Design and Analysis of the Performance of a Hybrid Cloud for Secure Data Storage with Data De-Duplication

M.PRIYA BHARATHI [#1], T.ANURADHA [*2], D.V.S RAVI VARMA [*3]

M.Tech Scholar [#1], Assistant Professor *[2], Assistant Professor *[3]

Department of Computer Science & Engineering,
Raghu Engineering College,
Dakamarri (V), Visakhapatnam, AP, India.

Abstract

Cloud Computing is one in each of the follow of using a group of remote servers hosted on internet to store, retrieve and access the information from the remote systems not from the local machines. As all the data is going to be stored on remote server, whenever any user who want to access the data, he will retrieve the data from that remote server. In the current cloud servers there is no facility like avoiding the duplicate data to be stored into the cloud, so this lead a major problem like huge maintenance cost while storing the data inside the cloud server. As he is accessing the data through a remote server, he need to pay the amount on rental basis, that is the reason why the cloud us also known as PAUZ (I.e. Pay As you UZe) server. So in order to avoid this knowledge/data duplication we'd like to use a replacement principle referred to as Data Deduplication.This is one among the simplest knowledge compression technique that was used for eliminating the duplicate copies of perennial knowledge and this was wide employed in recent cloud storage. Along with this de-duplication technique the data confidentiality must also be applied for the stored data, which is not available in the current cloud servers. So in this paper we have implemented any of the primitive cryptography technique to convert the plain text file into cipher text file and then store into the cloud server. As an extension we also implemented a new concept like granting access privileges for the user like read/ write and modify access by the private cloud at the time of user authorization after the initial registration. Also as an extension the admin will monitor all the user activities like upload, download and modify date and time details of the entire session ,which is not available in any of the current cloud servers.

Keywords

Authorization, Maintenance, Confidentiality, Data De-duplication, Encryption, Cryptography Algorithm, PAUZ.

1. Introduction

In the current days cloud computing plays a very crucial role in almost all aspects of real time environments like MNC companies, Many Industries, schools, hospitals, software companies and so on. Generally there are various types of services in the cloud computing which differ mainly with one another. The main services that are available in

the cloud computing domain are discussed in details as follows:

a) Infrastructure as a Service(IaaS)
b) Platform as a Service (PaaS)
c) Software as a Service(SaaS)

In the primitive cloud domain there is mainly three types of services which was developed based on different models. But as the cloud usage is increased a lot, there were many other services which were evolved. One among the newly evolved cloud services are as follows:

d) DataBase as a Service (DaaS)
e) Database Security/Secure as a Service(DSaaS)

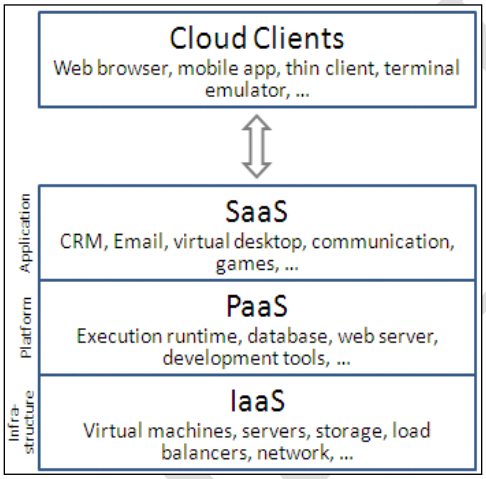

Figure. 1. Represents the Various Cloud Clients and their Individual Services

As we know that in the context of cloud server ,the cloud users try to store a crucial or sensitive information into a semi trusted infrastructures of third party cloud servers, there is no guarantee for that data is secure [1], [2]. This demand imposes clear data management choices: Original plain data ought to be accessible only by reliable parties that do not embrace cloud suppliers, intermediaries, and Internet; in any untrusted context, data ought to be encrypted. Satisfying these goals has fully completely different levels of quality betting on the type of cloud service. There are several solutions guaranteeing confidentiality for the storage as a service paradigm (e.g., [3], [4], [5]), whereas guaranteeing confidentiality at intervals the data as a service (DaaS) paradigm [6] remains associate open analysis house. Throughout this context, we have a tendency to propose a brand new Secure DaaS as a result of the primary answer that allows cloud tenants to require full advantage of DaaS qualities, like handiness, dependability, and elastic quality, while not exposing unencrypted data to the cloud provider.

