Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 7, July 2016, pg.245 – 252

Achieving Data Confidentiality by Usage of Hybrid Cloud and Deduplication

Guljar P. Shaikh[1], Prof. S.D.Chaudhary[2], Prof. P.S.Paygude[3]

[1,2,3] Department of Information Technology, India

[1] guljar.shaikh042@gmail.com; [2] sdchaudhary@bvucoep.edu.in; [3] pspaygude@bvucoep.edu.in

Abstract— Now a day's Data storage over cloud has been very fresh and popular technology. It offers you to store your large amount of data with less cost estimation and to retrieve that stored information anytime from cloud; so it mostly likes by people and researchers. But problem faced by many of the data storage systems is that storing duplicate copies of files having no use. So main confront to cloud services is that managing that ever-increasing quantity of data having only unique information or data copies. For making scalable data management, new technique or term comes in picture which is called as deduplication. In this term deduplication, it stores only unique file rather than keeping multiple copies of the data which having indistinguishable data. Maintaining data privacy and its secrecy is key term considered in any cloud environment. Providing data confidentiality for users or the owners is mainly accomplished by using method Convergent encryption as an alternative to the previous encryption technique. The concept named Duplicate Check which is used for penetrating the records that are duplicated and those are based on results which only stores unique contents. We have shown experimental results that gives us less utilization of storage space and simultaneously consuming less Network bandwidth.

Keywords— Deduplication, Security, confidentiality, duplication authorized check, Convergent Encryption, hybrid cloud.

I. INTRODUCTION

In today's world, there are many cloud storage system that are extensively used for storing your data or files safely and for accessing those stored data/files conveniently. For example, drop box, Google drive etc. are the well-known cloud storage system services. In recent times, Cloud storage systems are used for storing data which is increasing day by day. Data Deduplication is well-known technique that growing rapidly and also strongly used by storage. The Deduplication methods are used for shrinking storage area as well as network bandwidth. Data Deduplication is the way for removing duplicates/repeated files in cloud storage, it is specialized technique in data compression. Rather than maintaining more copies having similar contents, technique called Deduplication removes that repeated data and keeps only one unique copy and also referring other duplicate data to that copy. Using Deduplication technique gives benefits in terms of security and privacy. To achieve confidentiality of mission critical data with data Deduplication, the convergent encryption technique is used, which encrypt data before being sent out. For encryption or decryption of data it uses cryptographic hash value with the help of convergent key which is obtained by computing. Convergent Encryption permits cloud to carry out Deduplication on cipher texts which prevents files from unauthorized user access.

II. TABULATED LITERATURE SURVEY

This has been done in hierarchical way of finding scope in every article and then we have summarized it for identifying existing issues and then we acknowledged proper solution for solving those issues by using different deduplication techniques.

TABLE. I.
TABULATED LITERATURE SURVEY

Existing Paper Name	Characteristics	Issues
A Hybrid Cloud Approach for Secure Authorized deduplication	-Secure Authorization -Server side Deduplication -Duplicate Check	-Whole File based Matching.
Fast and secure laptop backups with encrypted deduplication	-Increase backup speed. -Reduce requirement of the storage space. -Supports client-side per user encryption.	- Network bandwidth can be a bottle-neck. - Costly Backing up.
DupLess: Server- Aided Encryption.	-Occupy less space for storage -Determination Of cross user Deduplication -Stronger security in opposition to External attacks. -High Performance.	-Crucial to slow down brute force attacks. -Low performance. -Increased storage requirements.
Secure Dedup: Server Deduplication with Encrypted Data for cloud Storage	-Achieves confidentiality - Allow block level Deduplication.	Deduplication with high cost
Convergent Key management	-Reduce storage area and bandwidth -Efficient -Management Of key reliably. -Data confidentiality	-Enormous storage overhead in key management. -Master key presents single-point-of- failure.
Private data deduplication Protocols in cloud storage	-Improved deduplication speed. -Fault tolerant -Less storage space	Improve efficiency.
Proofs of Ownership	-Saves Time -precise security -Require less bandwidth -Identification of attacks	-Impossible to verify input distribution
Twin clouds	-Secure computation -Low latency -execution environment is secured	-Personnel of cloud provider could misuse their capabilities -vulnerabilities in shared recourses.

Data Deduplication Scheme for cloud computing	-Improves the speed of data deduplication - Integrity for files.	Cloud storage server will regard all blocks as a new blocks and may store all blocks, resulting in storing duplicate blocks.
Weak leakage –resilient client side Deduplication of encrypted data in cloud storage	-Secure client –side deduplication scheme. - Cross-user security.	Encryption methods are not semantically secure

III. NOVEL APPROACH FOR DATA DEDUPLICATION

Ever since the utilization of storage competence has grown to be a significant matter in storage over cloud, erasing replicated information can use less space for storing files and it leading to more efficient techniques. This makes things easier and enhances process of management of data. Deduplication is a technique for using minimum storage space which is simply storing just distinctive data. Deduplication algorithm identifies duplicated data in comparison with stored data in storage by make use of hash values. And this Hashing reduces the complexity because the hash size is so lesser than the data. So in simple way we can say that the process of Deduplication decreases expense of storing the data since it make use of smaller disks and so less disk purchase will be required.

