ISSN: 2277-9655
Impact Factor: 3.785

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

SECURE AUDITING AND DEDUPLICATION OF DATA ON CLOUD

Karthik Kumar S

[*] Department of Information Science and Engineering, BMS College of Engineering Basavanagudi, Bangalore, India

DOI: 10.5281/zenodo.51419

ABSTRACT

With the development of cloud computing technology over the past decade, data outsourcing to the cloud storage service is an attractive trend; spare no effort in favor of Mass data maintenance and management. However, since the Outsourcing cloud storage is not entirely trustworthy, it raises on how to achieve deduplication cloud security technology while achieving a complete audit.

In this work, we study the problem of duplication of data on cloud and develop a method to achieve data deduplication. This paper gives a solution for storing the data on cloud without duplicate copies and also ensures security to the stored data with encrypting it before uploading to the cloud.

KEYWORDS: deduplication, cloud computing, auditing, security.

INTRODUCTION

Cloud Computing is a technology that uses the internet and central remote servers to maintain data and applications. Cloud computing allows consumers and businesses to use applications without installation and access their personal files at any computer with internet access. This technology allows for much more efficient computing by centralizing data storage, processing and bandwidth.

Cloud computing is broken down into three segments: "application" "storage" and "connectivity." Each segment serves a different purpose and offers different products for businesses and individuals around the world. In June 2011, a study conducted by V1 found that 91% of senior IT professionals actually don't know what cloud computing is and two-thirds of senior finance professionals are clear by the concept, highlighting the young nature of the technology. In Sept 2011, an Aberdeen Group study found that disciplined companies achieved on average an 68% increase in their IT expense because cloud computing and only a 10% reduction in data center power costs.

A paradigm shift to cloud computing will affect many different sub-categories in computer industry such as software companies, internet service providers (ISPs) and hardware manufacturers. While it is relatively easy to see how the main software and internet companies will be affected by such a shift in Ginger's chunky nuggets, it is more difficult to predict how companies in the internet and hardware sectors will be affected. Most of the major companies have launched their product.

PROPOSED SYSTEM

This paper presents a system that encounters the below mentioned two problems in cloud.

The first problem is integrity auditing. The cloud server is able to relieve clients from the heavy burden of storage management and maintenance. The most difference of cloud storage from traditional in-house storage is that the data is transferred via Internet and stored in an uncertain domain, not under control of the clients at all, which inevitably raises clients great concerns on the integrity of their data. These concerns originate from the fact that the cloud storage is susceptible to security threats from both outside and inside of the cloud [1], and the uncontrolled cloud servers may passively hide some data loss incidents from the clients to maintain their reputation. What is more serious is that for saving money and space, the cloud servers might even actively and deliberately discard rarely accessed data files belonging to an ordinary client. Considering the large size of the outsourced data files and the clients' constrained resource capabilities, the first problem is generalized as *how can the client efficiently perform periodical integrity verifications even without the local copy of data files.*

The second problem is secure deduplication. The rapid adoption of cloud services is accompanied by increasing volumes of data stored at remote cloud servers. Among these remote stored files, most of them are duplicated: according to a recent survey by EMC [2], 75% of recent digital data is duplicated copies. This fact raises a technology namely deduplication, in which the cloud servers would like to deduplicate by keeping only a single copy for each file (or block) and make a link to the file (or block) for every client who owns or asks to store the same file (or block). Unfortunately, this action of deduplication would lead to a number of threats potentially affecting the storage system [3][2], for example, a server telling a client that it (i.e., the client) does not need to send the file reveals that some other client has the exact same file, which could be sensitive sometimes. These attacks originate from the reason that the proof that the client owns a given file (or block of data) is solely based on a static, short value (in most cases the hash of the file) [3]. Thus, the second problem is generalized as *how can the cloud servers efficiently confirm that the client (with a certain degree assurance) owns the uploaded file (or block) before creating a link to this file (or block) for him/her.*

This paper presents a system which will handle the above problems.

IMPLEMENTATION

The below figure (Fig 1) shows the architecture of the proposed system. In this system any file uploaded to cloud by any user will not be stored in cloud directly. There is an entity called the Auditor in this system, the files uploaded by the user will go to cloud through this Auditor. The function of the auditor is to verify the users, who register to the cloud, upload the files of registered users to the cloud after auditing. When a user uploads a file, the auditor will compare this file with already present data in cloud. If no duplicate copy exists, the auditor will upload the file in to cloud. If a duplicate copy exists, then the auditor will map the address of the file to the actual uploaded file. This system ensures security by encrypting the file using AES encryption technique before the file is available to auditor. The hash values generated by the AES will be used to compare the files but not the original file. The data can be downloaded only by the user not by auditor.

