

Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage

Dr.S.Masood Ahamed¹, N.Mounika², N.vasavi³, M.Vinitha Reddy⁴

HOD, Department of Computer Science & Engineering,, Guru Nanak Institutions, Ibrahimpatnam, Hyderabad, India¹

B.Tech Student, Department of Computer Science & Engineering, Guru Nanak Institutions, Hyderabad, India²

B.Tech Student, Department of Computer Science & Engineering, Guru Nanak Institutions, Hyderabad, India³

B.Tech Student, Department of Computer Science & Engineering, Guru Nanak Institutions, Hyderabad, India⁴

ABSTRACT: In cloud storage services, deduplication technology was introduced to reduce the space and bandwidth requirements of services by eliminating redundant data and storing only a single copy of them. Deduplication is most effective when multiple users upload the same data to the cloud storage, but it raises issues relating to security and ownership. Proof-of-ownership schemes allow any owner of the same data to prove to the cloud storage server that he owns the data in a different way. Recently, several deduplication schemes are proposed to solve this problem by allowing each owner to share the same encryption key for the same data. However, most of the schemes suffer from security issues, since they do not consider the dynamic changes in the ownership of the data that occur frequently in a practical cloud storage service. In this paper, we propose a server-side deduplication scheme for encrypted data. It allows the cloud server to control access to outsourced data even when the ownership changes dynamically by providing convergent encryption and secure ownership group key distribution. This prevents data leakage not only to revoked users even though they previously owned that data, but also to any honest curious cloud storage server. In addition, the proposed scheme guarantees data integrity against any tag inconsistency attack. Thus, security is enhanced in the proposed scheme. The efficiency analysis results demonstrate that the proposed scheme is almost as efficient as the previous schemes, while the additional computational overhead is negligible.

KEYWORDS: Deduplication, cloud storage, encryption, proof-of-ownership

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, deduplication has been a well-known technique and has attracted more and more attention recently.

Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage.the deduplication a hashing function can be used to return a unique key for a particular data [1].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, deduplication eliminates redundant data by keeping only. One physical copy and referring other redundant data to that copy. Deduplication can take place at either the file level or the block level. For file-level deduplication, it eliminates duplicate copies of the same file.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

Deduplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files [1].

Providers of cloud-based storage such as Google drive can save on storage costs via deduplication: should two clients upload the same file, the service detects this and stores only a single copy [2]. Convergent encryption [2][9] has been proposed to enforce data confidentiality while making deduplication feasible. It encrypts/decrypts a data copy with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After key generation and data encryption, users retain the keys and send the cipher text to the cloud.

II.RELATED WORK

Bellare et al. [3] showed Data confidentiality by transforming the predictable message into unpredictable message. Introduces a key server as third party to generate the file tag for duplicate check. Bugiel et al. [6] provided an architecture consisting of twin clouds for secure outsourcing of data and arbitrary computations to an untrusted commodity cloud. Xu et al. [9] also addressed the problem and showed a secure convergent encryption for efficient encryption, without considering issues of the key-management and block-level deduplication.

Anderson,Le Zhang. [1] proposed backup solutions for fast and secure backups,used an encrypted deduplication algorithm. This algorithm supports client-end per-user encryption which is necessary for confidential personal data. It also supports a unique feature which allows immediate detection of common sub trees, avoiding the need to query the backup system for every file.

III.EXISTING SYSTEM

In the existing deduplication system, each user is issued a set of privileges during system initialization. Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files. Before submitting his duplicate check request for a file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this if and only if there is a copy of this file and a matched privilege stored in cloud.

EXISTING TECHNIQUE:

- Symmetric encryption technique.

TECHNIQUE DEFINITION:

- Symmetric encryption uses a common secret key to encrypt and decrypt information.

DRAWBACKS:

- The user needs to know private key.
- Less protect security.
- These deduplication systems cannot support differential authorization duplication check.

IV.PROPOSED SYSTEM

In the proposing system, we .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 privacy of sensitive data while supporting deduplication, 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 deduplication.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

PROPOSED TECHNIQUE:

- Convergent encryption technique.

TECHNIQUE DEFINITION:

- A user derives a convergent key from each original data copy and encrypts the data copy with the convergent key.

ADVANTAGES:

- The user don't needs to know private key
- Better protect security.
- These deduplication systems can support differential authorization duplication check.

V. SYSTEM ARCHITECTURE

Whenever someone wants to give information or take information from cloud they have to take permission i.e. authentication is done. If not a member they have to register first. Then the user will request private cloud to get a file token .Private cloud will issue file token and Convergent key generation takes place. With that key user will upload a file to the public cloud, then there will be a deduplication system to check whether the file already exist or not. If the file already exists then the file will not upload in public cloud.

