ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

A Framework for Integrity Auditing & Secure Data De-Duplication in Cloud

K.Swathi, Dr.N. Kasiviswanath

PG Student, Department of CSE, G.Pulla Reddy Engineering College (Autonomous), Kurnool, India

Professor & HOD, Department of CSE, G.Pulla Reddy Engineering College (Autonomous), Kurnool, India

ABSTRACT: Cloud storage systems are becoming more attractive trend for outsourcing data to cloud service for storage, which is beneficial in sparing efforts on heavy data maintenance and management. As cloud computing becomes prevalent, because enormous amount of data is being stored on the cloud. However, rapid growth of ever increasing volume of data has raised many new challenges. Data de-duplication is one of significant data compression technique for eliminating duplicate copies of repeating data and has been widely adopted in cloud storage to reduce the amount of storage space and save upload bandwidth. To preserve the confidentiality of susceptible data while supporting de-duplication, the Content hash keying an encryption mechanism has been proposed to encrypt the data before outsourcing it to cloud storage.

In this work, we review the issues of integrity auditing and secure data de-duplication. Specifically, we are focusing to achieve both data integrity and de-duplication in cloud, we come up with two secure systems, namely SecCloud and SecCloud+.SecCloud offers an auditing entity which maintains MapReduce cloud, assist clients to generate data tags before uploading as well as audit the integrity of data stored in the cloud. SecCloud+ is designed and motivated by the fact that clients always want to encrypt their data before uploading and authorize integrity auditing and secure de-duplication on encrypted data.

KEYWORDS: Cloud Storage, Data de-duplication, Integrity auditing.

I.INTRODUCTION

Cloud computing is computing within which giant clusters of remote server's square measure networked to permit centralized information storage and on-line access to pc resources. With cloud computing, giant pools of resources will be connected through non-public or public network. Cloud computing provides apparently unlimited "virtualized" resources to users as services across the entire net, at identical time activity platform and implementation details. Today's cloud servers provide each extraordinarily offered storage and significantly parallel computing resources at comparatively low prices. As cloud computing becomes common, as increasing quantity of information is being kept on the cloud and shared by users with mere privileges, which outline the access privileges of the information. One important challenges of cloud storage services are that the management of the ever-increasing volume of information.

Even though cloud storage system has been broadly adopted, it fails to accommodate some of the important emerging needs such as the ability to audit integrity of files by cloud clients and detecting duplicate files by cloud servers. We illustrate both problems below. Data de-duplication is an intelligent compression technique which eliminates redundant copies of data and maintains only a single copy has been implemented in cloud storage to reduce storage space and upload bandwidth. Capable as it is, an arising challenge is to perform secure de-duplication in cloud storage. Although content hash keying encryption method has been adopted for secure de-duplication, a critical issue of making content hash keying encryption practical is, to efficiently and reliably manage a huge number of convergent keys. One critical confront of today's cloud storage services is the management of the ever increasing volume of data. To make data management scalable we use de-duplication and content hash keying encryption for secure de-duplication services.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

Two types of de-duplication in terms of the volume: (a) block-level de-duplication, which find out and eliminate redundancies among data blocks.(b)file-level de-duplication, which determine redundancies between different files and eradicate these redundancies to decrease ability demands and The file can be separated into lesser fixed-size.

II.RELATED WORK

A. ENABLING PUBLIC VERIFIABILITY AND DATA DYNAMICS FOR STORAGE SECURITY IN CLOUD COMPUTING

Author: Qian Wang

Cloud Computing has been envision as the next generation design of IT Enterprise. It moves the application software and databases consolidated with large data centers, where the data is managed and services may not be fully trustworthy which brings many new challenges. Our research work examines the problem of assuring the integrity of data storage in Cloud Computing. In appropriate, we consider the task of allowing a third party auditor (TPA), on concern of the cloud client, to verify the integrity of the dynamic data on behalf of clients stored in the cloud. The introduction of TPA dismisses the contribution of client through the auditing of whether user's data stored in the cloud is truly unharmed. The support for data dynamics through the most general forms of data operation, such as block modification, insertion and deletion, is also more powerful step toward practicality, since services in Cloud Computing are not limited to archive or backup data only. While presiding work on ensure remote data integrity often needs the supports of either public verifiability or dynamic data operation. Proofs of Ownership in Remote Storage Systems.

