INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume3, Issue4)
Available online at www.ijariit.com

Authorized Deduplication of Files in Cloud Environment

Shrikrishna Kerur
M.Tech in CSE, KLE Dr. MSSCET,Belagavi
shrikrishnakerur@gmail.com

Dr. Anand N. Diggikar
Computer Science and Engg. KLE Dr. MSSCET, Belagavi
vm4anand@gmail.com

Abstract: Data deduplication is one of vital information compression procedures for disposing of copy duplicates of rehashing information, and has been generally utilized as a part of cloud storage to lessen the measure of storage room and spare data transfer capacity. To secure the classification of sensitive information while supporting deduplication, the convergent encryption method has been proposed to encode the information before outsourcing. To better ensure information security, this paper makes the primary endeavor to formally address the issue of approved information deduplication. Not quite the same as customary deduplication frameworks, the differential benefits of clients are additionally considered in copy check other than the information itself. We likewise display a few new deduplication developments supporting approved copy check in a half and half cloud design. Security examination exhibits that our plan is secure as far as the definitions indicated in the proposed security show. As a proof of an idea, we execute a model of our proposed approved duplicate check plan and direct testbed experiments utilizing our model. We demonstrate that our proposed authorized duplicate check scheme brings about negligible overhead contrasted with normal operations.

Keywords: Deduplication, Authorized Duplicate Check, Confidentiality, Hybrid Cloud, Convergent Key.

I. INTRODUCTION

Over the whole web, the boundless virtualized assets are given to clients as administrations by cloud computing and it likewise gives concealing stage and usage points of interest. These days the high accessibility of capacity and greatly parallel computing assets are given by cloud specialist organizations at low expenses. As distributed computing turns out to be all the more effective, the expansion in the measure of information is stored in the cloud and the stored information likewise shared by clients with determined benefits, which additionally called get to privileges of that stored information. One of the greatest difficulties for cloud storage administrations is the regularly expanding size of information.

Data deduplication is one of the notable procedure for making information administration more adaptable in cloud computing and furthermore pulled in more consideration as of late. Information deduplication is characterized as a specific information compression strategy for evacuating duplicate copies of rehashing information in information storage. This system is utilized to enhance the use of capacity and furthermore to diminish the quantity of bytes that must be sent in network information transfers. Rather than putting away various information duplicates with a similar substance, deduplication expels repetitive information by putting away just a single physical duplicate of information and referring other excess information duplicates to that physical duplicate. Deduplication can occur in two sorts, as document level or block level. For document level deduplication, it expels copy duplicates of a similar document. For block level, it expels copy pieces of information. It will happen in non-indistinguishable documents.

The information deduplication brings many advantages, yet the security and protection of privacy concerns likewise emerge. The sensitive information stored by clients are helpless to both outsider and insider assaults. Traditional encryption technique gives information secrecy yet it is incongruent with information deduplication. In traditional encryption, different clients will encode their information with their own particular keys. Thus, similar information duplicates of various clients will make distinctive ciphertexts. It prompts making information deduplication unimaginable.

II. RELATED WORK

A. Dupless: server aided encryption for deduplicated storage

The specialist organizations for cloud storage to be specific Mozy, Dropbox and others perform deduplication by putting away just a single duplicate of each record that is transferred to spare storage space. The customers ought to traditionally encrypt the documents, despite the fact that, the reserve funds are lost. This pressure is settled by message locked encryption (the indication of convergent encryption is generally noticeable). Albeit characteristically it experiences brute force attacks. The documents can be recouped by falling into a known set. Here a design is proposed. The architecture gives secure deduplicated storage that opposes

brute force attacks and acknowledges it in a dupless framework. In dupless, users encrypt through message based keys that are acquired from a key server through an unaware PRF convention. Consequently, the customers are empowered to store the information that is encrypted with an administration that is as of now existed. The administration will perform deduplication and additions solid privacy ensures. The deduplicated storage with encryption can accomplish space investment funds and execution, that is near the use of storage space comparing to the plaintext information.[1]

B. Message-locked encryption and secure deduplication

Message Locked Encryption (MLE) is a most recent cryptographic primitive, encryption and decryption are done under the key, in which it is gotten from a message. Secure deduplication is accomplished by MLE, an objective is set by many cloud specialist co-ops. The type of uprightness called as tag consistency and privacy, both are given. In light of this, both hypothetical and practical commitments are made. On the practical side, the MLE plans which include deployed plans with security investigation are given. On the hypothetical side, the test will be on solutions of a standard model. Associations are made with deterministic encryption. Hash capacities are secured in light of the worldview of test sample-then-extract and related inputs to convey under various classes of message sources and diverse presumptions. The output demonstrates that MLE is primitive for both hypothetical and practical intrigue.[2]

