ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 3, pp: (355-358), Month: July - September 2015, Available at: www.researchpublish.com

Eliminating Redundancy in Cloud – Databases Using Authorized Hybrid Cloud Approach

Satya Sudheer Varma Surimalla[1], Krishna. B[2]

[1]Student, [2]Assosciate Professor (M.Tech, CSE), Department of CSE Visakha Institute of Engineering and Technology, Visakhapatnam, India

Abstract: **Cloud computing promises to increase the rapidness of application deployment and delivery. Authorized Data De-duplication in cloud computing deals with elimination of excessive usage storage space due to duplicated data and providing privacy for user's .Each authorized user will get an individual token to access their file in the cloud and can perform duplicate check based on the privileges. Authorized user is able to use his/her own private keys to generate query and hence attributes are attached along with the file. Attributes are found in the private cloud; so, control immediately passes to the private cloud, where duplicate check can be performed. Data stored in the public cloud is accessed only by the authorized users with the help of different individual encryption privilege keys. Convergent and Symmetric encryption techniques produce identical cipher text that results in minimum overhead. Proof of reliability assures a verifier via an acknowledgement that a user's file is available.**

Keywords: **Convergent encryption, De-duplication, duplicate check, hybrid cloud.**

I. INTRODUCTION

Cloud computing provides a more cost effective and hassle free means to outsource data and computations. As cloud computing is becoming more prevalent in the present day scenario, there is a sharp rise in the volume of data that is being stored in the cloud resources and the data that is in the cloud is shared by multiple users with some specified privileges, which provide the access rights to a particular individual or a group of users. The privileges may vary with different applications such as privileges based on the role of user or privileges based on the time of usage. So the major problem faced by cloud storage providers is the management of increasing volume of data on the cloud along with providing an authorization of data and security for the data that is being accessed or stored. Data deduplication [5] is a common simple technique to make data management on the cloud scalable. Data deduplication is specialised compression technique to remove the redundant copies of repeating data in the cloud storage. The technique is will thus enhance the storage utilization and can also be useful for transferring the data over the network by reducing the number of bytes that are being sent. Although, data deduplication provides a lot of benefits, there are some security and privacy concerns as the sensitive data is susceptible to both insider and outsider attacks. So, to overcome these issues a special encryption technique called Convergent Key [6] encryption is also included in this approach.

This paper mainly focuses on two issues related to cloud storage: 1. Guarantee of authorized access: Guarantee of authorized access implies that only authorized person can access the data. 2. Hybrid cloud approach: Hybrid cloud approach for security of user's data.

II. LITERATURE SURVEY

Cloud computing is now an emerging market. Day by day application hosting on cloud increases rapidly causes huge data storage on cloud. Due to this the main challenge faced by cloud service provider is the management of this ever increasing bulk data.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 3, pp: (355-358), Month: July - September 2015, Available at: www.researchpublish.com

In archival storage systems, there are a lot of duplicate data copies or redundant data, which occupy unnecessary storage space which hinders the resource - utilization (such as the network bandwidth and storage) which results in extra burden on the cloud users. So for the data de-duplication, the goal of which is to minimize the duplicate data in the inter level sharing in a multi user environment, has been receiving broad attention both in research and industry in recent years. In this paper, we propose a Semantic Data De-duplication (SDD) is proposed, which makes use of the semantic information in the I/O path (such as file type, file format, application hints and system metadata) of the archival files to direct the dividing a file into semantic chunks (SC). While the main goal of SDD is to maximally reduce the inter file level duplications, directly storing variable SC's into disks will result in a lot of fragments and involve a high percentage of random disk accesses, which is very inefficient. So an efficient data storage scheme is also designed and implemented: SC's are further packaged into fixed sized Objects, which are actually the storage units in the storage devices, so as to speed up the I/O performance as well as ease the data management. Primary experiments have demonstrated that SDD can further reduce the storage space compared with current methods. With the advent of cloud computing, secure data deduplication has attracted much attention recently from research

Harnik, D. [1] proposes cross-user and multi user deduplication with a trust-based security mechanism which implements data redundancy elimination in shared data environment but it does not take security of sensitive data into view in considering the various security attacks and vulnerabilities which occur inwards and outwards the cloud.