B. PROOFS OF OWNERSHIP IN REMOTE STORAGE SYSTEMS

Author: Shai Halevi

Cloud storage systems are becoming more and trendier. A promising technology that makes cost down is de-duplication, which stores only a single copy of duplicated data. Client-side de-duplication attempts to identify de-duplication opportunities already at the client side and save the bandwidth of uploading copies of existing files to the server. In this work we discover attacks that exploit client-side deduplication, granting an attacker to gain access to arbitrary-size files of other users based on a hash signature of these files. More explicitly, an attacker who knows the hash signature of a file can assure the storage service that it owns that same file, hence the server lets the attacker download the entire file.

C. DupLESS: SERVER-AIDED ENCRYPTION FOR DE-DUPLICATED STORAGE

Author: Mihir Bellare

Cloud storage service providers such as Dropbox, Mozy, and others perform de-duplication to save space by only storing one copy of each file uploaded. Should clients frequently encrypt their files, convergent encryption resolves this tension. However it is inherently vulnerable to brute -force attacks that can recover files falling into a known set. We propose an architecture that provides secure de-duplicated storage opposing brute-force attacks, and realize it in a system called DupLESS. In DupLESS, clients encrypt the under message-based keys obtained from a key-server via an oblivious PRF protocol. It enables clients to store encrypted data with a current service, the service perform de-duplication on their behalf and yet achieves strong confidentiality. We show that encryption for de-duplicated storage can achieve performance and space savings near to that of using the storage service with plaintext data.

D. PROVABLE DATA POSSESSION AT UNTRUSTED STORES

Authors: Giuseppe Ateniese

We introduce a model for obvious knowledge possession (PDP) that permits a consumer World Health Organization stores the information at associate degree untrusted server to verify that the server possesses the first data while not retrieving it. The model generates probabilistic proofs of possession by sampling random sets of blocks from the server, which considerably reduces I/O prices. The consumer maintains a relentless quantity of information to verify the proof. The challenge/response protocol transmits a tiny low, constant quantity of information that minimizes network

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

communication. Thus, the PDP model for remote knowledge checking supports massive knowledge sets in wide - distributed storage systems.

III.PROPOSED MODEL

The distributed de-duplication systems upcoming aim is to truly store information within the cloud whereas achieving privacy and Consistency. Its main objective is to permit de-duplication and distributed storage of the information indirectly multiple storage servers. As an alternate encrypting the information to stay the privacy of the information, new structures placed on the top- secret intense technique is to separate data into shards. These shards can then be distributed transversally in multiple storage servers.

In our previous data de-duplication systems, the non-public cloud is bothered as a proxy to permit data owner/users to firmly perform duplicate analyze differential privileges. Such vogue is smart and has attracted attention from researchers. Information owners exclusively supply their information storage by utilizing public cloud while the data operation is managed in private cloud. Data de-duplication is one of necessary knowledge compression technique for eliminating duplicate copies of replicated data and has been widely employed in cloud storage to reduce the number of duplicate copies of files and also saves the bandwidth. To safeguard the confidentiality of sensitive data whereas supporting de-duplication, Cloud computing provides supposedly unlimited ,virtualized resources to users as services across the whole internet, hiding the platform and implementation details. Today's cloud service suppliers provide each very offered storage and massively parallel computing resources at relatively low prices.

INTEGRITY AUDITING:

Integrity auditing refers to take care of and assure the accuracy and consistency of information over its entire life-cycle and could be a vital side to the planning, implementation and usage of any system that stores, processes/retrieve information. The definition of obvious information possession (PDP) was introduced by Ateniese et al for reassuring that the cloud servers exactly possess the target files while not retrieving or downloading the entire information. Essentially, PDP could be a probabilistic proof protocol by sampling a random set of blocks and asking the servers to prove that they precisely possess these blocks and the verifier only maintaining a small amount of metadata is able to perform the integrity checking.

Another line of labor supporting integrity auditing is proof of retrievability (POR). Compared with PDP, POR not simply assures the cloud servers possess the target files, however conjointly guarantees their full recovery. the primary style goal of this work is to produce the potential of confirmative correctness of the remotely hold on information. The integrity verification additionally needs 2 features: 1) public verification, that permits anyone, not simply the clients originally hold on the file, to perform verification: 2) stateless verification, which is ready to eliminate the requirement for state information maintenance at the verifier side between the actions of auditing and information storage.

