IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

Avoiding Duplication of Encrypted Data Using Cloud

Ashweta Magar[1], Trupti Jagtap[2], Pradnya Gaikwad[3], Rashmi Singh[4]

Student, Computer Dept, ISB&M School of Technology, Pune, India [1, 2, 3, 4]

Abstract: Deduplication may be a storage saving technique that has been adopted by several cloud storage suppliers like Dropbox. In cloud storage services, deduplication technology are commonly accustomed cut back the world and {knowledge and data} live necessities of services by eliminating redundant knowledge and storing entirely one copy of them. Deduplication is best once multiple users supply an identical data to the cloud storage, but it raises issues concerning security and possession. Issues over information security still forestall several users from migrating information to remote storage. The standard resolution is to write in code the info before it leaves the owner's premises. Client-side information deduplication specifically ensures that multiple transfers of constant contentsolely consume network information measure and space for storing of one upload. We'll use server facet information deduplication. During this paper we have a tendency to planned novel server-side deduplication theme for encrypted information.

Keywords: Access control, big data, cloud computing, data deduplication, proxy re-encryption.

1. INTRODUCTION

CLOUD computing offers a replacement approach of data Technology services by rearranging numerous resources (e.g., storage, computing) and providing them to users based on their demands. Cloud computing provides an enormous resource pool by linking network resources along. Its fascinating properties, like measurability, elasticity, fault-tolerance, and pay-per-use. Thus, it's become a promising service platform. The most vital and widespread cloud service is information storage service. Cloud users transfer personal or confidential information to the information center of a Cloud Service supplier (CSP) and allow it to take care of these information. Since intrusions and attacks towards sensitive information at CSParen't evitable [3], it's prudent to assume that CSP can't be totally sure by cloud users. Moreover, the loss of management over their own personal information [1], [2] results in high information security risks, particularly information privacy leakages. Thanks to the fast development of knowledge mining and different analysis technologies, the privacy issue becomes serious [2]. Hence, a decent observe is to solely source encrypted information to the cloud so as to confirm information security and user privacy [5]. Deduplication has proven to realize highcost savings, e.g., reducing up to 90-95 p.c storage wants for backup applications [9] and up to sixty eight p.c in customary file systems [10]. Obviously, the savings, which maybe passed back directly or indirectly to cloud users, square measure important to the political economy of cloud business. {How to the approach to|a way to} manage encrypted information storage with deduplication in Associate in Nursing economical way could be a sensible issue. However, current industrial deduplication solutions cannot handle encrypted information. Existing solutions for deduplication suffer from brute-force attacks [7]. They can't flexibly support information access management and revocation at an equivalent time [4], Most existing solutions cannot guarantee responsible, security and privacy with sound performance.

2. LITERATURE SURVEY

1] Sun Yat-sen University, Kaohsiung, Taiwan Shi-Yuan Huang Taiwan Wen-Che Hsu National Sun Yat-sen University, Kaohsiung, Taiwan," Encrypted Data Deduplication in Cloud Storage", 2015

Data deduplication is a specialized data compression technique which makes all the data owners, who upload the same data, share a single copy of duplicate data and eliminates the duplicate copies in the storage. When users upload their data, the cloud storage server will check whether the uploaded data have been stored or not. If the data have not been stored, it will be actually written in the storage; otherwise, the cloud storage server only stores a pointer, which points to the first stored copy, instead of storing the whole data. Hence, it can avoid the same data being stored repeatedly.

2] Zheng Yan, Mingjun Wang, and Yuxiang Li, Xidian, Athanasios V. Vasilakos," Encrypted Data Management with Deduplication in Cloud Computing", 2325-6095/16/$33.00 © 2016 IEEE

Cloud computing offers a new way to deliver services by rearranging resources over the Internet and providing them to users on demand. It plays an important role in supporting data storage, processing, and management in the Internet of Things (IoT). Various cloud service providers (CSPs) offer huge volumes of storage to

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

maintain and manage IoT data, which can include videos, photos, and personal health records.

3] Junbeom Hur, Dongyoung Koo, Youngjoo Shin, and Kyungtae Kang," Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage", 1041-4347 (c) 2016 IEEE.

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 randomized convergent encryption and secure ownership group key distribution. This prevents data leakage not only to revoked users even though they previously owned that data, but also to an honest-but-curious cloud storage server. In addition, the proposed scheme guarantees data integrity against any tag inconsistency attack. Thus, security is enhanced in the proposed scheme. The efficiency analysis results demonstrate that the proposed scheme is almost as efficient as the previous schemes, while the additional computational overhead is negligible.

4] Chuan-Mu Tseng, Jheng-Rong Ciou, Tzong-Jye Liu," A cluster-based data deduplication technology", 2014 Second International Symposium on Computing and Networking
The proposed method excludes bloom filter's false positives that do not need to check all the index tables. It only needs to query one index table, effectively reducing the time to exclude the false positives.

Our approach:-
In this paper, we propose a scheme based on data ownership challenge and Proxy Re-Encryption (PRE) to manage encrypted data storage with deduplication. We aim to solve the issue of deduplication in the situation where the data holder is not available or difficult to get involved. Meanwhile, the performance of data deduplication in our scheme is not influenced by the size of data, thus applicable for big data.