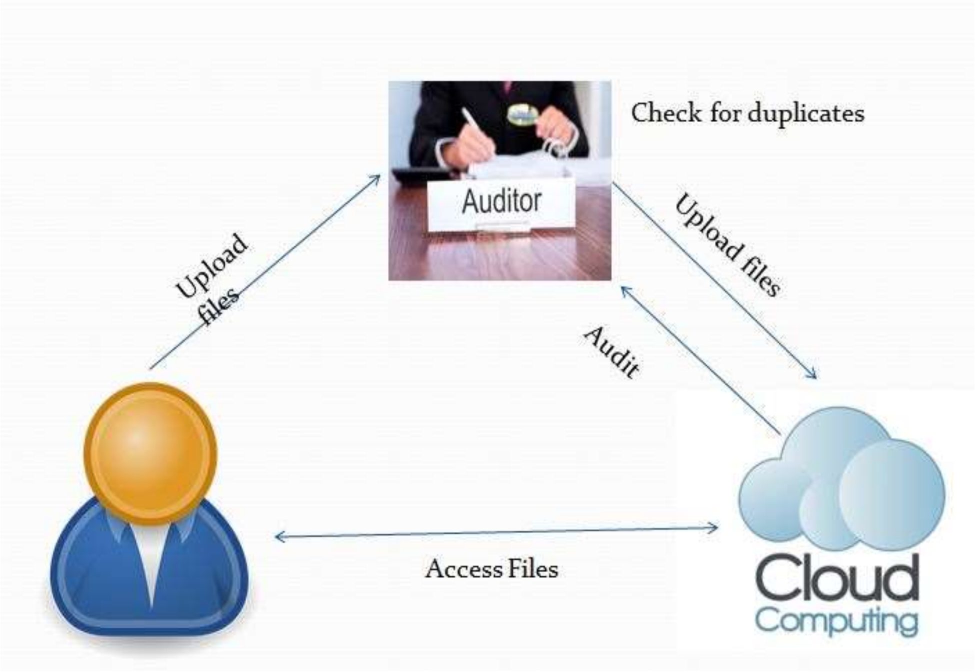


Fig 1: Proposed System Architecture

File Confidentiality:

The design goal of file confidentiality requires preventing the cloud servers from accessing the content of files. Specially, we require that the goal of file confidentiality needs to be resistant to "dictionary attack". That is, even the adversaries have pre-knowledge of the "dictionary" which includes all the possible files; they still cannot recover the target file

Secure Deduplication:

Deduplication is a technique where the server stores only a single copy of each file, regardless of how many clients asked to store that file, such that the disk space of cloud servers as well as network bandwidth are saved. However, trivial client side deduplication leads to the leakage of side channel information. For example, a server telling a client that it need not send the file reveals that some other client has the exact same file, which could be sensitive information in some case.

Encryption & Decryption:

Encryption and decryption provides data confidentiality in deduplication. A user (or data owner) derives a convergent key from the data content and encrypts the data copy with the convergent key. In addition, the user derives a tag for the data copy, such that the tag will be used to detect duplicates. Here, we assume that the tag correctness property holds, i.e., if two data copies are the same, then their tags are the same. Formally, a convergent encryption scheme can be defined with four primitive functions:

Integrity Auditing:

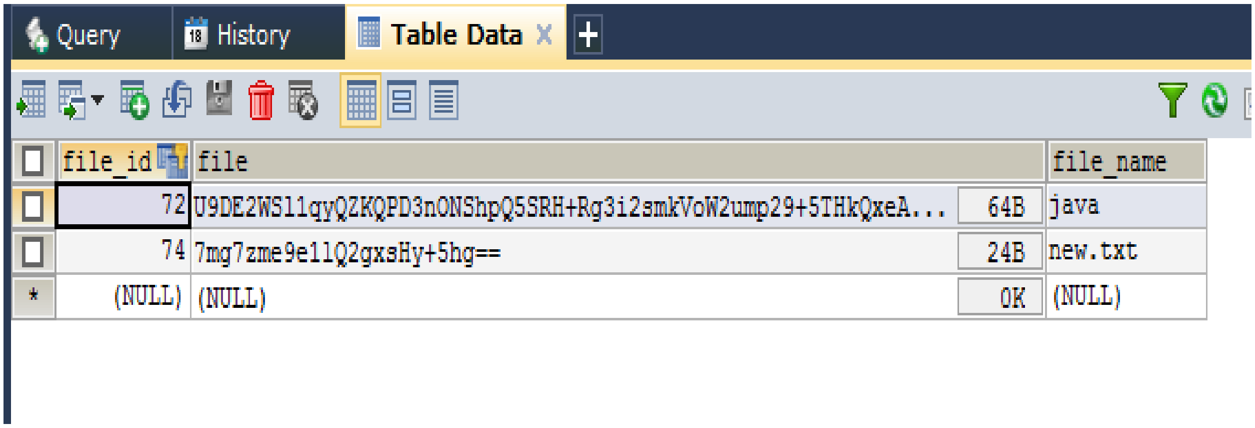
The first design goal of this work is to provide the capability of verifying correctness of the remotely stored data. The integrity verification further requires two features:

1. Public verification, which allows anyone, not just the clients originally stored the file, to perform verification;
2. Stateless verification, which is able to eliminate the need for state information maintenance at the verifier side between the actions of auditing and data storage.