SECURE DEDUPLICATION:

The second design goal of this work is secure de-duplication. In alternative words, it needs that the cloud server is ready to scale back the space for storing by keeping only 1 copy of identical file. Notice that, concerning to secure de-duplication, our objective is distinguished from previous work that we have a tendency to propose a way for permitting each de-duplications over files and tags. Data de-duplication could also be a specialized data compression technique for eliminating duplicate copies of repetition data. Connected and somewhat synonymous terms unit of measurement intelligent data compression and single-instance data storage. This technique is used to spice up storage utilization and would possibly even be applied to network data transfers to chop back the number of bytes that must be sent. Among the de-duplication technique, distinctive chunks of data, or store unit patterns, unit of measurement illustrious and hold on throughout a way of study. As a result of the analysis continues various chunks sq. measure compared to the hold on copy and whenever a match happens, the redundant chunk is replaced with a little low reference that points to the hold on chunk.

The file level and block level de-duplication is used for higher reliability. The secret splitting technique is used to protect the data. Our proposed structures support both traditional de-duplication methods. Privacy, credibility

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

and integrity can be achieved in our proposed system. In solution to kind of secret agreement attacks are considered. These are the attack on the data and the attack against servers. The data is secure when the opponent control limited number of storage servers.

IV.SYSTEM MODEL

The architecture involves three main entities

- ✓ Cloud service provider
- ✓ Data users and
- ✓ Auditor

Fig 1 .System architecture

Cloud Service Provider

- ✓ We introduce a Cloud Service Provider module. This is an cloud storage server that provides a data storage services in public cloud.
- ✓ The CS provides the data outsourcing service and stores data on behalf of the users.
- ✓ To reduce the storage cost, the CS eliminates the storage of redundant data via de-duplication and keeps only the unique data.
- ✓ Here, we assume that CS is always online and has abundant storage capacity and computation power.

Data Users Module

- ✓ A user is an entity that desires to source information storage to the S-CSP and access the information later.
- ✓ In a storage system supporting de-duplication, the user solely transfers distinctive information however doesn't transfer any duplicate information so as to avoid wasting the upload information measure, which can be owned by a similar user or completely different users.
- ✓ In the approved de-duplication system, every user is issued a group of privileges within the setup of the system. Every file is protected with the convergent cryptography key and privilege keys to appreciate the approved de-duplication with differential privileges.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

Auditor

Auditor who helps clients uploads and audits their outsourced data, which maintains a MapReduce cloud and acts like a certificate authority. This guess presumes that the auditor is associated with a pair of public and private keys. Its public key is made available to the other individuals in the system. The first design goal of this work is to provide the capability of verifying correctness of the remotely stored data. Public verification, which allows anyone, not just the clients originally stored the file, to perform verification.