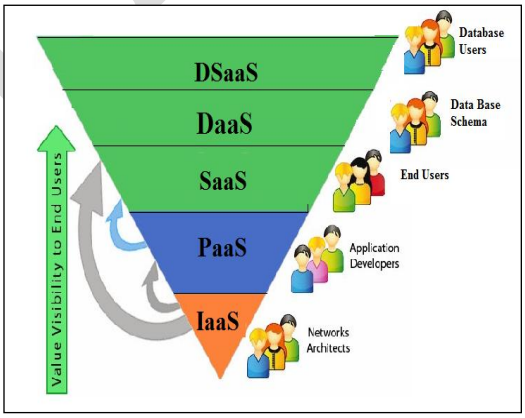

Figure. 2. Represents the Primitive Cloud Services along with DaaS and DSaaS

From the above figure 1, we can clearly find out the various types of cloud clients and their services like Software oriented services, Infrastructure oriented services and Platform oriented services. Now a day's almost all the MNCs or Industries are showing their

valuable interest in cloud computing in storing their valuable or sensitive data inside the cloud servers and try to access the data from that server at the time of need. As this cloud server came into the existence there was a lot of work reduction for the end users to store and update in their insite server. In the above figure, we found that web browsers, mobile app, emulator and terminals come under cloud clients. All these clients try to communicate with various applications like Email, CRM, Games and so on to perform their task under SaaS or Application as a service phase. And for executing or deployment of various services we need several development tools as well as data base and web servers for performing these operations, so all these will come under Platform as a Service. Now once the task is deployed on the cloud server, it should be access from a remote terminals with the help of virtual machines, load balancing devices and a many more which comes under IaaS service.

Also along with the above functionalities ,in this paper we try to make data management ascendible in cloud computing, a new technique called as data de-duplication [6] has been a widely deployed in order to attract more user's attention towards its usage. Data de-duplication is treated as one of the specialized knowledge compression technique for eliminating the duplicate files which are stored continuously in the cloud data base. De-duplication mainly eliminates the redundant knowledge by keeping only one original copy of data and removing the duplicate copies. De-duplication will happen at either the file level or the block level.

2. Related Work

In this section we will find the related work that was analyzed and studied in order to implement this current paper. This section will describe the work related to Data De-Duplication

2.1 About Data De-Duplication

De-duplication technique conjointly called as data De-duplication may well be a method of reducing storage needs by eliminating redundant information. Only one distinctive instance of the data is actually maintained on storage media, like disk or tape. Redundant information is replaced with a pointer to the distinctive information copy [7].

For example, a typical database server like MySQL or SQL Server may contain 200 instances of constant 1 GB (GB) file attachment. If that database server is saved or archived, all 200 instances unit of measurement saved, requiring 200 GB storage space. With information de-duplication, only one instance of the attachment is actually stored; each future instance is solely documented back to the one saved copy. Throughout this instance, a 200 GB storage demand can be reduced to only 10 GB as we may found all the remaining instances are duplicated in that server and by using this new technique we can able to reduce a lot of data storage .

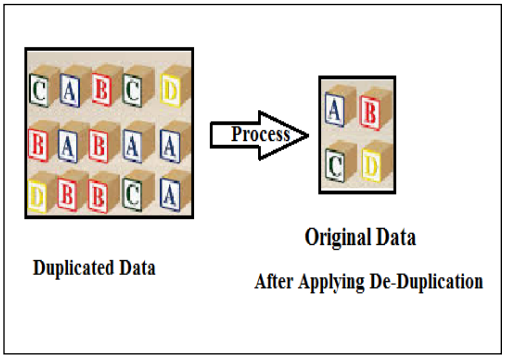

Figure. 3. Represents the Architecture of Data De-Duplication

From the figure 3, we can clearly find out the architecture flow of a data de-duplication technique and its advantages over the data base servers. For instance we took various alphabets as input with A, B, C, D as 4 alphabets and initially we took alphabets in a duplicated manner. Once after applying the de-duplication process all the duplicate letters are eliminated and only the distinct values like A, B, C, and D are only retrieved as a result at the other end. This example clearly justifies how much the data de-duplication has advantage over the various data base servers including the cloud database server.

3. Proposed Hybrid Model for a Secure Data De-Duplication

In this section, we describe a simple hybrid model for a secure data de-duplication along with proof of ownership model that was used in this current paper.

A) A Novel Model for Secure Data De-Duplication

In order to achieve data deduplication at a high level, we need to assume an enterprise network; consisting of a bunch of attached purchasers (for example, workers of a company) World Health Organization can use the S-CSP and store information with deduplication technique. In this setting, deduplication will be often utilized in these settings for information backup and disaster recovery applications in order to reduce a lot of duplicate wastage space. Such systems area unit widespread and area unit usually a lot of appropriate to user file backup and synchronization applications than richer storage abstractions. As per our requirement, system is divided into 3 entities like I.e. users, non-public cloud and S-CSP publically cloud as shown in Figure. 4. The S-CSP performs deduplication by verifying the contents between two users during the time of uploading.