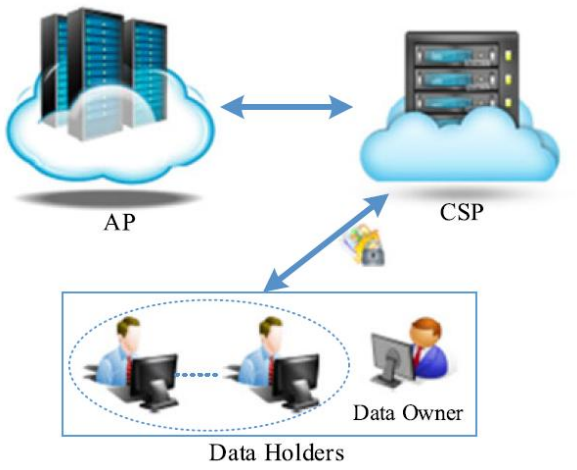
3. PROPOSED SYSTEM

I. System introduction
We propose Dekey, a brand new construction during which users don't have to be compelled to manage any keys on their own however instead firmly distribute the oblique key shares across multiple servers. Dekey victimisation the Ramp secret sharing theme and demonstrate that Dekey incurs restricted overhead in realistic environments we tend to propose a brand new construction referred to as Dekey, that provides potency and responsibility guarantees for oblique key management on each user and cloud storage sides. a brand new construction Dekey is planned to supply economical and reliable oblique key management through oblique key Deduplication and secret sharing. Dekey supports each file-level Deduplication. Security analysis demonstrates that Dekey is secure in terms of the definitions per the planned security model. Above all, Dekey remains secure even the individual controls a restricted range of key servers. we tend to implement Dekey victimisation the key sharing theme that allows the key management to adapt to totally different responsibility and confidentiality levels. Our analysis demonstrates that Dekey incurs restricted overhead in traditional upload/download operations in realistic cloud environments.

II. System module
- Data Deduplication
- Data Deletion
- Data Owner Management
- Encrypted Data Update
- Valid Data Replication

III. System features
- We motivate to save cloud storage and preserve the privacy of data holders by proposing a scheme to manage encrypted data storage with deduplication.
- Our scheme can flexibly support data sharing with deduplication even when the data holder is offline, and it does not intrude the privacy of data holders.
- We propose an effective approach to verify data ownership and check duplicate storage with secure challenge and big data support.
- We integrate cloud data deduplication with data access control in a simple way, thus reconciling data deduplication and encryption.
- We prove the security and assess the performance of the proposed scheme through analysis and simulation. The results show its efficiency, effectiveness and applicability.

Block diagram



Mathematical module
• Mathematical formulation if possible
Set of Input = {U, S, D, C}
Set of Output= {Data}
Server = {IP, MAC}

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

Where
Msg is non repeated Data
U is a set of users
u1, u2, u3.....un € U
There can be number of users in our system which interacts with our developed system and uses features of our system.
S is the Web Server
D is set of Devices
d1, d2,.....dn€ D
IP is an IP address of Server (S)
MAC is MAC address of Server (S)

4. CONCLUSION

Managing encrypted information with deduplication is very important and vital in observe for achieving a prospering cloud storage service, particularly for giant information storage. during this paper, One of the feature is, information is in encrypted kind thus privacy of user is maintained. we have a tendency to projected a sensible theme to manage the encrypted massive information in cloud with deduplication supported possession challenge and PRE. Our theme will flexibly support information update and sharing with deduplication even once the info holder's area unit offline. Encrypted information will be firmly accessed as a result of solely approved information holders will acquire the regular keys used for information secret writing. Intensive performance analysis and take a look at showed that our theme is secure and economical underneath there presented security model and really appropriate for giant information deduplication.

5. FUTURE SCOPE

We will try to save storage space on cloud in order to avoid traffic on server.

REFERENCES

[1] M. Dutch, "Understanding data deduplication ratios," SNIA Data Management Forum, 2008.
[2] W. K. Ng, W. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage," Proc. ACM SAC'12, 2012.
[3] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," Proc. StorageSS'08, 2008.
[4] N. Baracaldo, E. Androulaki, J. Glider, A. Sorniotti, "Reconciling end-to-end confidentiality and data reduction in cloud storage," Proc. ACM Workshop on Cloud Computing Security, pp. 21–32, 2014.
[5] P. S. S. Council, "PCI SSC data security standards overview," 2013.
[6] D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side channels in cloud services, the case of deduplication in cloud storage," IEEE Security & Privacy, vol. 8, no. 6, pp. 40–47, 2010.
[7] C. Wang, Z. Qin, J. Peng, and J. Wang, "A novel encryption scheme for data deduplication system," Proc. International Conference on Communications, Circuits and Ssytems (ICCCAS), pp. 265–269, 2010.
[8] Malicious insider attacks to rise, http://news.bbc.co.uk/2/hi/ 7875904.stm
[9] Data theft linked to ex-employees, http://www.theaustralian. com.au/australian-it/datatheftlinked-to-ex-employees/story-e6f rgakx-1226572351953
[10] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system," Proc. International Conference on Distributed Computing Systems (ICDCS), pp. 617–624, 2002