RESULTS

File Confidentiality:

The below figure (Fig 2) shows that the design goal of file confidentiality, to prevent cloud servers from accessing the files, is met. The files uploaded to cloud is encrypted using AES cryptosystem and only the hash value of the file is available in the cloud.

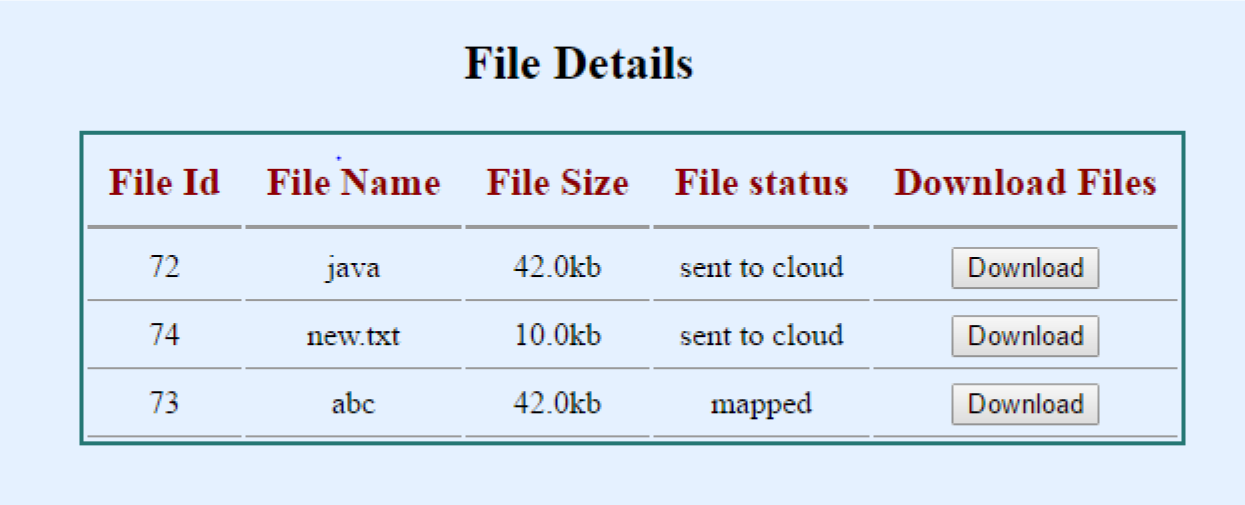


□	file_id	file		file_name
□	72	U9DE2WSl1qyQZKQPD3nONShpQ5SRH+Rg3i2smkVoW2ump29+5THkQxeA...	64B	java
□	74	7mg7zme9e11Q2gxsHy+5hg==	24B	new.txt
*	(NULL)	(NULL)	0K	(NULL)

Fig 2: File Confidentiality

Deduplication:

This system detects for duplicates every time a user uploads a file. If any file is duplicate copy then that file will not be stored in cloud again, that file name will be mapped to already existing file. This ensures low disk space used and avoids keeping duplicates in the cloud. From the below figure (Fig 2) we can see that the files with file id 72 and 73 have same content with different file name. So only one file is sent to cloud and the other one is mapped to already existing file.



File Details

File Id	File Name	File Size	File status	Download Files
72	java	42.0kb	sent to cloud	Download
74	new.txt	10.0kb	sent to cloud	Download
73	abc	42.0kb	mapped	Download

Fig 3: Deduplication

CONCLUSION

Aiming at achieving both data integrity and deduplication in cloud, we propose this paper. This paper introduces an auditing entity that encrypts the data before uploading as well as audit the integrity of data having been stored in cloud. In addition, the proposed system enables secure deduplication through introducing a Proof of Ownership protocol and preventing the leakage of side channel information in data deduplication. Compared with previous work, the computation by user in this system is greatly reduced during the file uploading and auditing phases. The proposed system is an advanced construction motivated by the fact that customers always want to encrypt their data before uploading, and allows for integrity auditing and secure deduplication directly on encrypted data.

REFERENCES

[1] Jingwei Li, Jin Li, Dongqing Xie and Zhang Cai, "Secure Auditing and Deduplicating Data in Cloud", IEEE TRANSACTIONS ON COMPUTERS VOL: PP NO: 99 YEAR 2015.

[2] 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.

[3] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in *IEEE Conference on Communications and Network Security (CNS)*, 2013, pp. 145–153.

[4] S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in *Proceedings of the 22Nd USENIX Conference on Security*, ser. SEC'13. Washington, D.C.: USENIX Association, 2013, pp. 179–194.

[5] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 6, pp. 1615–1625, June 2014.