International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 12, December 2015
Available at http://internationaljournalofresearch.org

A Hybrid Cloud Approach for Secure Authorized Reduplication

Kolakani Purnachandra Rao [1]& M. Venkata Reddy [2]

[1] M.Tech Student, Eswar College of Engineering, Narasaraopet, Guntur.

[2] Asst. Professor & Project Supervisor, Eswar College of Engineering, Narasaraopet, Guntur.

Abstract

Data reduplication is one of important data compression techniques for eliminating duplicate copies of repeating data, and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting reduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data reduplication. Different from traditional reduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. We also present several new reduplication constructions supporting authorized duplicate check in a hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct tested experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

I. Introduction

Cloud computing provides seemingly unlimited "virtualized" resources to users as services across the whole Internet, while hiding platform and implementation details. Today's cloud service providers offer both highly available storage and massively parallel computing resources at relatively low costs. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and shared by users with specified *privileges*, which define the access rights of the stored data. One critical challenge of cloud storage services is the management of the ever-increasing volume of data.

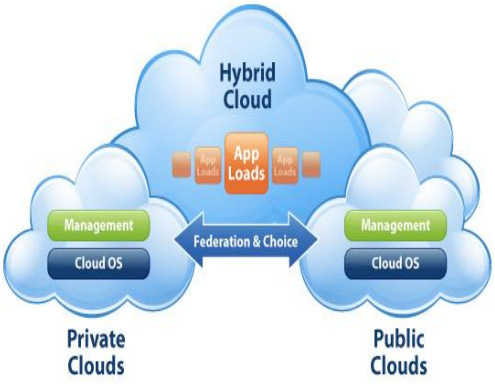
To make data management scalable in cloud computing, reduplication [17] has been a well-known technique and has attracted more and more attention recently. Data reduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. The technique is used 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, reduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy.

II. Literature Survey

In archival storage systems, there is a huge amount of duplicate data or redundant data, which occupy significant extra equipments and power consumptions, largely lowering down resources utilization (such as the network bandwidth and storage) and imposing extra burden on management as the scale increases. So data de-duplication, the goal of which is to minimize the duplicate data in the inter level, has been receiving broad attention both in academic and industry in recent years. In this paper, 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 SCes 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: SCes 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 reduplication has attracted much attention recently from research community. Proposed a reduplication system in the cloud storage to reduce the storage size of the tags for integrity check. To enhance the security of reduplication and protect the data confidentiality, Bellare et al. showed how to protect the data confidentiality by transforming the predictable message into unpredictable message. In their system, another third party called key server is introduced to generate the file tag for duplicate check. Stanek et al. presented a novel encryption scheme that provides the essential security for popular data and unpopular data. For popular data that are not particularly sensitive, the traditional conventional encryption is performed. Another two-layered encryption scheme with stronger security while supporting reduplication is proposed for unpopular data. In this way, they achieved better trade between the efficiency and security of the out-sourced data. Liet al. addressed the key management issue in block-level reduplication by distributing these keys across multiple servers after encrypting the files.

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 others provided externally .For example, an organization might use a public cloud service, such as Amazon Simple Storage Service(Amazon S3) for archived data but continue to maintain in house storage for operational customer data.



The concept of a hybrid cloud is meant to bridge the gap between high control, high cost “private cloud” and highly callable , flexible , low cost “public cloud”. “Private Cloud” is normally used to describe a VMware deployment in which the hardware and software of the environment is used and managed by a single entity. The concept of a “Public cloud” usually involves some form of elastic/subscription based resource pools in a hosting provider datacenter that utilizes multi-tenancy. The term public cloud doesn’t mean less security, but instead refers to multi-tenancy. The concept revolves heavily around connectivity and data portability. The use cases are numerous: resource burst-ability for seasonal demand, development and testing on a uniform platform without consuming local resources, disaster recovery, and of course excess capacity to make better use of or free up local consumption. 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 view, power status, and more from a “workloads” tab, and offers the ability to copy virtual machine templates to and from a remote public cloud offering.

IV. Hybrid Cloud for Secure Deduplication

At a high level, our setting of interest is an enterprise network, consisting of a group of affiliated clients (for example, employees of a company) who will use the S-CSP and store data with reduplication technique. In this setting, reduplication can be frequently used in these settings for data backup 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 richer storage abstractions. There are three entities defined in our system, that is, *users*, *private cloud* and S-CSP in *public cloud* . The S-CSP performs reduplication by checking if the contents of two files are the same and stores only one of them. The access right to a file is defined based on a set of *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, Project Lead, and Engineer), or we may define a *time-based* privilege that specifies a valid time period (e.g., 2014-01-01 to 2014-01-31) within which a file can be accessed. A user, say Alice, may be assigned two privileges "Director" and "access right valid on 2014- 01-01", so that she can access any file whose access role is "Director" and accessible time period covers 2014-01- 01. 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 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, a semi trusted third party which will aid in performing reduplicable encryption by generating file tokens for the requesting users. We will explain further the role of the private cloud server below. Users are also provisioned with per-user encryption keys and credentials.

