid cloud combines multiple clouds (private, community of public) wherever those clouds retain their distinctive identities, however area unit sure along as a unit. A hybrid cloud might provide standardized or proprietary access to knowledge and applications, still as application immovableness.
Community cloud: A community cloud is one wherever the cloud has been organized to serve a standard operate or purpose.
It may be for one organization or for many organizations; however they share common issues like their mission, policies, security, regulative compliance desires, and so on. A community cloud could also be managed by the constituent organization(s) or by a third party.

1.2 Architecture of Cloud Computing

The goal of cloud computing is to use ancient supercomputing, or high- performance computing power, usually employed by military and analysis facilities, to perform tens of trillions of computations per second, in consumer-oriented applications like money portfolios, to deliver customized data, to supply information storage or to power giant, immersive on-line laptop games.

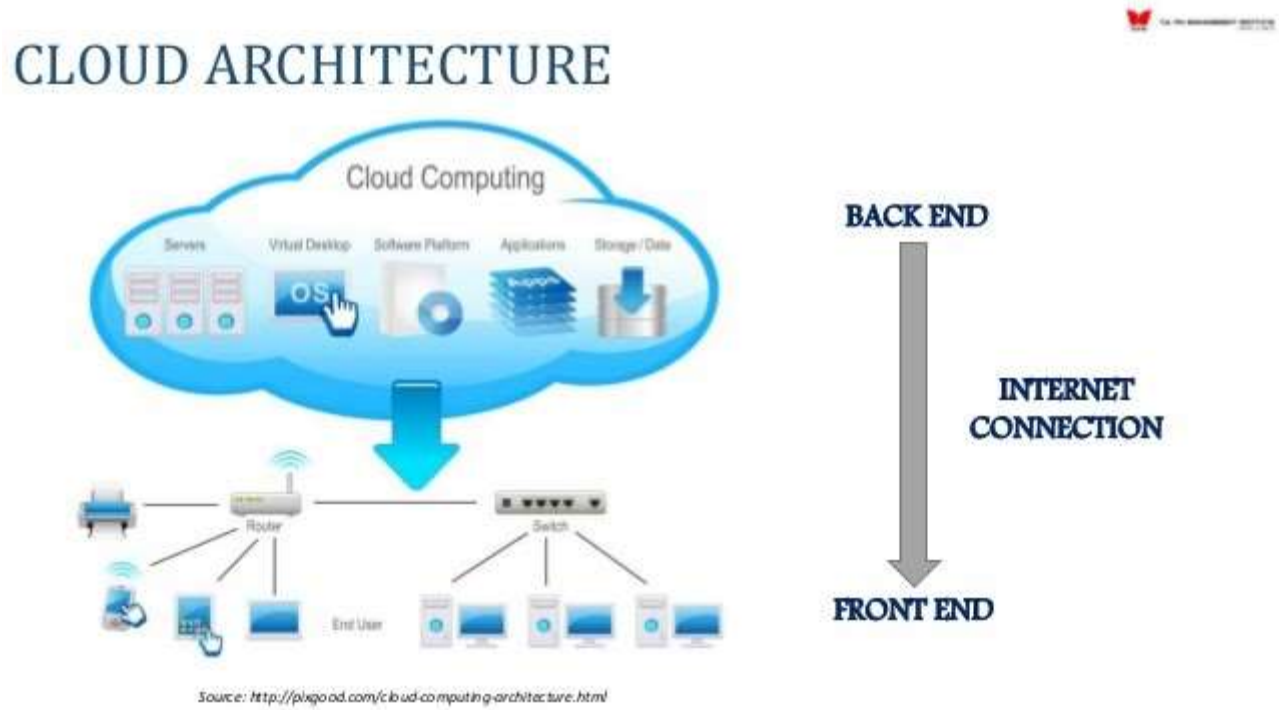

Chart -1: Cloud Architecture

To do this, cloud computing uses networks of large groups of servers typically running low-cost consumer PC technology with specialized connections to spread data-processing chores across them. This shared IT infrastructure contains large pools of systems that linked together. Often, virtualization techniques are used to maximize the power of cloud computing.

2. Comparison

Sr. No.	Title	Conference	Publication Year	Topic Reviewed
1.	Fast and secure laptop backups with encrypted de-duplication.	LISA'10 Proceedings of the 24th international conference on Large installation system administration	2010	This paper describes an algorithm, which takes advantage of the data, which is common between users to increase the speed of backups, and reduce the storage requirements. This algorithm supports client-end per-user encryption, which is necessary for confidential personal data.

2.	Cloud-Assisted Mobile-Access of Health Data With Privacy and Auditability	IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS	MARCH 2014	This paper build privacy into mobile healthcare systems with the help of the private cloud. This system offers salient features including efficient key management, privacy-preserving data storage, and retrieval, especially for retrieval at emergencies, and auditability for misusing health data. Author propose to integrate key management from pseudo random number generator for unlink ability, a secure indexing method for privacy preserving keyword search which hides both search and access patterns based on redundancy, and integrate the concept of attribute based encryption with threshold signing for providing role-based access control with auditability to prevent potential misbehavior, in both normal and emergency cases.
3.	Private Data Deduplication Protocols in Cloud Storage	Proceedings of the 27th Annual ACM Symposium on Applied Computing	2012	Private data deduplication protocol, a deduplication technique for private data storage is introduced and formalized. A private data deduplication protocol allows a client who holds a private data proves to a server who holds a summary string of the data that he/she is the owner of that data without revealing further information to the serve.
4.	Security proofs for identity-based identification and signature schemes	International Conference on the Theory and Applications of Cryptographic Techniques	MAY 2004	These is a framework that on the one hand helps explain how these schemes are derived and on the other hand enables modular security analyses, thereby helping to understand, simplify, and unify previous work.
5.	A reverse deduplication storage system optimized for reads to latest backups IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED S	Proceedings of the 4th Asia-Pacific Workshop on Systems	2013	It introduces fragmentation that degrades read performance. RevDedup, a deduplication system that optimizes reads to the latest backups of virtual machine (VM) images using reverse deduplication. In contrast with conventional deduplication that removes duplicates from new data, RevDedup removes duplicates from old data, thereby shifting fragmentation to old data while keeping the layout of new data as sequential as possible.