The access right to a file is outlined supported a group of privileges. The precise definition of a privilege varies across applications. for instance, we have a tendency to might outline a role based privilege [6], [7] in line with job positions (For example , The Director, Project Lead, and The Engineer), or we have a tendency to might outline a time-based privilege that specifies a sound period (e.g., 2015-02-01 to 2015-02-28) among that a file will be accessed. A user, say Alice, is also appointed 2 privileges "The Director" and "access right valid on 2015- 02-01", so she will access any file whose access role is "The Director" and also accessible period which covers the year 2015- 02-01. Every privilege is portrayed within the type of a brief message known as token. Every file is related to some file tokens that denote the tag with nominative privileges. A user computes and sends duplicate-check tokens to the general public cloud for authorized duplicate check.

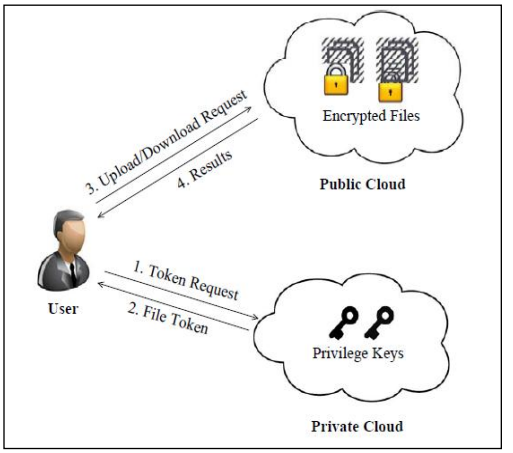

Figure. 4. Represents the Architecture of Data Dedupliation Environment Model

In this paper, we'll solely take into account the file level deduplication for simplicity. In another word, we refer a knowledge copy to be a full file and file-level deduplication which eliminates the storage of any redundant files. Actually, block-level deduplication will be simply deduced from file-level deduplication that is analogous to [8]. Specifically, to transfer a file, a user initially performs the file-level duplicate check. If the file may be a duplicate, then all its blocks should be duplicates as well; otherwise, the user can perform the block-level duplicate check and find out the best blocks which are not at all duplicated or copied. Each data copy (i.e., a file or a block) is related to a token for the duplicate check.

S-CSP: This is often AN entity that has a knowledge storage service publically cloud. The S-CSP provides the data outsourcing service and stores knowledge on behalf of the users. To scale back the storage price, the S-CSP eliminates the storage of redundant knowledge via deduplication and keeps solely distinctive knowledge. During this paper, we assume that S-CSP is usually on-line and has abundant storage capability and computation power [9].

Knowledge Users: A user is an entity that desires to source data storage to the S-CSP and access the data later. In a very storage system supporting deduplication, the user solely uploads distinctive knowledge however will not transfer any duplicate knowledge to avoid wasting the transfer bandwidth, which can be closely-held by a similar user or completely different users. Within the licensed deduplication system. Here in this system each and every user is provided with a set of privileges within the setup of the system. Every file is protected with the convergent secret writing key and privilege keys to appreciate the licensed deduplication with differential privileges[10].

Non-public Cloud : Compared with the standard 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 knowledge user/owner aspect square measure restricted and also the public cloud isn't absolutely trusty in observe, non-public cloud is ready to supply knowledge user/owner with AN execution setting and infrastructure operating as AN interface between user and the public cloud. The non-public keys for the privileges square measure managed by the non-public cloud, who answers the file token requests from the users. The interface offered by the non-public cloud permits user to submit files and queries to be firmly hold on and computed severally.

B) Proof of Owner Ship Model

The notion of Proof of oWnership (PoW) permits users to prove their possession of data copies to the storage server. Generally, PoW is implemented as an interactive formula (POW) pass a prover (i.e., user) and a friend (i.e., storage server). The friend derives a brief denoted by worth $\phi(M)$ from a knowledge copy M. To prove the possession of the data copy M, the prover has to send ϕ' to the friend such that

$$\phi' = \phi(M).$$

The formal security definition for prisoner of war roughly follows the threat model during a contentdistribution network, wherever AN assaulter doesn't apprehend the entire file, however has accomplices World Health Organization have the file.

