iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** III **Month of publication:** March 2017

DOI: http://doi.org/10.22214/ijraset.2017.3052

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Survey Paper- Distributed Duplication of Data and Reliability

Ms. Vijayalaxmi Bhisikar[1], Ms. Nisarga Chafle[2]
[1,2]*Final Year B.E Dept. of CSE, SRMCEW, RTMNU, Nagpur, India.*

Abstract: ***Data deduplication is a method for removing multiple copies of same data, and has been widely used in cloud storage to reduce storage space and uploading speed. However, there is only one copy for each file stored in cloud even if such a file is owned by a multiple number of users. As a result, deduplication system improves storage utilization while reducing reliability. Mostly, the challenge of privacy for sensitive data also arises when they are outsourced by users to cloud. Aiming to address the above security challenges, this paper makes the first attempt to formalize the notion of reliable deduplication system. We propose new distributed deduplication systems with higher reliability in which the data chunks are distributed across multiple cloud servers. The data confidentiality and tag consistency of data are also achieved by introducing a deterministic secret sharing scheme in distributed storage systems, instead of using convergent encryption as in previous deduplication systems. Security analysis depicts that our deduplication systems are secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement the proposed systems and demonstrate that the incurred overhead is very limited in realistic environments.***

Keywords: Deduplication, Reliability, Distributed System.

I. INTRODUCTION

With the huge growth of digital data, deduplication techniques are widely employed to backup data and minimize network and storage overhead by detecting and eliminating redundancy among data. 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 has received much attention from academics and industry because it can greatly improves storage utilization and save storage space, especially for the applications with high deduplication ratio such as storage systems. A number of deduplication systems have been proposed based on various deduplication strategies such as client-side or server-side deduplications, file-level or block-level deduplications. A brief review is given in Section 6. Especially, with the advent of cloud storage, data deduplication techniques become more attractive and critical for the management of ever-increasing volumes of data in cloud storage services which motivates enterprises and organizations to outsource data storage. To third-party cloud providers, as evidenced by many real-life case studies . According to the analysis report of IDC, the volume of data in the world is expected to reach 40 trillion gigabytes in 2020 . Today's commercial cloud storage services, such as Drop box, Google Drive. have been applying deduplication to save the network bandwidth and the storage cost with client-side deduplication. There are two methods of deduplication in terms of the size: (i) file-level deduplication, which discovers redundancies between different files and removes these redundancies to reduce capacity demands, and (ii) block level deduplication, which discovers and removes redundancies between data blocks. The file can be divided into smaller fixed-size or variable-size blocks. Using fixedsize blocks simplifies the computations of block boundaries, while using variable-size blocks (e.g., based on Rabin fingerprinting [3]) provides better deduplication efficiency. Though deduplication technique can save a large amount of storage space for the cloud storage service providers, it reduces the reliability of the system. Data reliability is actually a very critical issue in a deduplication storage system because there is only one part of each file stored in the server shared by all the owners. If such a shared file/chunk was lost, a disproportionately large amount of data becomes inaccessible because of the unavailability of all the files that share this file/chunk. If the value of a file were measured in terms of the amount of file data that would be lost in case of losing a single file, then the amount of user data lost when a file in the storage system is corrupted

A. Cloud Computing

Cloud computing is an Internet based development of computer technology Cloud.Cloud computing combine virtualization, on-demand deployment, Internet delivery of services, and open source software. Cloud computing uses various approaches, concepts, and best practices that have been established Cloud computing is architecture for providing computing service via the internet on

demand and pay per use access to a pool of shared resources viz. various types of networks, storage, servers, services and applications, without physically acquiring.