6.	Secure deduplication with efficient and reliable convergent key management	IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS	JUNE 2014	It introduce a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them to the cloud. 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. This paper propose Dekey, a 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. Dekey incurs limited overhead in realistic environments.
7.	A Review on Cloud Storage Performance to Improve File Accessing Efficiency	International Journal of Engineering Research & Technology	NOV 2014	Files are re-uploaded into server that decreases the bandwidth and increases the server workload. So, to remove the redundant or duplicate copies De-Duplication technique is used in cloud storage. To optimize the transmission node performance the Index Name Server (INS) architecture is used. Beside from this file compression, chunk matching, real time feedback control and load balancing techniques are also discussed. Using these techniques the cloud storage performance increases also the storage workload is reduced.
8.	Decentralized Access Control with Anonymous Authentication of Data Stored in Clouds. International Journal of Inventions in Computer Science and Engineering	International Journal of Inventions in Computer Science and Engineering	2014	The cloud verifies the authenticity of the server without knowing the user's identity before storing data. Author added feature of access control in which only valid users are able to decrypt the stored information. The scheme prevents replay attacks and supports creation, modification, and reading data stored in the cloud and address user revocation. Authentication and access control scheme is decentralized and robust, unlike other access control schemes designed for clouds, which are centralized. The communication, computation, and storage overheads are comparable to centralized approaches.

9.	Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data	IEEE INFOCOM	2011	Sensitive data has to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely differentiate the search results. The efficient principle of "coordinate matching", i.e., as many matches as possible, to capture the similarity between search query and data documents, and further use "inner product similarity" to quantitatively formalize such principle for similarity measurement.
10.	A Secure Client Side Deduplication Scheme in Cloud Storage Environments	IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS	2014	Security and privacy are among top concerns for the public cloud environments. Towards these security challenges, we propose and implement, on Open Stack Swift, a new client-side deduplication scheme for securely storing and sharing outsourced data via the public cloud. First, it ensures better confidentiality towards unauthorized users. Second, by integrating access rights in metadata file, an authorized user can decipher an encrypted file only with his private key.

Table -1: Literature Survey

4. CONCLUSION

In this paper, the notion of licensed knowledge deduplication projected to shield the info security by together with differential privileges of users within the duplicate check. We also presented many new deduplication constructions supporting authorized duplicate sign up hybrid cloud design, in which the duplicate-check tokens of files area unit generated by the personal cloud server with personal keys. Security analysis demonstrates that our schemes area unit secure in terms of insider and outsider attacks per the projected security model. As a signal of construct, we tend to enforce an image of our projected licensed duplicate check theme and conduct test bed experiments on our image. We showed that our licensed duplicate check theme incurs minimal overhead compared to focus coding and network transfer.

5. REFERENCES

[1]. Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee,Wenjing Lou A Hybrid Cloud Approach for Secure Authorized Deduplication IEEE Transactions on Parallel and Distributed Systems: PP Year 2016

[2]. Mr Vinod B Jadhav Prof Vinod S Wadne Secured Authorized Deduplication Based Hybrid Cloud Approach International Journal of Advanced Research in Computer Science and Software Engineering

[3]. Abdul Samadhu, J. Rambabu, R. Pradeep Kumar, R. Santhya Detailed Investigation on a Hybrid Cloud Approach for Secure Authorized Deduplication International Journal for Research in Applied Science and Engineering Technology (IJRASET)

[4]. Jadapalli Nandini, Rami reddy Navateja Reddy Implementation Deduplication System with Authorized Users International Research Journal of Engineering and Technology (IRJET)

[5]. Sharma Bharat, Mandre B.R. A Secured and Authorized Data Deduplication with Public Auditing International Journal of Computer Applications (09758887)

[6]. Wee Keong Ng SCE, NTU Yonggang Wen SCE, NTU Huafei Zhu Private Data Deduplication Protocols in Cloud Storage SAC12 March 2529, 2012, Riva del Garda, Italy. Copyright 2011 ACM 9781450308571/12/03

[7]. Shweta D. Pochhi, Prof. Pradnya V. Kasture Encrypted Data Storage with Deduplication Approach on Twin Cloud International Journal of Innovative Research in Computer and Communication Engineering

[8]. Backialakshmi. N Manikandan. M SECURED AUTHORIZED DEDUPLICATION IN DISTRIBUTED SYSTEM IJIRST International Journal for Innovative Research in Science and Technology Volume 1 Issue 9 February 2015

[9]. Bhushan Choudhary, Amit Dravid A Study On Secure Deduplication Techniques In Cloud Computing International Journal of Advanced Research in Computer Engineering and Technology (IJARCET) Volume 3, Issue 12, April 2014

[10]. James S. Plank Lihao Xu Optimizing Cauchy Reed-Solomon Codes for Fault-Tolerant Network Storage Applications The 5th IEEE International Symposium on Network Computing and Applications (IEEE NCA06), Cambridge, MA, July, 2006.