Securing Data in Cloud Through Data Sharing And Deduplication

Maregowda R[1] Mahesh N[2]
[2]Department of Computer Science & Engineering
[1,2]CIT, Gubbi, Tumkur

Abstract— Cloud computing provides a set of resources and services offered through the Internet. Cloud service providers are delivering the service from data centers located throughout the world and providing virtual resources via internet. This services can be divided into 3 categories: IaaS, PaaS and SaaS. User's information is stored in the cloud globally. In cloud information storage framework, users store their information in the cloud and won't really have the information locally. Client's information is typically prepared remotely in obscure machines that clients don't possess or operate. In this paper influence the JAR programmable capacities to make a dynamic and traveling item, and to guarantee that any access to clients' information will trigger confirmation and generate the JAR locally. To make stronger client's control, likewise give appropriated reviewing components. In addition to this a privacy protection technique by name is proposed. This paper also introduces a structure for distributed accountability and reviewing, which is utilized to secure client's information and also tracking the actual usage of information in the cloud. Specifically in this paper, a logging mechanism is accommodated to provide the client's information with access policies, and guarantees that any access to their information will trigger confirmation, by this system data owner may know his/her information is taken care of according to his access policies. Information deduplication is one of important data compression strategies for eliminating with duplicates copies in the cloud storage to decrease the amount of storage space. To ensure the confidentiality of sensitive information while supporting deduplication, the convergent encryption strategy has been proposed to encrypt the information before outsourcing. To better information security, this paper addresses the issues of approved information deduplication.

Key words: Cloud computing, JAR Cryptography, Deduplication

I. INTRODUCTION

Cloud computing provides set of IT resources which are deployed in the network and deliver the services over a network. The organizations are pay per use these IT resources. User information is accessed from the hybrid cloud anywhere or anytime. The main characteristic is the cost effective. The large amount of information is stored in the cloud.

Cloud computing and virtualization are like two faces of the same coin. Virtualization is one of the most important concept in the cloud computing because users will access services on the cloud. There are two types of virtualizations that is fullvirtualization and paravirtualization. Fullvirtualization technique is the complete installation of one device is run on another device also the entire system is emulated. All software's run on the server inside a virtual machine. The paravirtualization technique is the numerous operating systems to run on distinct hardware device at the similar time.

Types of cloud computing:

Public Cloud: Public cloud is one of the standard cloud computing model in the hybrid cloud. Public cloud available to all users public over the Internet. The public cloud provides services to all users may be free or pay per usage model. The public cloud services provider makes resources such as application and storage. The examples of public clouds are Amazon Web Service(AWS), GoogleApp Engine etc. The benefits of public cloud are no wastage of resources because of users to pay what you use these resources.

Private Cloud: Private cloud is one of the cloud computing model provides deliver the services over the private network. Private cloud is a internal cloud or corporate cloud. It is a simple storage service provider also users data under the control of organization. The private cloud normally used by large organizations and their own private network. The benefits of private cloud computing is same as public cloud scalability, self-service, delivers users information to multiple systems.

Hybrid Cloud: The hybrid cloud is a cloud computing standard model consolidating information from both private cloud and public cloud will become the most popular selection for enterprises. As it were mixture cloud can likewise be characterized as the blend of both public and private cloud which is joined via VPN. The hybrid cloud service providers are AWS, GoogleApp Engine, Jelastic, CloudBees and Eucalyptus etc.

Users are information stored in the cloud using CIA framework in remote system. Users Information is stored in cloud correctness, reliability and accessibility of the information being put away on the appropriated cloud server must be guaranteed. The users approved information is stored in the cloud and unapproved information is can't be stored in the cloud. Registered users are information stored and retrieved from the cloud. Avoid the users information can't be altered from the any attacker[4]. The cloud computing service provider are AWS, GoogleApp engine windows azure. All cloud services provides web based online service.

The users are helpless before their cloud service providers for the ease of access and integrity of their information. Then again, in spite of the fact that the cloud frameworks are substantially more capable and dependable than individualized computing gadgets, expansive scope of both outer and inward dangers for information trustworthiness still exist. Users information can't be add duplicate files in the cloud. Eliminate duplicate files in the cloud. It is more point of interest for personal users information to store their information redundantly over numerous physical servers to lessen the information trustworthiness and accessibility dangers. Therefore appropriated conventions for capacity accuracy certification will be of most significance of accomplishing powerful and secure cloud storage frameworks.

The main aim of the paper is to users information stored in the cloud server using encrypted format and

generate JAR file with encrypted key. Eliminate duplicate entries of repeating data in the cloud server and has been reduce the amount of storage space and save bandwidth. Users are try to download the information form cloud server with the help of extract the JAR, verify the key and decrypt the content.

The objectives of the paper are as follows:

- Provides large security for user's information.
- Encrypted information files and images stored in the cloud server.
- Generate JAR file and key.
- Automatically generate logging mechanism.
- Decrypt the content and extract JAR.

II. EXISTING SYSTEM

Users now days are utilizing cloud services to store their large amount of information.