4. Implementation Modules

Implementation is the stage where theoretical design is converted into practical manner. Generally in the implementation stage we will divide the application into number of modules in order to make the application develop very easily. We have implemented the proposed concept on Java Platform in order to show the performance this hybrid cloud approach. The front end of the application takes JSP, HTML and Back End takes My SQL Server 5.0 along with a Real Cloud Service provider called as DRIVEHQ Cloud Service provider. This cloud service provider will provide a space up to 2 GB for storing the files which is used by the application. The application is divided mainly into following 4 modules. They are as follows:

1) Admin Module
2) Private Cloud Module
3) Data User Module
4) De-Duplication Module

1) Admin Module

In this module, admin is the main person who has a facility to monitor all the activities that are been processed by cloud user. The facilities are like upload, download and update. Here the admin can view the accessing information like date, time and filename and system IP address where the file has been processed by user. The admin has no other facility other than monitoring the user activities. Here the admin can view the log details of the data user who try to access the data that is stored inside the cloud server.

2) Private Cloud Module

In this module the private cloud has the facility to grant access permission for the end users like read/write and update at the time of user getting access privilege. He is the person who initially sends the token for the data user and data owner in order to get login into the account with their details. If the cloud user or owner don't have token key then they are not allowed to access any of the files which are stored inside the cloud server. He can also view the uploaded file information. This private cloud access like an S-CSP in our current application where all the central authority lies in the hands of a private cloud.

3) Data User Module

In this module, the data user is the person who tries to access the data which is stored by the data owner. He initially get registered with all the basic details and then waits for the approval by the private cloud, once he gets acceptance from the private cloud then he will try to download the data or update the data based on his access rights given by the private cloud. If any user who don't have access privileges for any operation need to request the private cloud to grant the access for him, once the private cloud grant the access permission then only he can operate the data operations.

4) De-Duplication Module

In this module the data will not be duplicated to store in the cloud data base. If the file which is already kept in the server by an user like 'X' can't be stored once again by any other user including 'X'.This method is known as data de-duplication technique where the data once uploaded on cloud server can't be uploaded once again in the cloud server.

5. Conclusion

In this paper, we for the first time have design a new cloud server with advanced facilities like encryption of data which is stored into the cloud server and also an advanced feature like data de-duplication,in which no duplicate data is stored inside the cloud. For this we have used a live cloud service like drivehq cloud service in order to prove these advanced features that are proposed in this current paper. Our proposed system uses this new de-duplication technique in order to main concept for resource allocations. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

6. References

[1] W. Jansen and T. Grance, ―Guidelines on Security and Privacy inPublic Cloud Computing,‖ Technical Report Special Publication800-144, NIST, 2011.

[2] M. Armbrust et al., ―A read of Cloud Computing,‖ Comm. of theACM, vol. 53, no. 4, pp. 50-58, 2010.

[3] H. Hacigu¨mu¨ s¸, B. Iyer, C. Li, and S. Mehrotra, ―ExecutingSQL over Encrypted knowledge within the Database-Service-ProviderModel,‖ Proc. ACM SIGMOD Int'l Conf. Management knowledge, June2002.

[4] A.J. Feldman, W.P. Zeller, M.J. Freedman, and E.W. Felten,―SPORC: cluster Collaboration victimisation Untrusted Cloud Resources,‖Proc. Ninth USENIX Conf. operative Systems style andImplementation, Oct. 2010.

[5] R.A. Popa, C.M.S. Redfield, N. Zeldovich, and H. Balakrishnan,―CryptDB: protective Confidentiality with Encrypted QueryProcessing,‖ Proc. twenty third ACM Symp. operative Systems Principles,Oct. 2011.

[6] S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In *Proc. USENIX FAST*, Jan 2002.

[7] 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.

[8] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. *J. Cryptology*,22(1):1–61, 2009.

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

[10] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. Role-based access control models. *IEEE Computer*, 29:38–47, Feb 1996.

Issue 6, Vol. 6 (November 2016)
ISSN 2249-6149

7. About the Authors

M.PRIYA BHARATHI is currently pursuing her 2 years M.Tech in Department of Computer Science and Engineering at Raghu Engineering College, Dakamarri (V), Visakhapatnam, AP, India. Her area of interest includes Computer Networks and Software Engineering.

T.ANURADHA is currently working as an Assistant Professor in Department of Computer Science and Engineering at Raghu Engineering College, Dakamarri (V), Visakhapatnam, AP, India. Her research interest includes Data Mining, Computer Networks and Database Management Systems.

D.V.S.RAVI VARMA is currently working as an Assistant Professor in Department of Computer Science and Engineering at Raghu Engineering College, Dakamarri (V), Visakhapatnam, AP, India. His research interest includes Data Mining, Computer Networks and Computer Graphics.