Functionalities of User and Auditor

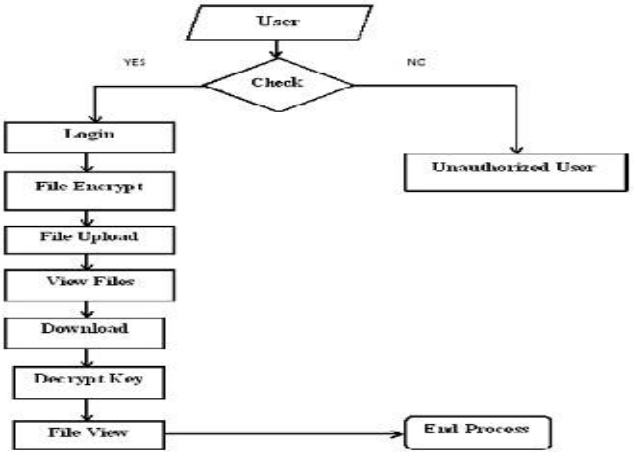


Fig 2: User functionalities

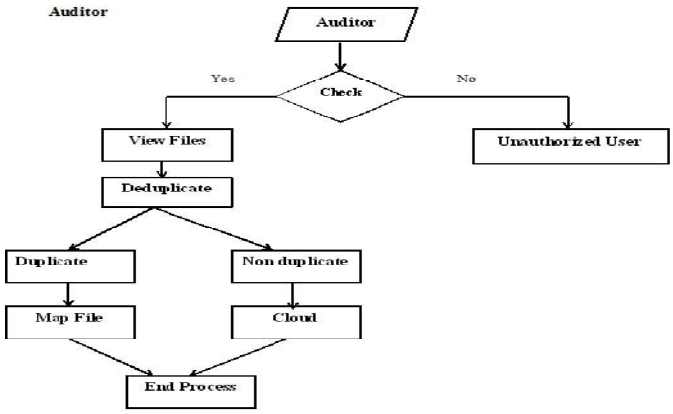
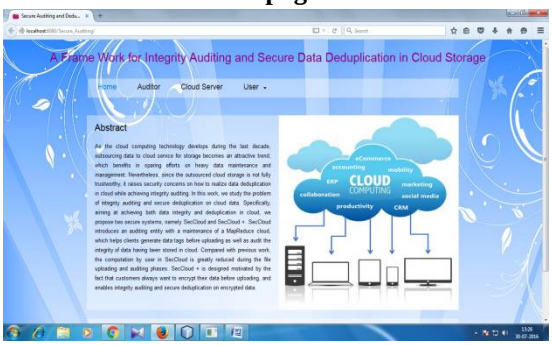


Fig 3: Auditor functionalities

V.RESULTS

1. Home page



Description: This is the Homepage consists of Three modules Auditor, Cloud server and user.

2. User registration page

Description: First the user has to get registered to become an authorized person.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

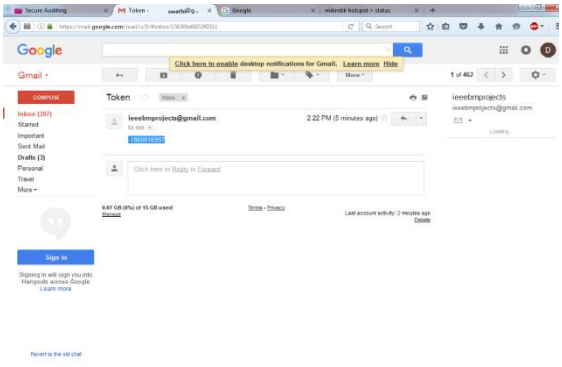
3. Cloud Home page

Description: Cloud server only has the right to Activate or de-activate the account of the user.

4. User Login

Description: Now after activating the account, the User can login.

5. Token generation

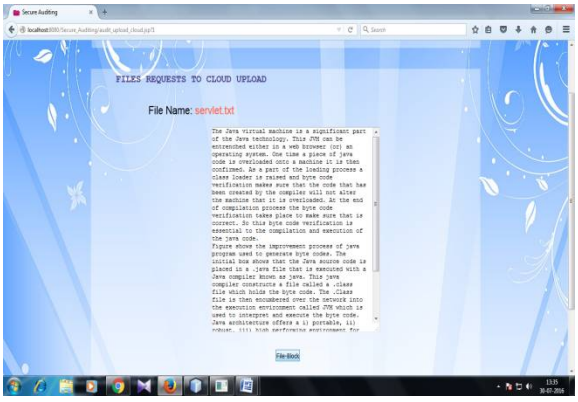


Description: Authorized user receives a token and that token has to be entered.

6. File Upload

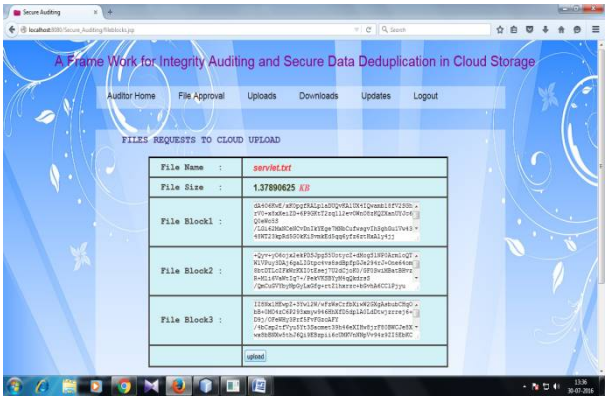
Description: Now the user browses a file and Upload it.

7. File Upload request



Description: Now Auditor checks for duplication and if it is unique he uploads it to cloud server.

8. Conversion of file to block



Description: File is divided into blocks and hashes are generated to that blocks.