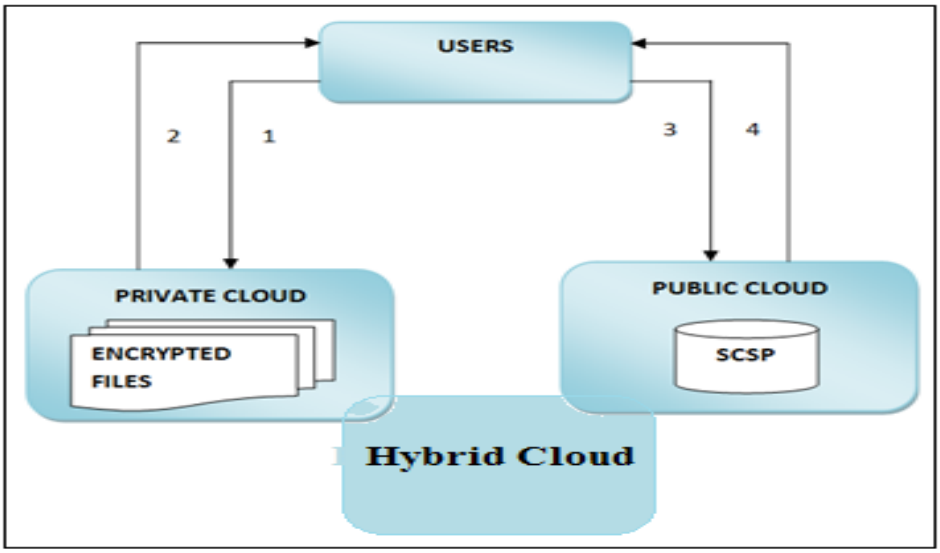


Figure 1. Architecture of Secured deduplication

A. Steps:

i. Token Request for file uploading.
   a. User send request to private cloud for File Token.
   b. Private cloud generates File token and done File tagging
   c. Then result is sent to user.

ii. Response to user.

iii. User sends Request for uploading and downloading files.
   a. Request is forwarded to private cloud it checks for possible deduplication and response back.
   b. Public cloud forwards that appropriate response to the clients.

iv. Results : Based on response Uploading and downloading done.

B. Architecture Explained:

i. *S-CSP*: Storage provider it is entity which makes available storage services of data inside a public cloud. S-CSP offers outsourcing data services and then it stores that data in support of users of the data. All the files uploaded by the users are captured in SCSP by its nature of File type, file Tag and individual File Token generated at time of upload request.

ii. *Users:* The users are the people who wish to outsource their data in public cloud to S-CSP then it access that stored data later whenever required. In this system of storage, a user uploads only unique single files by applying deduplication approaches although it is impossible for them to upload any duplicated data files as those files are denied by private cloud only on upload request by checking duplicate check constraints.

iii. *Private Cloud:* Private cloud concept initiated for smooth the progress of consumer's secure utilization of service provided by the cloud. It acts as a bridge between the owner of File and a public cloud. All operations performed on File like file tagging, token generation, duplicate checking and generation of secrete keys are executed on private cloud. As per its results the response for file originality is given to cloud.

iv. *Public cloud*: Public cloud is similar to S–CSP where the user uploads and downloads their files. Whenever the user wants to download a particular file asks for the secret key and this key is generated as well as stored in private cloud and that has been sent to their registered email address or phone no for providing more security to the files. Every user can access its corresponding files if file token is equivalent with that token which is stored at the S-CSP server. In case of duplication; Reference pointer has been generated and that File reference for downloading given to users.

Operation of public Cloud:

1. File Uploading.
2. File Downloading.

C. Advantages:

i. Maintain data confidentiality.
ii. Stores large data in structured way.
iii. It builds less overhead comparatively usual operations on file.
iv. Provide more Security for files.

IV. RESULTS

A. Storage Space

Before deduplication technique comes into scene there was issue faced by many of user that problem is storing data over cloud and consuming more Space over there. So Deduplication is invented to consume less space and more storage of data. We have shown the difference between required spaces with normal storage and with deduplication storage.

