ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
IJARSMT
Volume 1, Issue 7, December 2015

A Review on Secure Authorized Deduplication Based Hybrid Cloud using Compression Techniques

Ashwini Garole[1], Sharmila Wagh[2]

Dept. Of CSE, Modern Education Society's College of Engineering, Pune, Maharashtra, India[1]
Dept. Of CSE, Modern Education Society's College of Engineering, Pune, Maharashtra, India[2]

ABSTRACT— Data deduplication is one of the important data compression techniques for eliminating duplicate copies of repeating data and has been widely used in cloud storage in order to minimize the amount of storage space and save bandwidth. On the other hand, there have been recently wide privacy considerations as data could come in contact with those third party servers also to unauthorized parties. To protect the confidentiality of important data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check beside the data itself. In this review we report a prototype of proposed authorized duplicate check scheme and conduct tested experiments using various prototype and tries to minimize the data duplication that occurs in hybrid cloud storage by using various techniques**.**

KEYWORDS- Deduplication, authorized duplicate check, hybrid cloud, confidentiality.

I. INTRODUCTION

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. Although data deduplication brings a lot of benefits, security and privacy concerns arise as users' sensitive data are susceptible to both insider and outsider attacks. Traditional encryption, while providing data confidentiality, is incompatible with data deduplication. Specifically, traditional

ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
IJARSMT
Volume 1, Issue 7, December 2015

encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different cipher texts, making deduplication impossible.

II. LITERATURE SURVEY

Convergent encryption [1] for making the feasible deduplication and maintain the data confidentiality used convergent encryption technique. It encrypts decrypts a data copy with a convergent key, the content of the data copy obtained by computing the cryptographic hash value of. After the data encryption and key generation process users retain the keys and send the cipher text to the cloud. Since the encryption operation is determinative and is derived from the data content, similar data copies will generate the same convergent key and hence the same cipher text. A secure proof of ownership protocol is used to prevent the unauthorized access and also provide the proof to user regarding the duplicate is found of the same file. Authorized deduplication technique is proven by Jin Li [2] which avoid the duplicate content in cloud storage system and incurs minimal overhead as compared to the normal operation by using convergent key encryption. It also provides the security to the given data. C.Ng [3] presented reverse deduplication technique for read to latest backup .W. K. Ng, Y. Wen, and H. Zhu[4]. Proposes private data deduplication Protocols in cloud storage for Enhance the efficiency of data S. Halevi, D. Harnik [5] proposes Proofs of Ownership in Remote Storage Systems which contain Performance measurements indicate that the scheme incurs only a small overhead compared to naive client-side deduplication. Bugiel [6] presented twin clouds: architecture for secure cloud computing for Client uses the trusted Cloud as a proxy that provides a clearly defined interface to manage the outsourced data, programs, and queries. Token generation technique and identity based signature for provide security to the give data in cloud storage.[7]. A hybrid cloud is a combination of private cloud and public cloud in which the data which is most critical that resides on a private cloud and the data which is easily accessible is resides on a public cloud hybrid cloud is helpful for reliability, extensibility and fast deployment and cost saving of public cloud with more security with private cloud [8]. The complex challenge of cloud storage or cloud computing is the arrangement of large volume of data duplication is a process of eliminating of duplicate data in de-duplication techniques redundant data removed leaving single instance of the data to be stored. In the previous old system the data is encrypted back to outsourcing [9]. By using de-duplication technique in hybrid cloud the encryption technique becomes simpler. As we all know that the network has large amount of data which is being shared by many users. Many large networks uses data cloud to store the data and share that data on the network [10].

III. RELATED WORK

However, previous deduplication systems could not support differential authorization duplicate check, which was important in many applications. In such 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. [11] 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. For example, in a company, many different privileges will be assigned to employees. In order to save cost and efficiently management, the data will be moved to the storage server provider (S-CSP) in the public cloud with specified privileges and the deduplication

ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
IJARSMT
Volume 1, Issue 7, December 2015

technique will be applied to store only one copy of the same file. Because of privacy consideration, some files will be encrypted and allowed the duplicate check by employees with specified privileges to realize the access control. [12]Traditional deduplication systems based on convergent encryption, although providing confidentiality to some extent; do not support the duplicate check with differential privileges. In other words, no differential privileges have been considered in the deduplication based on convergent encryption technique. It seems to be contradicted if we want to realize both deduplication and differential authorization duplicate check at the same time.