II. LITERATURE SURVEY

A. *Cloud Computing Security: From Single to Multi-Clouds*

Once of the outcomes that they propose is to use a Byzantine blemish tolerant replication tradition inside the cloud. Hendricks et al. express that this outcome can sidestep data pollution made by a couple parts in the cloud. On the other hand, Cachinet al. declare that using the Byzantine blemish tolerant replication tradition inside the cloud is inadmissible in light of the way that the servers having a spot with cloud suppliers use the same structure foundations and are physically set in the same spot [1]. According to Garfinkel, another security danger that may happen with a cloud supplier, for instance, the Amazon cloud organization, is a hacked mystery key or data intrusion. If some person becomes acquainted with an Amazon account mystery key, they will have the ability to get to most of the account's events and resources. In spite of the way that cloud suppliers are aware of the noxious insider risk, they expect that they have essential responses for alleviate the issue [1]. Rocha and Correia [1] center possible aggressors for Iaas cloud suppliers. For outline, Grosse et al. [1] propose one outcome is to keep any physical access to the servers. In any case, Rocha and Correia [1] battle that the aggressors depicted in their work have remote get to and needn't trouble with any physical access to the servers. Grosse et al. [1] propose a substitute result is to screen OK to get access to the servers in a cloud where the customer's data is secured. In any case, Rocha and Correia [1] declare that this segment is profitable for watching laborer's behavior to the extent whether they are after the assurance course of action of the association or not, in any case it is not fruitful in light of the way that it distinguishes the issue after it has happened.

B. *Ensuring Data Integrity and Security In Cloud Storage*

A substitute way to deal with secures the data using various pressing and encryption computations and to disguise its region from the customers that stores and recuperates it. The primary complexity is that the system presented by OlfaNasraoui [2] is an application based structure like which will keep running on the clients own system. This application will allow customers to exchange record of different associations with security quirks including Encryption and Compression. The exchanged records may be gotten to from wherever using the application which is given. The security of the OlfaNasraoui [2] model has been examination on the reason of their encryption estimation and the key organization. It has been watched that the encryption count have their own specific qualities; one computation gives security to the detriment of fittings, other is strong however uses more number of keys, one takes also taking care of time. This region exhibits the diverse parameters which accept a vital part while selecting the cryptographic computation. The Algorithm found most ensuring is AES Algorithm with 256 bit key size (256k) [2].

C. *Factors Affecting the Adoption of Cloud Computing: An Exploratory Study*

Lukas Malina and Jan Hajny present a novel security ensuring security answer for cloud organizations. They oversee customer anonymous access to cloud advantages and conferred stockpiling servers. Their answer outfits enlisted customers with anonymous access to cloud organizations. Our answer offers anonymous confirmation. This infers that customers' near and dear qualities (age, authentic enlistment, powerful portion) can be exhibited without revealing customers' identity. In this way, customers can useorganizations with no danger of profiling their behavior. On the other hand, if customers break supplier's deals with, their entitlement to get access rights is renounced. They dismember current insurance securing responses for cloud organizations and structure our answer centered around dynamic cryptographic fragments. Their answer offers unacknowledged access, un-join limit and the security of transmitted data. What's more, we complete our answer and we yield the test comes to fruition and differentiation the execution and related courses of action [16].

D. *Key-Aggregate Cryptosystem for Scalable Data Sharing In Cloud Storage IEEE Transactions on Parallel and Distributed Systems*

Bryan Ford discussed on exchange issues of circulated processing like ice burgs in cloud. Conveyed registering is drawing in from organization and efficiency perspectives,however brings threats both known and dark. Understood and fervently information security perils, in light of programming vulnerabilities, insider ambushes, and side-channels for case, may be only the tip of the ice sheet. As different, unreservedly made cloud organizations confer interminably easily and compellingly multiplexed gear resource pools, unpredictable associations between weight conforming and other delicate instruments could incite component insecurities or emergencies. Non-direct layering structures, where choice cloud organizations may appear to be independent yet give significant,