C. Boosting efficiency and security in proof of ownership for deduplication

For lessening the measure of storage the deduplication method is utilized. Numerous clients need to store same substance for various reasons. In this way putting away and keeping up a solitary duplicate of a record is adequate. In any case, executing these idea prompts different security risks. In that most common one: an owner asserting to have such sort of a record. The paper contributions are: at first present, the Proof of Ownership (POW) conspire that contains all elements of cutting edge solutions however just a little overhead experienced to the contender. At second security system relies on upon data hypothetical and not on the computational presumptions. Feasible streamlining methods additionally proposed to enhance the execution of the plan. Toward the end, the nature of proposed conspire is supported by the broad benchmarking.[7]

D. Twin clouds: an architecture for secure cloud computing

Here a design is proposed for self-assertive calculations and secure outsourcing of information to a commodity cloud that is not trusted. In this technique, the client initially communicates with a cloud that must be trusted and in which it will encrypts the information and confirms the stored information and furthermore plays out the operations in the commodity model that is not trusted. At to start with, calculations will part in such a way the cloud, to the point that is trusted is utilized for the majority of the security basic operations in the period of less time-basic setup, though the inquiries for the outsourced information are parallel prepared on encoded information by quick commodity cloud.[3]

III.PROPOSED SYSTEM

The proposed framework with deduplication strategy is at a high level. Since proposed framework setting is an enterprise network. In which it comprises of a gathering of individuals associated to utilize S-CSP and to store information utilizing deduplication method. In this kind of setting, deduplication as often as possible backups for information reinforcement and catastrophe recuperation applications. In the meantime, it will extraordinarily reduce the storage space. Such sort of frameworks is likewise more reasonable for synchronize applications than richer storage capacity and furthermore utilized for client document backup. In the proposed framework, there are three elements, they are characterized as clients, S-CSP in public cloud and private cloud. In which the public cloud containing S-CSP performs deduplication check. It will check whether the two records have a similar substance and stores just a single of the on the off chance that they are with same substance.

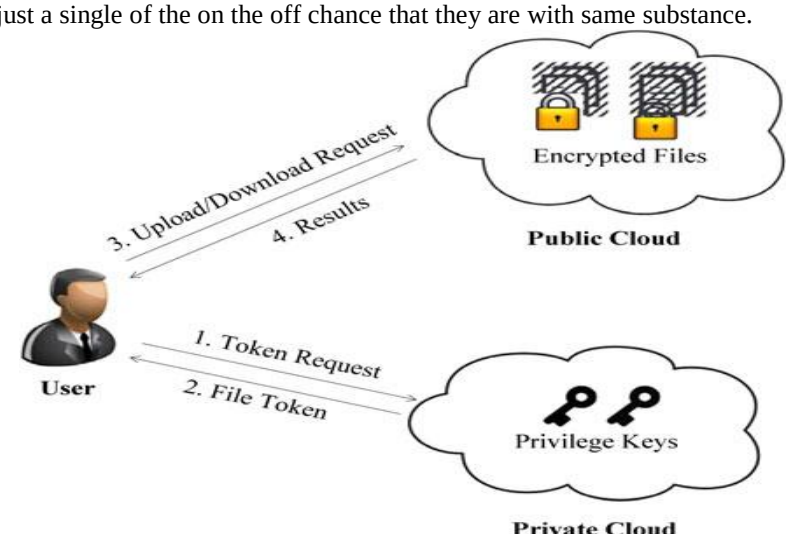

Fig. 1 System Architecture for Authorized Deduplication

To encode the information before it will outsource, the convergent encryption system is utilized. For the better insurance of information security, the paper makes its initially endeavor by formally tending to the authorized information deduplication. Compared with past deduplication frameworks, aside from information itself promote the diverse privileges of clients are likewise considered for copy check. In hybrid cloud architecture, new deduplication developments underpin the authorized copy check.

IV. MODULES

A. Admin Module:

The administrator is the one, who will maintain and take care of all the activities of the users. Admin can log in with his secret username and password that is only known to him. Admin will maintain a log for all the users who are registered to the cloud service and also the history of user uploads, downloads. Admin has the power to activate and deactivate users, as the users registered for the first time the details of the user will be sent to admin. Admin will check and respond back as activating the account by checking user privileges. Once the activation did the user will get the token id for his account for further operations performed in the cloud. The token id generated and maintained in private cloud. The token id has been sent to the user's email id along with the permissions to the operations like upload and download

B. User Registration Module:

The user has to be registered with proper details to get access to the cloud storage if the user registers with all the details he has to wait for account activation by the admin. Once the account is activated then the user can login with the valid username and password. Token id will be generated and sent to the mail id of the user along with file permissions. The user has to enter the valid token id and then only he can perform upload, download operations with respect to his permissions given by admin.

C. Data Upload Module:

In data upload module authorized user can upload files by selecting the file from the local storage. Then the user has to submit for upload, in uploading process, the content of the file computes hash as file tag using the SHA-1 algorithm. Then the duplicate check request will be sent to the cloud, the content of the file is encrypted with the convergent encryption using AES algorithm with the convergent key. Then if the file with the same content already exists in the cloud, it will show the duplicate file exists and it will not be uploaded. The duplication check will be performed with the help of comparing the generated key with the ciphertext. If both key and ciphertext matches then the duplicate file will be detected. If not, the file will be uploaded to the public cloud in encrypted form. Data can be uploaded in 2 forms: Full file upload and block file upload, in full file upload the whole content of the file will be checked for deduplication. While in the block file upload the content of the file will be checked block by block before uploading. If all blocks are duplicate then the file will not be uploaded if any block is original then that block will be eligible to upload.

D. Data Download Module:

The data download will be done only with the authorized users who have the proper privileges to access the file and download. The user can access the files that are stored in encrypted form. The user can view the file before downloading. Once the user clicks on download option the file will be downloaded in the decrypted form.

RESULTS

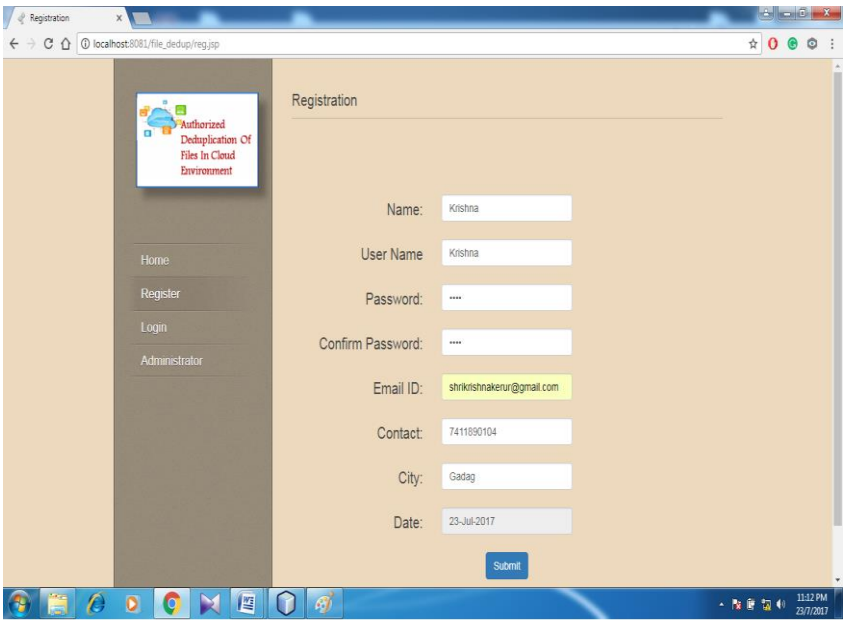

Fig. 2 User Registration

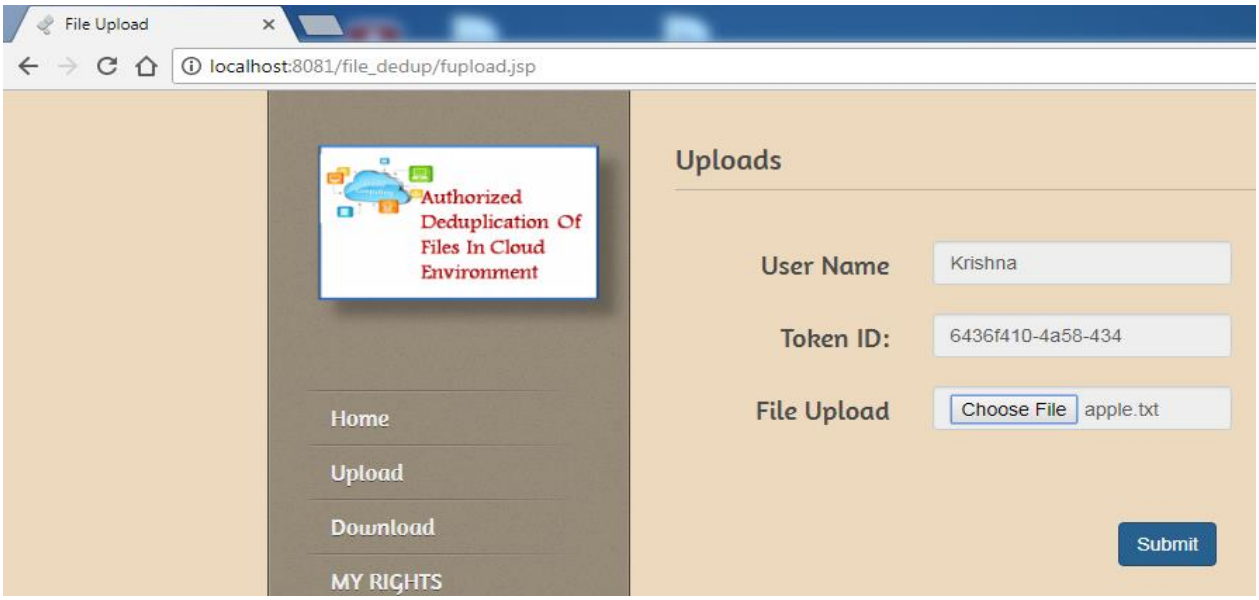

Fig. 3 User uploading file with valid token id

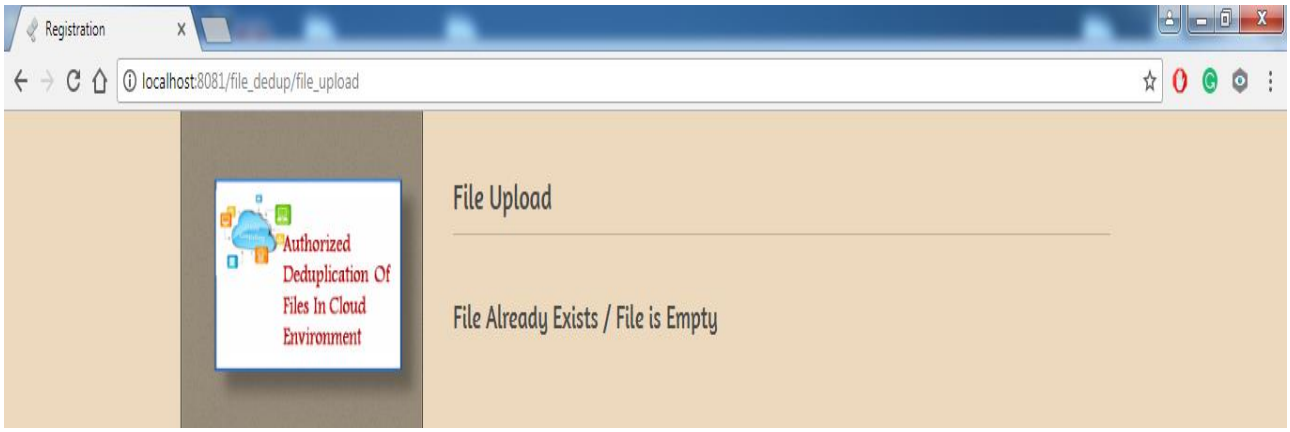

Fig. 4 File duplicate check

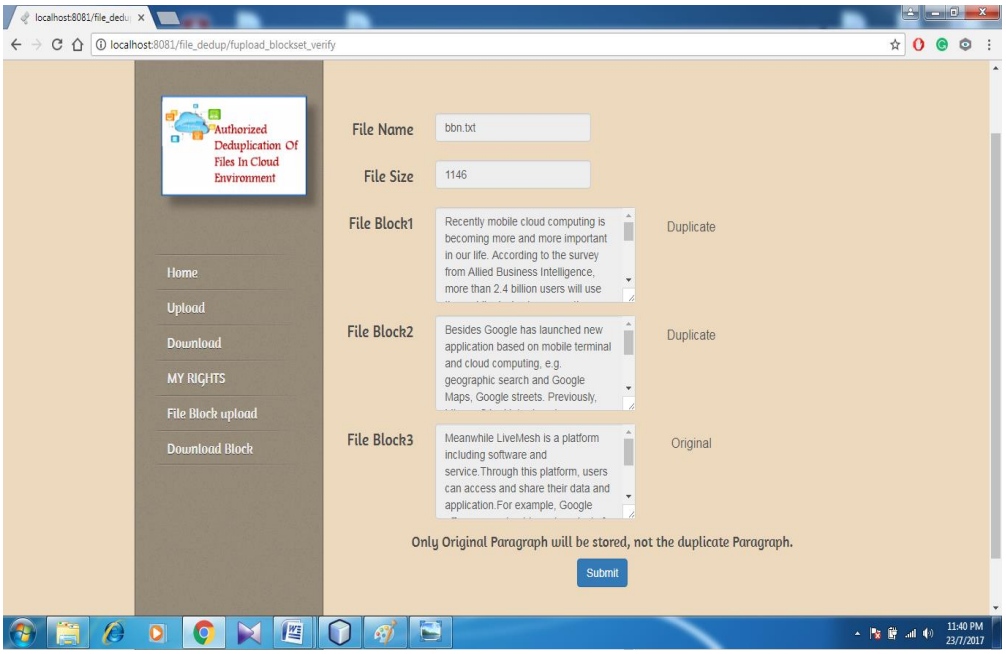

Fig. 5 File block upload and duplicate check

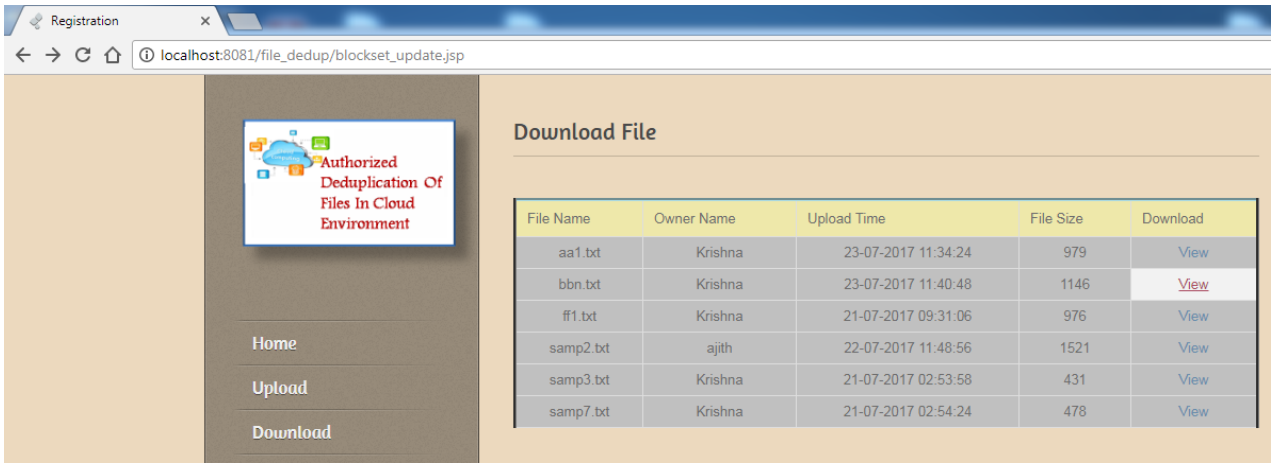

Fig. 6 List files to view and download

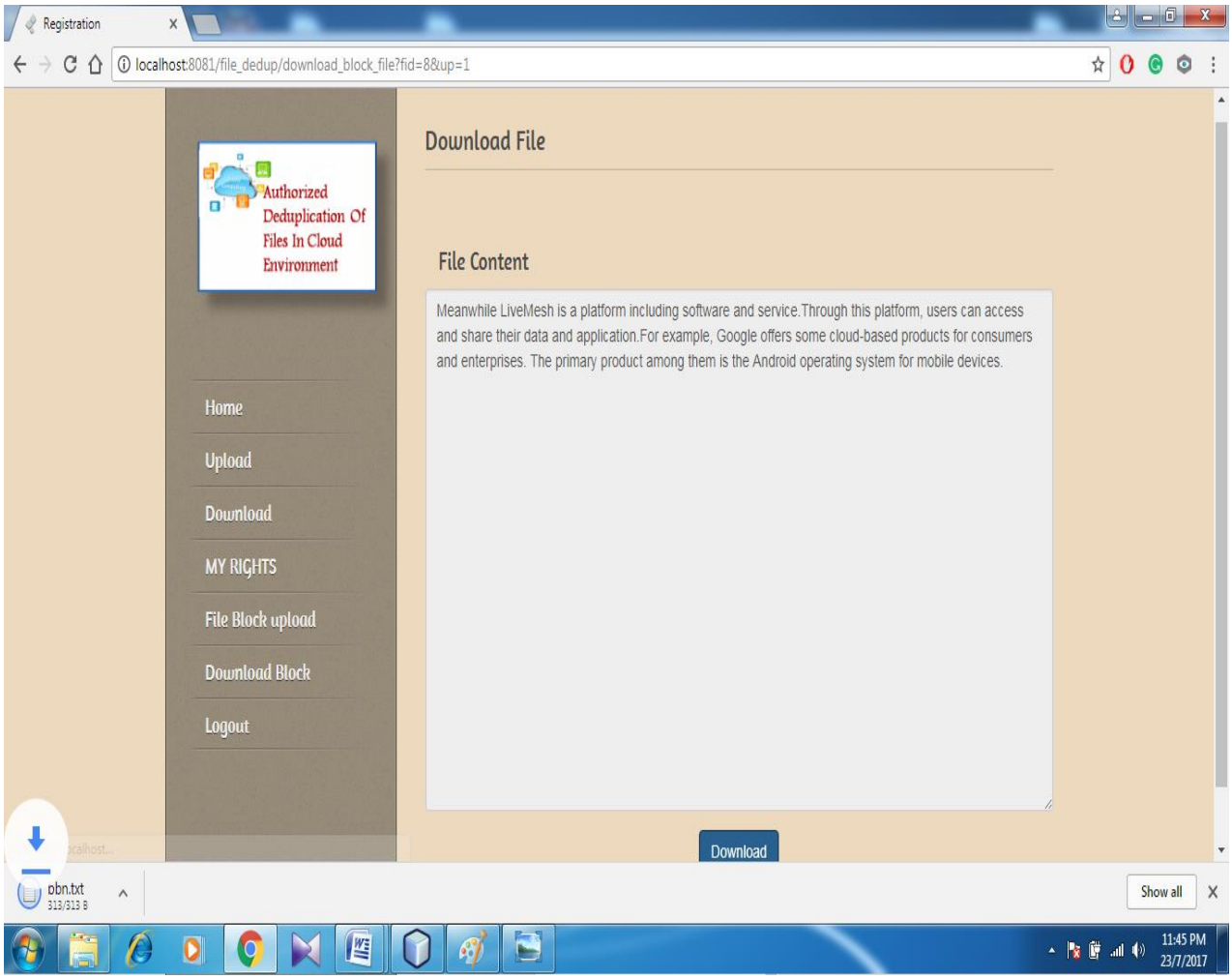

Fig. 7 File can be viewed and downloaded by authorized user

CONCLUSIONS

The authorized information deduplication thought was proposed here to ensure the information security that incorporates differential privileges of users in the copy check. The deduplication developments underpins authorized copy check in the design of hybrid cloud. In which the private cloud server with the private keys creates copy check tokens of records. At security side, the data that is uploaded by users will be stored in the encrypted form. Henceforth, the data will be accessed and decrypted by only

authorized users. Here by, the authorized deduplication check with security will be achieved. Security examination demonstrates that the proposed scheme is secure from outsider and insider attacks.

REFERENCES

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Server aided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Sec. Symp., 2013.
[2] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Proc. 32nd Annu. Int. Conf. Theory Appl. Cryptographic Techn., 2013.
[3] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider, "Twin clouds: An architecture for secure cloud computing," in Proc. Workshop Cryptography Security Clouds, 2011.
[4] 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.
[5] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in Proc. ACM Conf. Comput. Commun. Security, 2011.
[6] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," in Proc. IEEE Trans. Parallel Distrib. Syst., 2013.
[7] R. D. Pietro and A. Sorniotti, "Boosting efficiency and security in proof of ownership for deduplication," in Proc. ACM Symp. Inf., Comput. Commun. Security, 2012.
[8] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for cloud storage," Tech. Rep. IBM Research, Zurich, ZUR 1308-022, 2013.
[9] J. Xu, E.-C. Chang, and J. Zhou, "Weak leakage-resilient clientside deduplication of encrypted data in cloud storage," in Proc. 8th ACM SIGSAC Symp. Inform. Comput. Commun. Security, 2013.
[10] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," IACR Cryptology ePrint Archive, 2013.
[11] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan, "Sedic: Privacy-aware data intensive computing on hybrid clouds," in Proc. 18th ACM Conf. Comput. Commun. Security, 2011.