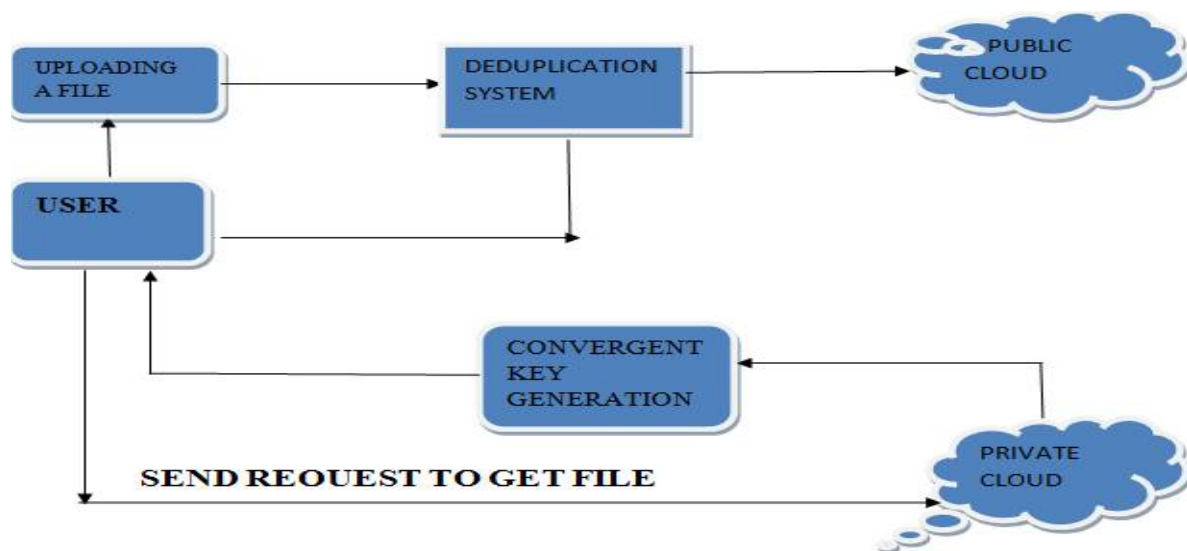


Fig1: System architecture

AUTHENTICATION:

The process of identifying an individual usually based on a username and password. In security systems, Authentication merely ensures that the individual is who he or she claims to be, but says nothing about the access rights of the individual. In authentication module is used to security purpose. Here this module only for user, after registration user enter the username and password. This input is check into the database, whether input is correct or not. If input is correct then allow to next process otherwise consider as a non authenticated user.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

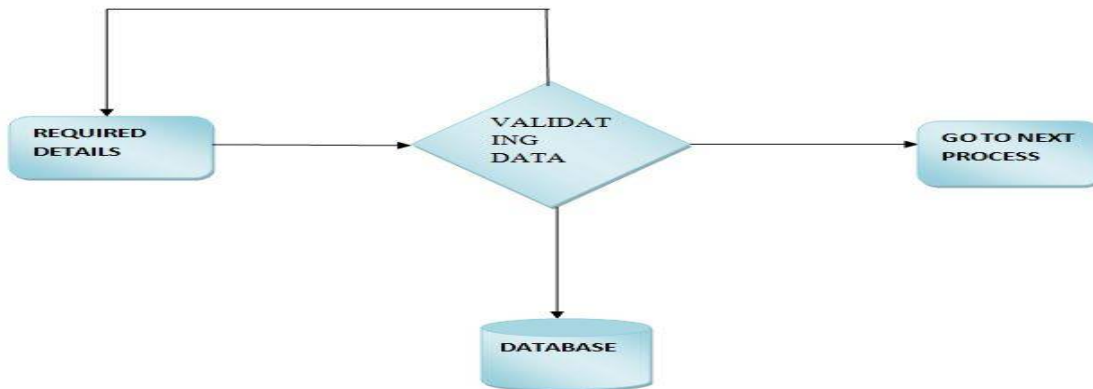
Vol. 6, Issue 4, April 2017



a) Authentication

REGISTER:

In this Module If he is a new user he needs to enter the required data to register the form and the data will be stored in server for future authentication purpose.



b) Register

CONVERGENT KEY GENERATION

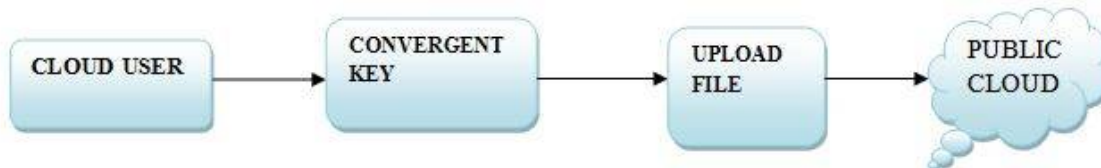
In this module, if user wants to upload a file user needs to get key from private cloud.