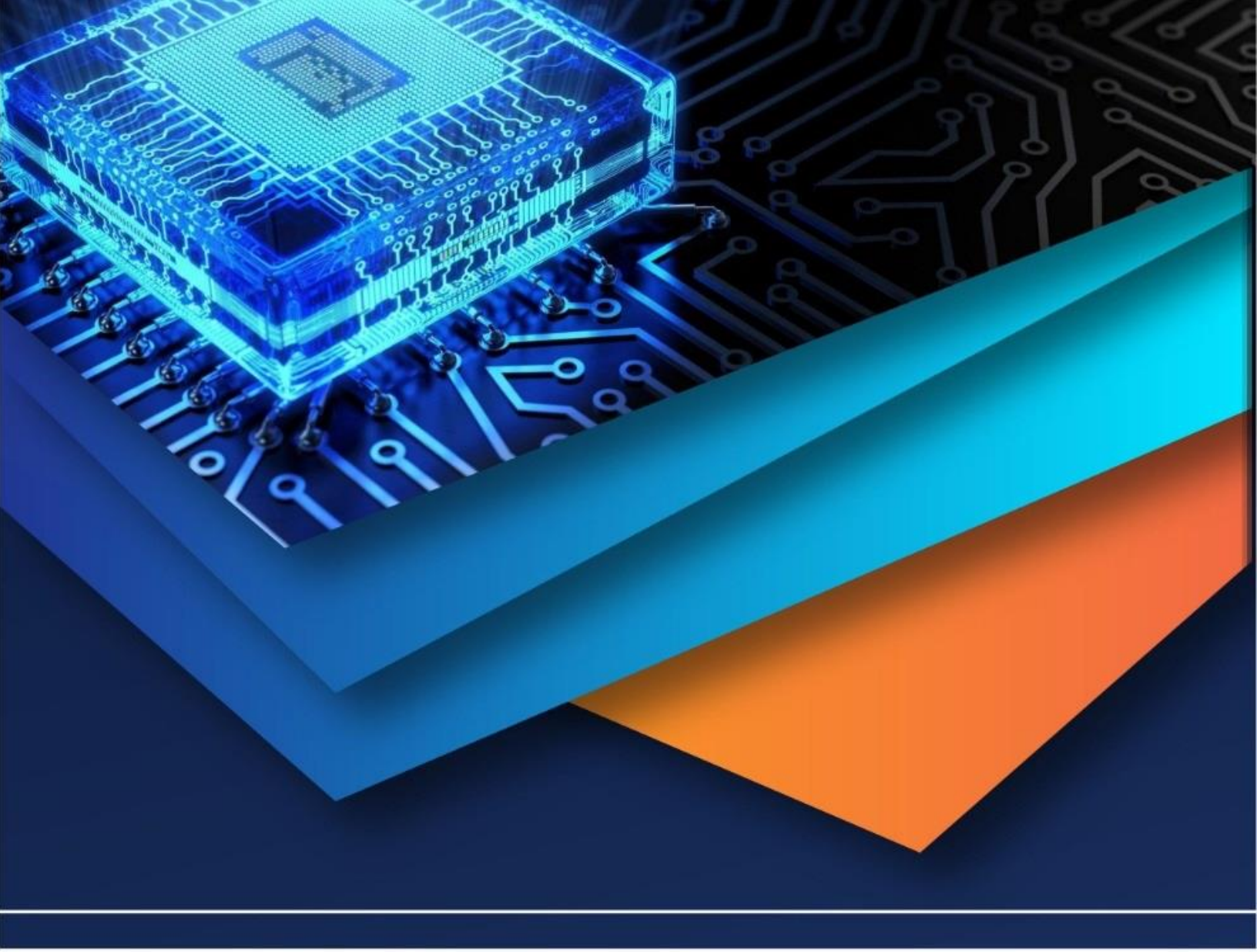
disguised resource conditions may make startling and possibly tragic dissatisfaction connections, reminiscent of budgetary industry crashes. Finally, dispersed processing mixes successfully troublesome propelled preservation challenges, in light of the way that simply the supplier of a cloud-based application or organization can account a live, utilitarian copy of a cloud knick-knack and its data for whole deal social defending. This paper explores these by and large un-saw risks, displaying the guard that we should study them before our monetary fabric gets the opportunity to be indistinguishably dependent on a favorable however possibly unstable preparing model [19].

III. CONCLUSION

This Project implements the distributed de-duplication systems to improve the Reliability of data while achieving the Confidentiality of the users' outsourced data without an encryption mechanism. It support file-level and fine-grained block-level data de-duplication. The security of tag consistency and integrity is achieved. We implemented our de-duplication systems using 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.