A. Architecture For Authorized Reduplication:

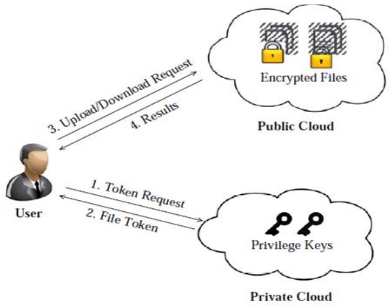


Fig Architecture for Authorized reduplication

This paper, we will only consider the file level reduplication for simplicity. In another word, we refer a data copy to be a whole file and file-level reduplication which eliminates the storage of any redundant files. Actually, block-level reduplication can be easily deduced from file-level reduplication, Specifically, to upload a file, a user first performs the file-level duplicate check. 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 reduplication 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 reduplication, 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 reduplication 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 reduplication with differential privileges.

Private Cloud. Compared with the traditional reduplication 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.

Notice that this is a novel architecture for data reduplication in cloud computing, which consists of a twin clouds (i.e., the public cloud and the private cloud). Actually, this hybrid cloud setting has attracted more and more attention recently. For example, an enterprise might use a public cloud service, such as Amazon S3, for archived data, but continue to maintain in-house storage for operational customer data. Alternatively, the trusted private cloud could be a cluster of virtualized cryptographic co-processors, which are offered as a service by a third party and provide the necessary hardware based security features to implement a remote execution environment trusted by the users.

Public Cloud: Public cloud entity is used for the storage purpose. User upload the files in public cloud. Public cloud is similar as S-CSP. When the user want to download the files from public cloud, it will be ask the key which is generated or stored in private cloud. When the users key is match with files key at that time user can download the file, without key user can not access the file. Only authorized user can access the file. In public cloud all files are stored in encrypted format. If any chance unauthorized person hack our file, but without the secrete or convergent key he doesn't access original file. On public cloud there are lots of files are store each user access its respective file if its token matches with S-CSP server token.

Operations performed on Hybrid Cloud

File Uploading : When user want to upload the file to the public cloud then user first encrypt the file which is to be upload by make a use of the symmetric key, and send it to the Public cloud. At the same time user generates the key for that file and sends it to the private cloud. in this way user can upload the file in to the public cloud.

File Downloading: When user wants to download the file that he/she has upload on the public cloud. he/she make a request to the public cloud. then public cloud provide a list of files that many users are upload on it. Among that user select one of the file form the list of files and enter the download option. at that time private cloud sends a message that enter the key for the file generated by the user. then user enters the key for the file that he/she is generated.

Then private cloud checks the key for that file and if the key is correct that means the user is valid. only then and then the user can download the file from the public cloud otherwise user can't download the file. When user download the file from the public cloud it is in the encrypted format then user decrypt that file by using the same symmetric key.

V. Review

The notion of authorized data reduplication was proposed to protect the data security by including differential privileges of users in the duplicate check. We also presented several new reduplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate check tokens of files are generated by the private cloud serve with private keys. Security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model.

As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct tested experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

VI. Conclusion

In this paper, the idea of authorized data reduplication was proposed to protect the data security by including differential authority of users in the duplicate check. In public cloud our data are securely store in encrypted format, and also in private cloud our key is store with respective file. There is no need to user remember the key. So without key anyone can not access our file or data from public cloud.

VII. Future Scope

It excludes the security problems that may arise in the practical deployment of the present model. Also, it increases the national security. It saves the memory by reduplicating the data and thus provide us with sufficient memory. It provides authorization to the private firms and protect the confidentiality of the important data.

VIII. 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 reduplication. In EUROCRYPT, pages 296–312, 2013..

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serve raided encryption for reduplicated 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 reduplication 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).

IX. Authors Profile

Kolakani Purnachandra Rao is a M.Tech(CSE) pursuing student in Eswar college of Engineering, Narasaraopet, Guntur.

M. Venkata Reddy M.Tech in Computer Science & Engineering. He is presently working as an Asst. Prof. in Eswar College of Engineering, Narasaraopet, Guntur, India. He is having about 6 years of teaching experience.