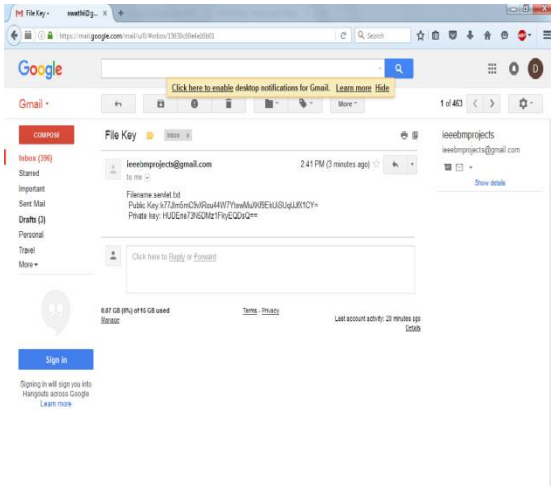
ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

9. File Keys



Description: A pair of keys (public and private keys) are sent to the user to perform modification.

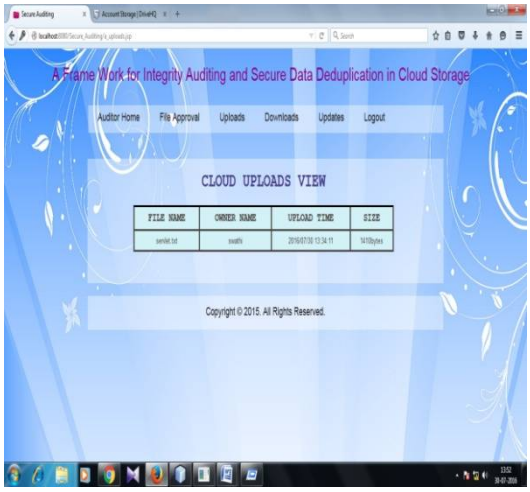
10. File Download

Description: These are list of files available for Downloads**.**

11. File For Download

Description: User has to provide pair of keys.

12. Cloud Upload view:



Description: This is view after uploading files.

VI.CONCLUSION

Aim to achieve both data integrity and data de-duplication in cloud storage; we proposed two secure clouds SecCloud and SecCloud+. SecCloud introduces an auditing entity which maintains a MapReduce cloud helps clients to generate data tags before uploading at the same time audits the integrity of data being stored in the cloud. In addition, SecCloud enables secure de-duplication through introducing a Proof of Proprietorship protocol by preventing the leakage of side channel information in data de-duplication. Compared with earlier work, the computation by user in SecCloud is significantly reduced during the file uploading and auditing phases. SecCloud+ is an advanced construction motivated by the fact that clients always want to encrypt their data before uploading and allows for integrity auditing and secure de-duplication directly on encrypted data.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

REFERENCES

[1] Jingwei Li, Jin Li, Dongqing Xie and Zhang Cai, "Secure Auditing and Deduplicating Data in Cloud", IEEE Transactions on Computers 2015.
[2]Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, ,A Hybrid Cloud Approach for Secure Authorized Deduplication', IEEE Transactions on Parallel and Cloud Systems,2014.
[3]H. Wang, "Proxy provable data possession in public clouds," IEEE Transactions on Services Computing, vol. 6, no. 4, pp. 551–559, 2013.
[4]M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: "Server aided encryption for deduplicated storage". In USENIX Security Symposium, 2013.
[5]Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, "Cooperative provable data possession for integrity verification in multicloud storage," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 12, pp. 2231–2244, 2012.
[6]S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of Ownership in Remote Storage Systems", in Proc. ACM Conf. Comput. Commun. Security, Y. Chen, G. Danezis, and V. Shmatikov, Eds., 2011.
[7]G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner,Z. Peterson, and D. Song, "Remote data checking using provable data possession," ACM Trans. Inf. Syst. Secur., vol. 14, no. 1, pp. 12:1–12:34,2011.
[8]M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski,G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Communication of the ACM, vol. 53, no. 4, pp.50–58, 2010.
[9]C. Erway, A. K¨upc¸ ¨u, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in Proceedings of the 16th ACM Conference on Computer and Communications Security, ser. CCS '09. New York, NY, USA: ACM, 2009.
[10]F. Seb´e, J.Domingo-Ferrer, A.Martinez-Balleste, Y. Deswarte, and J.-J.Quisquater, "Efficient remote data possession checking in critical information infrastructures," IEEE Trans. on Knowl and Data Eng,vol.20, no. 8, pp. 1034–1038, 2008.
[11]H. Wang, "Proxy provable data possession in public clouds," IEEE Transactions on Services Computing, vol. 6, no. 4, pp. 551–559, 2013.
[12]Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, "Cooperative provable data possession for integrity verification in multicloud storage," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 12, pp. 2231–2244, 2012.