But cloud is third party provider so security related issues is the major problem here. Digital documents are stored in the cloud without apply any cryptography or generate one encrypted key. That encrypted key is stored in cloud. Using encrypted key users download the particular file. Data deduplication is the important data compression technique and eliminate duplicate entries in cloud server to increase the amount of storage space and bandwidth. To protect the privacy of aware information while sustaining deduplication,

Here using decentralized information accountability framework is produced to about the authenticate users access the information in the cloud, this is termed as Cloud Information Accountability (CIA) structure, and this is also accompanied with tracking mechanisms. The main concept is information accountability in the cloud computing. CIA structure provides end-to-end accountability in a strongly distributed manner. One of the fundamental innovative elements of the CIA structure lies in its capacity of keeping up lightweight and intense responsibility that consolidates parts of access control, utilization control and authentication. Decentralized traceable should be achieved, which means a user can retrieve the details of the file stored in the cloud. Information is updated based on user request and authorized users. Records should be easily retrievable from the cloud whenever needed.

Problems on existing system: Information taking care of outsourced by the direct CSP to different components in the cloud. Users are information stored in the cloud server in the form of plain text. Attackers are try to modify the users information because i.e plaintext. Create duplicate copies in the cloud server, acquired large amount of space. The critical challenge of cloud storage services is the management of the ever – increasing volume of data.

III. PROPOSED SYSTEM

The proposed CIA framework gives end – to - end responsibility in a very scattered design. One of the fundamental imaginative components of the CIA system lies in its capacity of keeping up lightweight and capable responsibility that consolidates parts of validation, access and usage control. Users information stored in the AWS cloud in encrypted format using MD5 algorithm. Authorized users information stored in the cloud, form of generate encrypted key and cipher text content using AES algorithm. Multimedia contents are stored in the cloud using long text data type, multimedia information also stored in the cloud in encrypted format. User's information eliminates the duplicate copies in the hybrid cloud.

Our main contributions are as follows:

Generates a automatic logging mechanisms in the cloud to track the usage. User information stored in cloud in the form of JAR files which is created by Java Virtual Machine, only authorized user can download the file. Authenticated users are extracting the key, JAR and decrypt the content then download the file. Logs should to be reliable and carefully designed to avoid illegal insertion, deletion, and change by malicious parties. Log records can access by both users and data owner. Eliminate the duplicate entries, save storage space and bandwidth.

CIA framework: The main features of Cloud Information Accountability (CIA) framework is the creates JAR files and user information stored in the cloud server also helps to tracking the data owner information and find the authorised users or not. The unauthorised users can't be able upload the information into the cloud server. Connected with the responsibility highlight, two unmistakable modes for examining is composed to be specific push and pull mode. The push mode sends the subtle elements of the client who got to the record and the pull mode refers to recover the logs and review when required. The CIA framework provides generate a automated logging and auditing mechanisms. The logger job is to automatically record the logging access of the file [1].

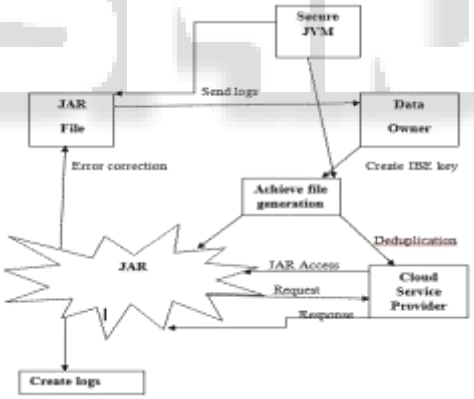

Fig. 1: Proposed system architecture

Java archive files (JARs): JAR is a package file format which aggregates resources files such as digital documents and multimedia contents into one file to distribute application software on the java program. JAR files are major file records, based on the ZIP file system and have the .JAR file extension. The components in a JAR record can be packed, which, together with the ability to download a whole application in a single request, makes downloading a JAR file substantially more advantageous than independently downloading the numerous uncompressed documents which would frame a solitary Java Application. To extract the contents of a JAR file clients can utilize any standard unzip programming, or the JAR command which accompanies each Java Virtual Machine: "jar-xf foo.jar".

The reason we use JARs, It is well known by servlet containers and java which helps major in standardization and ease of deployment also improves the performance of

application. If we open the JAR files we get encrypted Unicode character of the original file which is cumbersome to unauthorized persons to decrypt.

IV. IMPLEMENTATION

Distinct auditing modes: Push mode auditing technique is used to serve the data owner/user. By using push mode auditing technique the data owner/user can retrieve the information of the access logs. This mechanism is called push mode because the access log records are periodically update the records every time when the data is been accessed, so that the data owner/user can view the log records and gets the details of the status of his files stored in the cloud server. Pull mode auditing technique is used to serve the user/data owner, who uses the cloud resources or the files stored by the data owner in the cloud. By using pull mode auditing technique the user can retrieve the access log files on demand from the cloud server. This mechanism is called pull mode because the access log regards are retrieved by the user on demand. The type of log maintained is called access log in the CIA framework in order to ensure review of data usage by the users.