REFERENCES

[1] Mohammed A. Alzain , Eric Pardede , Ben Soh , James A. Thom "Cloud Computing Security: From Single To Multi-Clouds", 45th Hawaii International Conference On System Sciences 2012.
[2] OlfaNasraoui, Member, IEEE, MahaSoliman, Member, IEEE, EsinSaka, Member, IEEE, Antonio Badia, Member, IEEE, And Richard Germain "Ensuring Data Integrity And Security In Cloud Storage", IEEE TRANSACTIONS ON CLOUD AND DATA ENGINEERING, VOL. 20, No. 2, February 2013.
[3] Qin Liu, ChiuC.Tan, Jiewu, and Guojun Wang "Reliable Re-Encryption in Unreliable Clouds", IEEE Communications Society Subject Matter Experts For Publication In The IEEE Globecom 2011 Proceedings.
[4] Wei-Tek Tsai, Xin Sun, JanakaBalasooriya "Service-Oriented Cloud Computing Architecture", 2010 Seventh International Conference On Information Technology
[5] Cheng-Kang Chu, Sherman S. M. Chow, Wen-GueyTzeng, Jianying Zhou, And Robert H. Deng, Senior Member, IEEE "Key-Aggregate Cryptosystem For Scalable Data Sharing In Cloud Storage", IEEE Transactions On Parallel And Distributed Systems. Volume: 25, Issue: 2. Year: 2014
[6] Mell-Peter, Grance-Timothy "The NIST Definition Of Cloud Computing", September 2011.
[7] C. Cachin, I. Keidar And A. Shraer "Trusting The Cloud", ACM SIGACT News, 40, 2009, Pp. 81-86. Clavister, "Security in The Cloud", Clavister White Paper, 2008.
[8] H.Mei, J. Dawei, L. GuoliangAnd Z. Yuan "Supporting Database Applications As A Service", ICDE'09:Proc. 25thintl.Conf. On Data Engineering, 2009, Pp. 832-843.C. Wang, Q. Wang, K. Ren and W. Lou, "Ensuring Data Storage Security In Cloud Computing", ARTCOM'10: Proc. Intl. Conf. On Advances In Recent Technologies In Communication And Computing, 2010, Pp. 1-9.
[9] Keiko Hashizume, David G Rosado, Eduardo Fernandez-Medina And Eduardo B Fernandez "An Analysis Of Security Issues For Cloud Computing", Hashizume Et Al. Journal Of Internet Services And Applications 2013.
[10] Gehana Booth, Andrew Soknacki, and Anil Somayaji "Cloud Security: Attacks and Current Defenses", 8th ANNUAL SYMPOSIUM ON INFORMATION ASSURANCE (ASIA'13), JUNE 4-5, 2013, ALBANY, NY.
[11] BrentLagesse"Challenges In Securing The Interface Between the Cloud And PervasiveSystems" IEEE Pervasive Computing, Vol. 8, Pp. 14-23, October 2009. [Online].
[12] C. Wang, Q. Wang, K. Ren and W. Lou, "Ensuring Data Storage Security in Cloud Computing", ARTCOM'10: Proc. Intl. Conf. On Advances In Recent Technologies In Communication And Computing, 2010, Pp. 1-9.
[13] Wayne A. Jansen Cloud Hooks: "Security And Privacy Issues In Cloud Computing Proceedings", Of The 44th Hawaii International Conference On System Sciences-2011.
[14] LukasMalina and Jan Hajny "Efficient Security Solution for Privacy-Preserving Cloud Services", 6TH INTERNATIONAL CONFERENCE ON TELECOMMUNICATIONS SIGNAL PROCESSING YEAR 201
[15] SushmitaRuj, Milos Stojmenovic, AmiyaNayak "Decentralized Access Control With Anonymous Authentication Of Data Stored In Clouds", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS VOL:25 NO:2 YEAR 2014.
[16] Morgan, Lorraine Conboy, Kieran "FACTORS AFFECTING THE ADOPTION OF CLOUD COMPUTING: AN EXPLORATORY STUDY", Proceedings of the 21st European Conference on Information Systems 2012
[17] SaritaMotghare, P.S.Mohod "International Journal of Advanced Research In Computer Science", Volume 4, No. 4, March-April 2013
[18] Bryan Ford "Icebergs in the Clouds: The Other Risks Of Cloud Computing" SIGCOMM, August 2010
[19] Cheng-Kang Chu, Sherman S. M. Chow, Wen-GueyTzeng, Jianying Zhou, And Robert H. Deng "Key Aggregate Cryptosystem For Scalable Data Sharing In Cloud Storage IEEE Transactions On Parallel And Distributed Systems. "Volume: 25, Issue: Year: 2014.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY