A NOVEL Authorized De-duplication MECHANISM Based on Hybrid Cloud

[1]K.Lavanya , [2]G.Chandrika , [3]D.Padma
Assistant Professor, Dadi Institute of Engineering and Technology
klavanya@diet.edu.in, chandrika@diet.edu.in, dpadma@diet.edu.in

Abstract
Data deduplication is a crucial data compression technique for getting rid of multiple copies of repeated data, and it is frequently applied in cloud storage to conserve bandwidth and reduce the quantity of storage space. Before outsourcing, it has been suggested that the data be encrypted to ensure the secrecy. This study attempts to formally address the issue of authorised data deduplication for the first time in order to better safeguard data security. In contrast to conventional deduplication systems, duplicate check also takes into account user privilege differences in addition to the data itself. In a hybrid cloud architecture, we also propose a number of innovative deduplication constructs that provide authorised duplicate checking. According to the definitions given in the suggested security model, security analysis shows that our method is secure. as evidence of

Index Terms: Hybrid cloud, authorised duplicate check, and deduplication.

Introduction

In order to reduce the amount of identical copies of the same data, data de-duplication is an essential data compression technique used in cloud computing. This process is used in conjunction with the de-duplication approach, which locates and stores identical data, to minimise data transmission over networks, data during the process of analysis, and to increase effective use of storage space.

Other data are compared to the stored copy as the process goes on, and anytime a match is made, the identical data is replaced with a brief reference to the stored data. A hybrid cloud is made up of both private and public clouds, with the most important data being stored on a private cloud and the easiest-to-access data being stored on a public cloud. Hybrid clouds are advantageous for public clouds' dependability, extensibility, rapid deployment, and cost-savings while providing higher security for private clouds [1], [2].

The organisation of a huge volume of data is a challenging task for cloud computing or storage. De-duplication techniques remove superfluous data from the remaining data as part of the process of removing duplicate data, network, but frequently what happens is that the data being downloaded and posted on the network contain the same data. In these situations, data confidentiality and cloud security are disturbed.

By eliminating data redundancy, the hybrid cloud offers the functionality, scalability, dependability, quick deployment, and cost-savings of public cloud storage [4].

Literature Survey

Due to the enormous amount of duplicate or redundant data present in archival storage systems, which consume a significant amount of additional hardware and energy, resource utilisation (such as network bandwidth and storage) is significantly reduced, adding to the management burden as scale increases. Data de-duplication, whose objective is to reduce duplicate data at the interlevel, has so received a lot of attention recently in both academia and business. This study proposes semantic data de-duplication (SDD), which uses the semantic information in the I/O path of the archived files (such as file type, file format, application hints, and system metadata) to drive the splitting of a file into semantic chunks (SC). the primary objective of

In order to create the file tag for the duplicate check, a new third party called key server is established. A novel encryption method that offers both popular and unpopular material the necessary security has been presented by Stanek et al. The classic ordinary encryption is used for widely used data that is not extremely sensitive. For unpopular data, a different two-layered encryption approach that supports deduplication and has higher security is recommended. They were able to better balance productivity and security of the outsourced data in this way. Block-level deduplication's key management problem was solved by Liet al. by dispersing the keys among several servers after the files had been encrypted.

Existing System

Data owners and users can run duplicate checks safely and with varied levels of access thanks to data deduplication systems that employ the private cloud as a middleman. Such a design is useful and has caught the interest of many academics. The data operation

is handled in a private cloud, and the data owners only outsource their data storage to the public cloud.