C) Convergent Key Generation

FILE UPLOADING

User can upload a file into the private cloud by using convergent key.



d) File Uploading

International Journal of Innovative Research in Science, Engineering and Technology

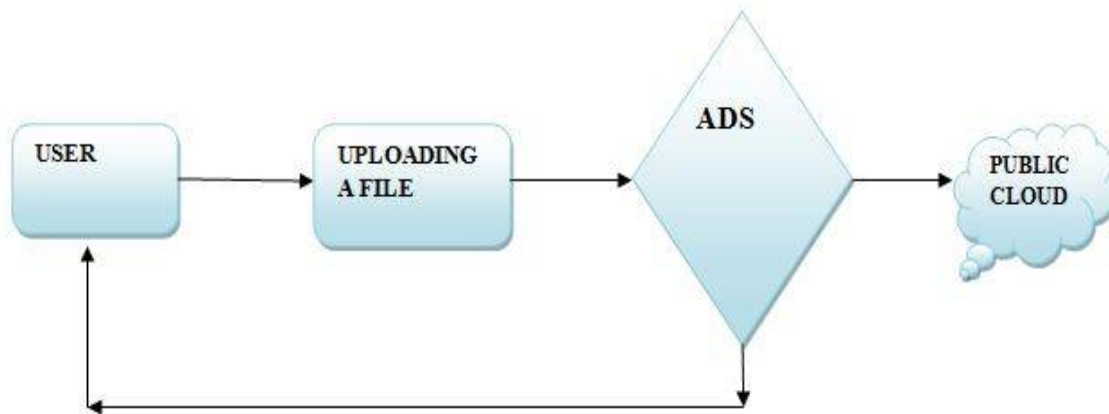
(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

AUTHORIZED DUPLICATE CHECK SCHEME (ADS)

The public cloud performs duplicate check directly and tells the user if there is any duplicate. Public Cloud can store and retrieve file. De-duplication has a removing duplicate file. Its will find out duplicate file.



e) Authorized Duplicate Check Scheme

VI. ALGORITHM USED

CONVERGENT ENCRYPTION TECHNIQUE

A user derives a convergent key from each original data copy and encrypts the data copy with the convergent key. The key generation algorithm that maps a data copy to a convergent key. The symmetric encryption algorithm that takes both the convergent key and the data copy as inputs and then outputs a cipher text. The decryption algorithm that takes both the cipher text and the convergent key as inputs and then outputs the original data copy and the tag generation algorithm that maps the original data copy and outputs a tag.

VII. APPLICATIONS

CtrlS Real Cloud:

The CtrlS Real Cloud has a multi-layered management model. The cloud controller server enables everything, from system architecture to VM root access, to be managed via the user interface and API. Real Cloud enables you to put up applications and manage them, all remotely and with utmost ease.

Cloud Layer Services:

Discover the promise of cloud, not the compromises. Cloud Layer includes virtual servers, remote storage and a robust content delivery network that leverage our core advantages and longtime leadership in automated, on-demand, self-managed infrastructure.

International Journal of Innovative Research in Science, Engineering and Technology

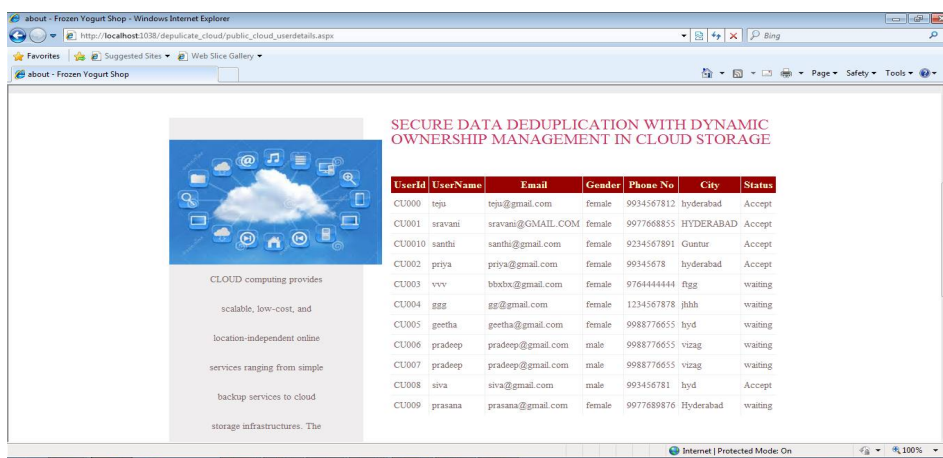
(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