Several deduplication schemes have been proposed showing how deduplication allows very appealing reductions in the usage of storage resources. The problems of coalescing and identifying identical files in the distributed file system, for the purpose of reclaiming storage space consumed by incidentally redundant content.[13] Recently proposed an encryption scheme that guarantees semantic security for unpopular data and provides weaker security and better storage and bandwidth benefits for popular data.[14] Recently, reported DupLESS a server-aided encryption for deduplicated storage which uses a modified convergent encryption scheme with the aid of a secure component for key generation [15] Researchers also presented the hybrid cloud techniques to support privacy-aware data-intensive computing. Proposed system considers addressing the authorized deduplication problem over data in public cloud. [16-17] the author reported that the use of multi-cloud providers to maintain security has received less attention from the research community than has the use of single clouds. Their work aims to promote the use of multi-clouds due to its ability to reduce security risks that affect the cloud computing user. [18] Security architecture for cloud to secure the system using client side encryption and compression. Their application allows them to upload and download files which are encrypted and decrypted at client side. Encryption algorithm used is AES 128. [19] They describe new public-key cryptosystems which produce constant-size cipher texts such that efficient delegation of decryption rights for any set of cipher texts are possible. [20] This research paper also analyzes the key research and challenges that presents in cloud computing and Offers best practices to service providers as well as enterprises hoping to leverage cloud service to improve their bottom line in this severe economic climate. [21]. Security and privacy are among top concerns for the public cloud environments. Towards these security challenges proposed a new client-side deduplication scheme for securely storing and sharing outsourced data via the public cloud. It ensures better confidentiality towards unauthorized users. Every client computes a per data key to encrypt the data that he wants to store in the cloud. As such, the data access is managed by the data owner and also introduces a new cryptographic method for secure Proof of Ownership (PoW), for improving data security in cloud storage systems, providing dynamic sharing between users and ensuring efficient data deduplication.[22]

IV. EXISTING WORK

In the existing system, the system protecting the data confidentiality by transforming the predictable message into unpredictable message. Here another third party called key server is introduced to generate the file tag for duplicate check in which the data confidentiality is not more secured. Convergent encryption ensures data privacy in deduplication.The system is formalized this primitive as message-locked encryption, and explored its application in space-efficient secure outsourced storage in which the system will take more time duration to encrypt and decrypt.The jobs arrival pattern is not predictable and the capacities of each node in the cloud differ, for the load balancing problem,

ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
IJARSMT
Volume 1, Issue 7, December 2015

and workload control is difficult to improve system performance and maintain stability. Cloud computing is efficient and scalable but maintaining the stability of processing so many jobs in the cloud computing environment is a very complex problem with load balancing receiving much attention for researchers.

V. PROPOSED WORK

In our system we implement a project that includes the public cloud and the private cloud and also the hybrid cloud which is a combination of the both public cloud and private cloud. In general if we used the public cloud we can't provide the security to our private data and hence our private data will be at loss. So that [23] we have to provide the security to our data for that we make a use of private cloud also. When we use private clouds the greater security can be provided. In this system we also provide the data deduplications. This is used to avoid the duplicate copies of data. User can upload and download the files from public cloud but private cloud provides the security for that data. That means only the authorized person can upload and download [24] the files from the public cloud. For that user generates the key and stores that key onto the private cloud. At the time of downloading user requests to the private cloud to access that particular file.

System Model: Now we see the architecture of our system. in our architecture there are three modules .

[1] User

[2] Public cloud

[3] Private cloud.etc

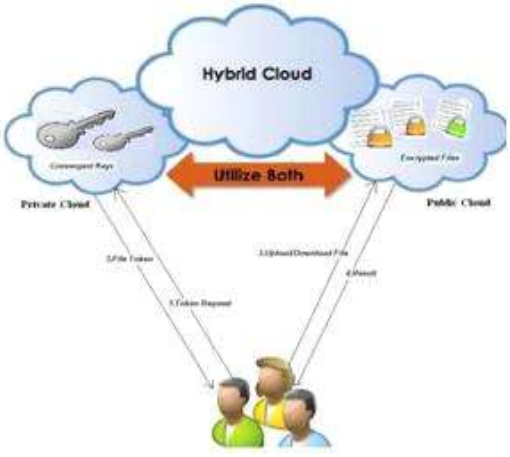

Fig 1: Architecture of Authorized Deduplication

VI. CONCLUSION

Various new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, the duplicate-check tokens of files are generated by the private cloud server with keys. Security analysis demonstrates that developed schemes are secure for insider and outsider attacks specified in the proposed security model. As a confirmation, we implemented a prototype of our proposed authorized duplicate check scheme and carried out tested experiments on our prototype. We reported our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer. Hybrid clouds offer a good flexibility to businesses while offering

ISSN (Online) : 2454-4159
International Journal of Advanced Research in Science Management and Technology
IJARSMT
Volume 1, Issue 7, December 2015

choice in terms of keeping control and security. Hybrid clouds are usually deployed by the organizations willing to push part of their workloads to public clouds either for cloud bursting purposes or for projects requiring faster implementation because hybrid clouds varies as the needs of company and structure of implementation.

REFERENCES

[1] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. Reclaiming space from duplicate files in a serverless distributed file system. In *ICDCS*, pages 617–624, 2002.

[2] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.

[3] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.

[4] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S Ossowski and P. 2012.

[5] R. D. Pietro and A. Sorniotti . Boosting efficiency and security in proof of ownership for deduplication. In H. Y. Youm and Y. Won, editors, ACM Symposium on Information, Computer and Communications Security 2012.

[6] 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.

[7] M. Bellare, C. Namprempre , and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 2009.

[8] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In *Proc. of USENIX LISA*, 2010.

[9] D. Ferraiolo and R. Kuhn. Role-based access controls. In *15th NIST- NCSC National Computer Security Conf.*, 1992.

[10] 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.

[11] 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.

[12] R. D. Pietro and A. Sorniotti. Boosting efficiency and security in proof of ownership for deduplication. In H. Y. Youm and Y. Won, editors, *ACM Symposium on Information, Computer and Communications Security*, pages 81–82. ACM, 2012.

[13] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer.(2002), Reclaiming space from duplicate files in a serverless distributed file system.,*In ICDCS*.

[14] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl.(2013), A secure data deduplication scheme for cloud storage, *In Technical Report*

[15] Bellare, M., Keelveedhi, S., Ristenpart, T.(*2013)* Message-locked encryption and secure deduplication, *In: Advances in Cryptology*

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

[17] K. Zhang, X. Zhou, Y. Chen, X.Wang, and Y. Ruan.(2011), Sedic: privacyaware data intensive computing on hybrid clouds. *In Proceedings of the 18th ACM conference on Computer and communications security, USA*

[18] A.AlZain et al., "Cloud Computing Security: From Single to Multi-Clouds" 2012 45th Hawaii International Conference on System Sciences.

[19] Olfa Nasraoui et al., "Ensuring Data Integrity And Security In Cloud Storage", IEEE, VOL. 20, No. 2, February 2013.

[20] Cheng-Kang Chu, et al., "Key-Aggregate Cryptosystem for Scalable Data Sharing in Cloud Storage" IEEE Transactions on Parallel and Distributed Systems. Volume: 25, Issue: 2. Year :2014

[21] Rabi Prasad Padhy, et al., 2011 vol. 1 No. 2 "Cloud Computing: Security Issues and Research Challenges" IJCSITS conference.

[22] Nesrine Kaaniche, Maryline Laurent(2014), A Secure Client Side Deduplication Scheme in Cloud Storage Environments. *6th international conference on new technologies,mobility and security.*

[23] 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.

[24] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In *IEEE Transactions on Parallel and Distributed Systems*, 2013.