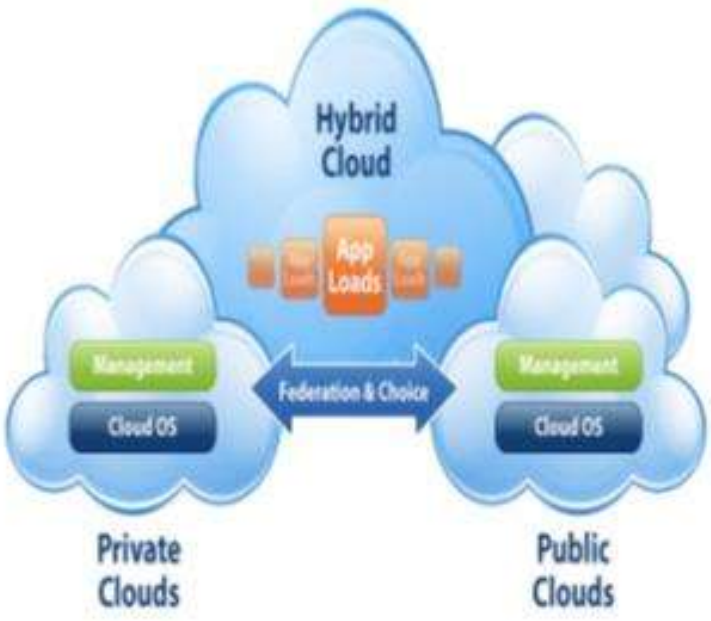
Hybrid Cloud for Secure Deduplication

At a high level, the environment we are interested in is an enterprise network, which consists of a number of affiliated clients (for example, workers at a corporation) that will utilise the S-CSP and store data using the deduplication approach. Deduplication can be effectively employed in these circumstances to significantly reduce the amount of storage needed for applications like data backup and disaster recovery. These systems are common and frequently better suited than deeper storage abstractions for user file backup and synchronisation applications. In our system, three entities have been defined: users, private clouds, and S-CSP in public clouds. By determining if the contents of two files are identical and storing only one of them, the S-CSP conducts deduplication. A file's access privilege is 01-01, allowing her to open any file whose access role is "Director" and accessible time period is January 2014 through January 1.

Each permission is represented by a token, which is a brief message. Each file has a few file tokens attached to it that signify the given tag. Duplicatecheck tokens are computed and sent by the user.

For a legitimate duplication check, to the public cloud. By creating file tokens for the asking users, the private cloud server, a semitrusted third party, will assist in performing deduplicable encryption. The function of the private cloud server will be covered in more detail below. Additionally, per-user encryption keys and credentials are provided to users.

For the sake of simplicity, we shall solely discuss file level deduplication in this work. In other words, we refer to a data copy as a complete file and file-level deduplication that gets rid of any superfluous files from storage. In reality, file-level deduplication can be used to readily determine block-level deduplication. In particular, before uploading a file, a user runs a file-level duplicate check. When a file is duplicated, all.



• S-CSP. This organisation offers a public cloud data storage service. In addition to storing data on behalf of consumers, the S-CSP offers a data outsourcing service. The S-CSP uses deduplication to remove redundant data from storage and retains only unique data in order to lower storage costs. In this study, we assume that S-CSP has a lot of storage and processing power and is always online.

Data consumers. An entity that wants to contract with the S-CSP for the storage of data it will later access is referred to as a user. In a storage system that supports deduplication, the user only uploads unique data—which may belong to the same user or to other users—and does not post any duplicate data in order to conserve upload bandwidth.

In actuality, recent times have seen a growing amount of interest in this hybrid cloud environment. For instance, a business might continue to maintain internal storage for operational customer data while using a public cloud service, such Amazon S3, for archived data.

A cluster of virtualized cryptographic co-processors that are provided as a service by a third party and offer the required hardware-based security characteristics to construct a remote execution environment trusted by the users might also serve as the trusted private cloud.

Conclusion

Data deduplication with permission is a concept. It was suggested that the differential privileges of users be included in the duplicate check in order to safeguard data security. Additionally, we provided a number of innovative deduplication designs that facilitate authorised duplicate check in hybrid cloud architecture and produce duplicate-check tokens for files, by the personal cloud server using personal keys. According to the suggested security model's insider and outsider attack specifications, security analysis shows that our methods are secure. We put our suggested authorised duplicate check technique into practise as a proof of concept and ran test bed experiments on our prototype. We demonstrated that, when compared to convergent encryption and network transfer, our authorised duplicate check technique incurs the least amount of overhead.

References

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

[2] M. Bellare, S. Keelveedhi, and T. Ristenpart.Message- locked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart.Dupless:Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.

[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] 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.

[6] Bugiel, S., N¨urnberger, S., Sadeghi, A.-R., Schneider, T.: Twin Clouds: An architecture for secure cloud computing (Extended Abstract). In: Workshop on Cryptography and Security in Clouds (WCSC 2011), March 15-16 (2011)

[7] Chung, K.-M., Kalai, Y., Vadhan, S.: Improved delegation of computation using fully homomorphic encryption. In: Rabin, T. (ed.) CRYPTO 2010. LNCS, vol. 6223, pp. 483–501. Springer, Heidelberg (2010)

[8] Cloud Security Alliance. Top threats to cloud computing, v. 1.0 (2010)