Access log: The access log is one of the best module in the cloud computing. Logs are security provides for the users information stored in the cloud server. These logs accessed for the both users and data owners whenever needed for retrieve the information from cloud server. An access log performs two operations; logging actions and enforcing access control. Log records in the access log are generated at the regular time when the file is been accessed or downloaded. The access log also contains some important fields like file name, Action, user name, Date and the ip address. The file name field indicates the name of the file being tried to access by the user. Action posses the type of action performed on the file either upload action or Download action. User name field is used to indicate the owner of the uploaded file in case of download indicate the user name. The date field indicates the day on which the action performed to breach the security. Last but most important field is the ip address field, this field is used to record/store the ip address of the hacker‘s system, by the help of which the hacker‘s location can be traced easily.

V. RESULTS AND ANALYSIS

The view of JAR files uploaded to the cloud server when the user or data owner uploads their raw data. The application encrypts the original files into JAR files as shown in the figure and stored in the cloud server. The cloud service provider or Hacker not able to view the raw data from the JAR since the data is been encrypted by the JVM i.e. 8byte Unicode character. These files can only be converted by the application using public key which was used before by the JVM while creating the JAR.

This enforces double protection to the files first need to explore JAR which is difficult without our application latter the decryption of the data is difficult without knowing the key and algorithm.

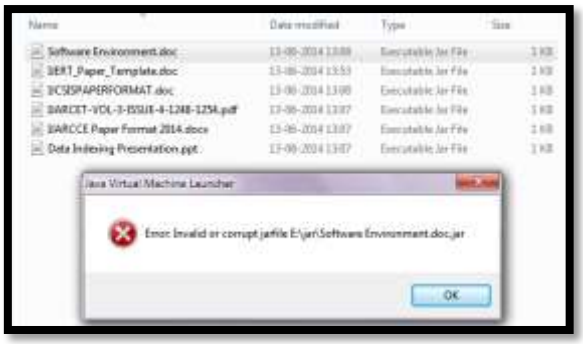

Fig. 2: The Error Message when Accessed JAR Files Stored in Cloud Server

VI. CONCLUSION & FUTURE SCOPE

In this paper, we analyze the problem of information security issues in the cloud data storage, which is essentially a distributed storage system. The primary objectives are securing the data put away in the cloud server furthermore to protect the cloud information from the unauthorized access. In this method provides total security for user’s information and automatically logging mechanisms. User’s information stored in cloud server using MD5 and advanced encryption standard algorithm also eliminates the duplicate entries in the cloud server.

The utilization of JARs in this attempt to store documents gives double assurance of the information records, These JARs takes less time to transfer and download once it is been gone under JVM, so utilizing this will expand proficiency and it have secured way. Users are try to download the information from the cloud server, extract the JAR also decrypt the information using advanced encryption standard algorithm.

In this work text and image files are used. But in future video files can also be used and tested with the same mechanism as a future work.

REFERENCES

[1] Cloud Computing, Principles and Paradigms by John Wiley & Sons.
[2] Kuyoro S. O., Ibikunle F. & Awodele O. “Cloud Computing Security Issues and Challenges”, International Journal of Computer Networks (IJCN), Volume (3): Issue (5)
[3] Ensuring Distributed Accountability for Data Sharing in the Cloud Author, Smitha Sundareswaran, Anna C.Squicciarini, Member, IEEE, and Dan Lin, IEEE Transactions on Dependable and Secure Computing ,VOL 9,NO,4 July/August 2012
[4] Te-Shun Chou, “Security threats on cloud computing vulnerabilities”, International Journal of Computer Science & Information Technology, Vol 5, No 3, June 2013
[5] S. Pearson , Y. Shen, and M. Mowbray," A privacy Manager for Cloud Computing," Proc. Int'l Conf. Cloud Computing (cloudcom), pp.90-106,2009.
[6] S. Pearson and A. Charlesworth, “Accountability as a Way Forward for Privacy Protection in the Cloud,” Proc. First Int‟l Conf. Cloud Computing, 2009.
[7] R. Corin, S. Etalle, J.I. den Hartog, G. Lenzini, and I. Staicu, “A Logic for Auditing Accountability in Decentralized Systems,” Proc. IFIP TC1 WG1.7

Workshop Formal Aspects in Security and Trust, pp. 187-201, 2005.

[8] B. Chun and A. C. Bavier ,"Decentralized Trust Management and Accountability in Federated System," Proc. Ann. Hawaii Int'l Conf. System Science (HICSS), 2004.

[9] P. Syam Kumar, R. Subramanian, "Homomorpic Distributed Verification Protocol for Ensuring Data Storage in Cloud Computing", Proc. First Int‟l Conf. Cloud Computing, 2009

[10] A. Juels and J. Burton S. Kaliski, "PORs: Proofs of Retrievability for the Files", ACM CCS Conference, 2012