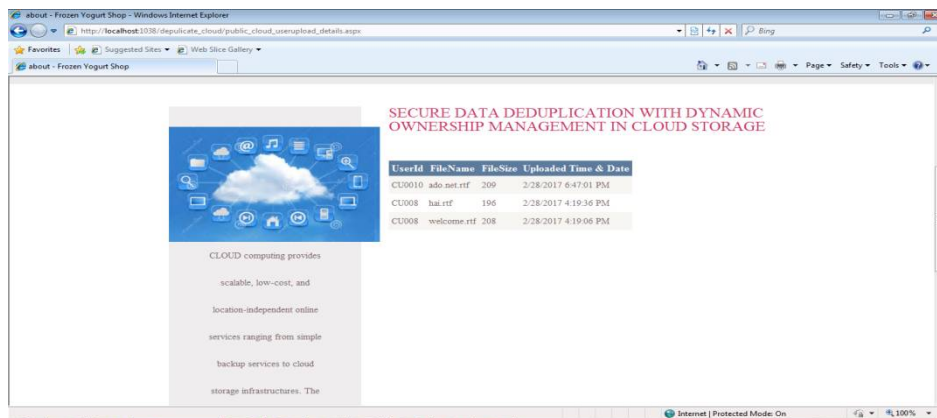
VIII. RESULT

Where the public cloud view user details which gives who viewed the uploaded file and public cloud view user upload details give details of users who uploaded files into public cloud.



UserId	UserName	Email	Gender	Phone No	City	Status
CU000	toja	toja@gmail.com	female	9934567812	hyderabad	Accept
CU001	sravani	sravani@GMAIL.COM	female	9977668855	HYDERABAD	Accept
CU010	santhi	santhi@gmail.com	female	9234567891	Guntur	Accept
CU002	priya	priya@gmail.com	female	99345678	hyderabad	Accept
CU003	vvv	bbxbx@gmail.com	female	9764444444	ftgg	waiting
CU004	ggg	gg@gmail.com	female	1234567878	jhhh	waiting
CU005	geetha	geetha@gmail.com	female	9988776655	hyd	waiting
CU006	pradeep	pradeep@gmail.com	male	9988776655	vizag	waiting
CU007	pradeep	pradeep@gmail.com	male	9988776655	vizag	waiting
CU008	siva	siva@gmail.com	male	993456781	hyd	Accept
CU009	prasana	prasana@gmail.com	female	9977689876	Hyderabad	waiting

Fig 2: public cloud view user details



UserId	FileName	FileSize	Uploaded Time & Date
CU0010	ado.net.rtf	209	2/28/2017 6:47:01 PM
CU008	hai.rtf	196	2/28/2017 4:19:36 PM
CU008	welcome.rtf	208	2/28/2017 4:19:08 PM

Fig 3: public cloud view user upload details

IX. CONCLUSION

In this paper, the notion of authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check. In which the duplicate-check tokens of files are generated by the private cloud server with private keys.

REFERENCES

- [1] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted de-duplication," in Proc. 24th Int. Conf. Large Installation Syst. Admin., 2010, pp. 29–40.
- [2] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Sec. Symp., 2013, pp. 179–194.



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

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

- [3] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Proc. 32nd Annu. Int. Conf. Theory Appl. Cryptographic Techn., 2013, pp. 296–312.
- [4] M. Bellare, C. Namprempe, and G. Neven, "Security proofs for identity-based identification and signature schemes," J. Cryptol., vol. 22, no. 1, pp. 1–61, 2009.
- [5] M. Bellare and A. Palacio, "Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks," in Proc. 22nd Annu. Int. Cryptol. Conf. Adv. Cryptol., 2002, pp. 162–177.
- [6] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider, "Twin clouds: An architecture for secure cloud computing," in Proc. Workshop Cryptography Security Clouds, 2011, pp. 32–44.
- [7] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system," in Proc. Int. Conf. Distrib. Comput. Syst., 2002, pp. 617–624.
- [8] D. Ferraiolo and R. Kuhn, "Role-based access controls," in Proc. 15th NIST-NCSC Nat. Comput. Security Conf., 1992, pp. 554–563.
- [9] J. Xu, E.-C. Chang, and J. Zhou. "Weak leakage-resilient client-side deduplication of encrypted data in cloud storage". In ASIACCS, pages 195–206, 2013.