Yunchuan Sun, Junsheng Zhang [2] – presents an approach towards data security and privacy in cloud computing which is only limited to private data but it did not take public clouds into consideration. Dawei Sun , Guiran Chang, Lina Sun, Xingwei Wang [3] propose a system for analyzing security, privacy and trust issues in cloud computing environments which only focuses on security issues but did not consider elimination of redundancy.On the similar lines, P. Anderson and L. Zhang [4] proposes a redundancy elimination system for laptop and mobile backup systems. The backup taken is in compressed and encrypted format. This paper mainly focuses on increase the speed of backup, and reduces the storage requirements.

III. OVERVIEW OF THE HYBRID CLOUD CONCEPTS

HYBRID CLOUD:

A hybrid cloud is a cloud computing environment in which an organization provides and manages some resources in-house and has other resources provided externally .For example, an organization might use a public cloud service, such as Amazon Web Services(Amazon S3) for archived data but continue to maintain in-house storage for operational customer data

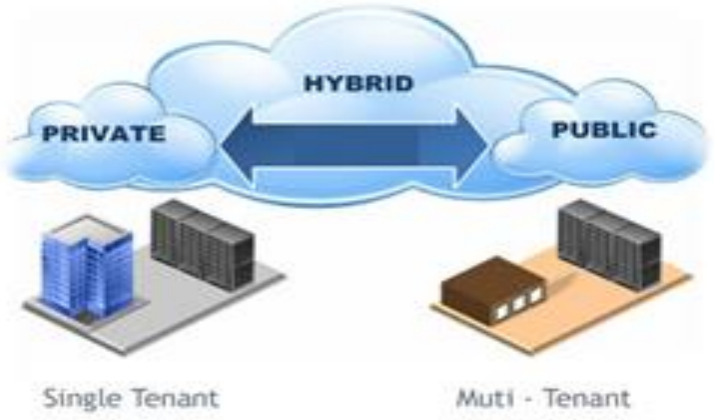

Fig: Hybrid Cloud Environment

The concept of a hybrid cloud is meant to bridge the gap between high level control, high cost "private cloud" and highly scalable, flexible and low cost "public cloud". "Private Cloud", for example, VMware deployment in which the hardware and software of the entire cloud environment is used and managed by a single entity. The concept of "Public cloud"

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 3, pp: (355-358), Month: July - September 2015, Available at: www.researchpublish.com

usually involves some form of subscription based resource pools [7] in a hosting provider data center that utilizes multi-tenant policy. The term public cloud doesn't mean less security, but instead refers to multi-tenancy. The concept is introduced to enhance connectivity and data portability. VMware has a key tool for "hybrid cloud" use called "vCloud connector". It is a free plug-in that allows the management of public and private clouds within the vSphere client. The tool offers users the ability to manage the console

IV. HYBRID CLOUD FOR SECURE DEDUPLICATION

At a higher levels i.e., in enterprise networks, consisting a group of affiliated clients (for example, employees of a company) who will use the S-CSP (Storage Cloud Service Provider) and store data with deduplication technique. In this context, deduplication can be frequently used in this environment for data backup, archiving and disaster recovery applications while greatly reducing storage space. Such systems are widespread and are often more suitable to user file backup and synchronization applications than general storage purposes. There are three entities defined in our system; they are users, private cloud and S-CSP in public cloud. The S-CSP performs deduplication by comparing the contents of two files for similarities in the content and stores only one copy of them. The access right to a file will defined based on a set of access privileges. The exact definition of a privilege varies across applications. For example, we may define a role-based privilege according to job positions (e.g., Director, Manager, and Engineer), or we may define a time-based privilege that specifies a valid time period (e.g., 01/01/2015 to 31/12/2015) within which a file can be accessed. A user, say Alice, may be assigned two privileges "Director" and "access right valid up to 01/02/2015", so that she can access any file whose access role is "Director" and accessible time period covers 01/02/2015. Each privilege is represented in the form of a short message called token. Each file is associated with some file tokens, which denote the tag with which the token is specified. A user computes and sends duplicate-check tokens to the public cloud for authorized duplicate check. Users have access to the private cloud server, which is a semi-trusted third party which will aid in performing deduplicable encryption by generating file tokens for the requesting users.

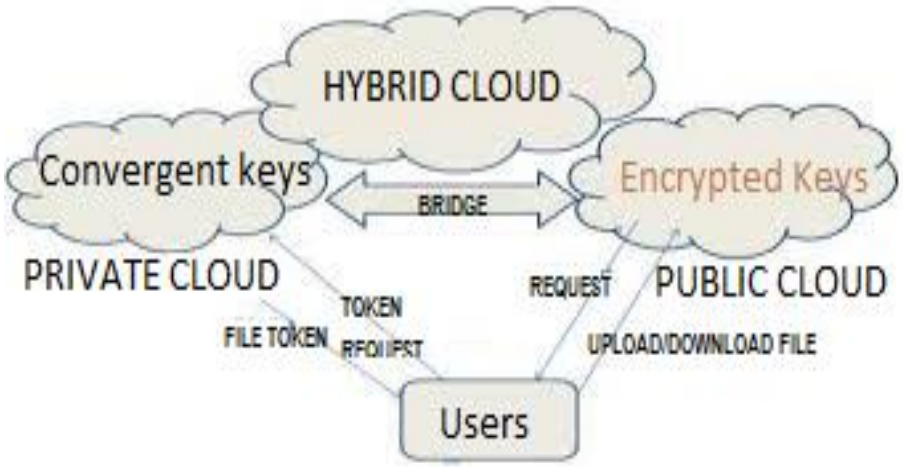

Fig: Architecture of the proposed system

In this paper, we have taken the consideration of the file level deduplication for simplicity. For this, we refer a data copy to be a whole file and file-level deduplication which eliminates the storage of the redundant files. Operationally, to upload a file, a user first performs the file-level duplicate check by sending a request to the CSP. If the file is a duplicate, then all its blocks must be duplicates as well; otherwise, the user further performs the block-level duplicate check and identifies the unique blocks to be uploaded. Each data copy (i.e., a file or a block) is associated with a token for the duplicate check.

• S-CSP. This is an entity that provides a data storage service in public cloud. The S-CSP provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via deduplication and keeps only unique data. In this paper, we assume that S-CSP is always online and has abundant storage capacity and computation power.

• Data Users. A user is an entity that wants to outsource data storage to the S-CSP and access the data later. In a storage system supporting deduplication, the user only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users. In the authorized deduplication system, each user is issued a set of privileges in the setup of the system. Each file is protected with the convergent encryption key and privilege keys to realize the authorized deduplication with differential privileges.

• Private Cloud. Compared with the traditional deduplication architecture in cloud computing, this is a new entity introduced for facilitating user's secure usage of cloud service. Specifically, since the computing resources at data user/owner side are restricted and the public cloud is not fully trusted in practice, private cloud is able to provide data user/owner with an execution environment and infrastructure working as an interface between user and the public cloud. The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users. The interface offered by the private cloud allows user to submit files and queries to be securely stored and computed respectively. This is a novel architecture for data deduplication in cloud computing, which consists of a twin clouds (i.e., the public cloud and the private cloud). Actually, this hybrid cloud concept has attracted more attention from the industry recently. Alternatively, the trusted private cloud could be a cluster of virtual cryptographic co-processors, which are offered as a service by a third party and provide the necessary hardware based security features to implements a remote execution environment trusted by the users.

V. CONCLUSION

The concept of authorized data deduplication was proposed to protect the data security by including differential privileges of users along with an inclusion of duplicate check. We also presented several new deduplication mechanisms that support authorized duplicate check in hybrid cloud architecture, in which the duplicate check tokens of files are generated by the private cloud server with private keys. The security analysis of proposed system demonstrates that our schemes are secure in terms of insider and outsider attacks which are more susceptible in most of the existing systems. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct test-bed experiments on our prototype with the help of Drop-Box API and CloudHQ. We conclude that our authorized duplicate check scheme incurs minimal overhead as observed over convergent encryption and network transfer.

REFERENCES

[1] Harnik, D. ,IBM Haifa Res. Lab., Haifa, Israel Pinkas, B. ; Shulman-Peleg, A. Side Channels in Cloud Services: Deduplication in Cloud Storage,Volume:8 Issue:6 Date Nov.-Dec. 2010,

[2] Yunchuan Sun, Junsheng Zhang., International Journal of Distributed Sensor NetworksVolume, Article ID 190903,2014

[3] Dawei Sun , Guiran Chang, Lina Sun, Xingwei Wang , Surveying and Analyzing Security, Privacy and Trust Issues in Cloud Computing Environments, Procedia EngineeringVolume 15, 2011.

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

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