TABLE. II.
REPRESENTATION OF SPACE

Number Of Files (in Thousands)	Normal Storage space(in GB)	Deduplication Storage space(in GB)
3	50	0.2
6	100	0.3
9	150	0.4
13	250	0.6
15	300	0.6
25	400	0.7
35	500	0.8
50	600	0.9
60	700	1
70	800	1

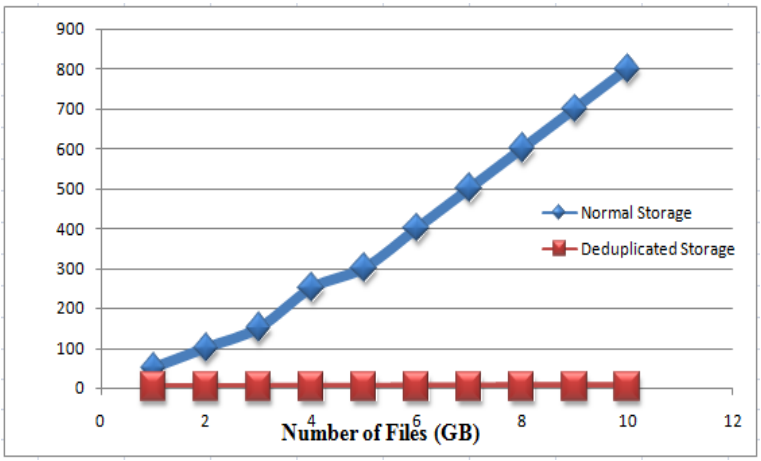


Figure.2. Analysis of Storage space

B. *Execution time*:

The execution time for uploading any file includes duplication check and encryption of file is as shown in below figure. The graph analysis for execution time shown

TABLE. III.
REPRESENTATION OF EXECUTION TIME

Approach	Execution Time (MS)	File Size (KB)
Existing approach	80	160
Our Approach	40	160

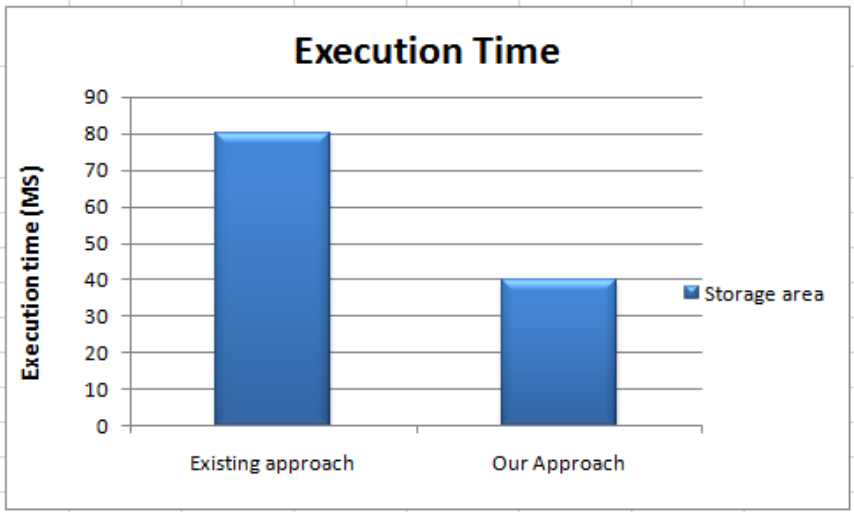
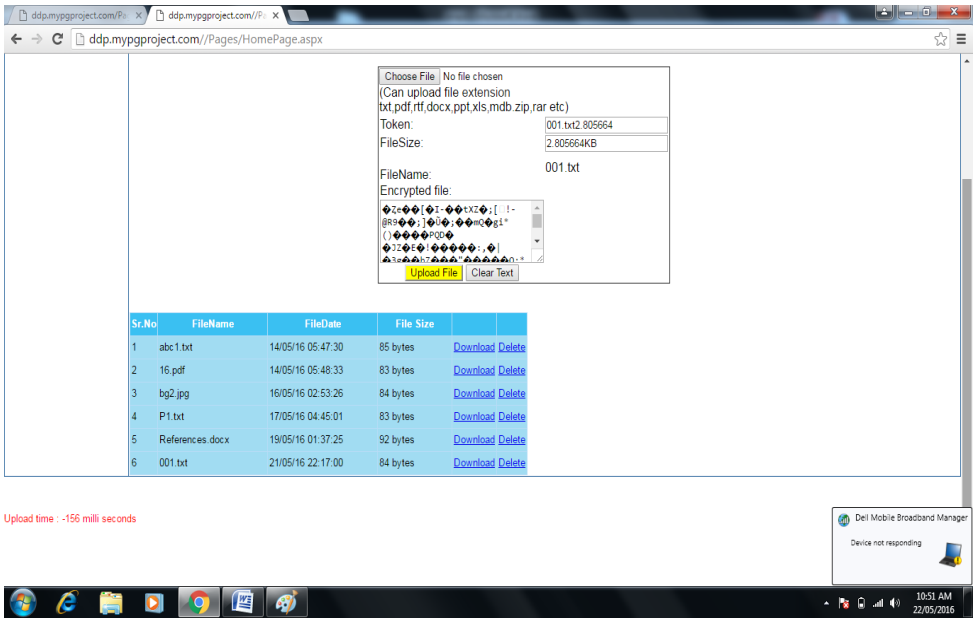


Figure.3. Representation of Execution Time

C. *File Size:*

We have shown result of File size in aspects of:

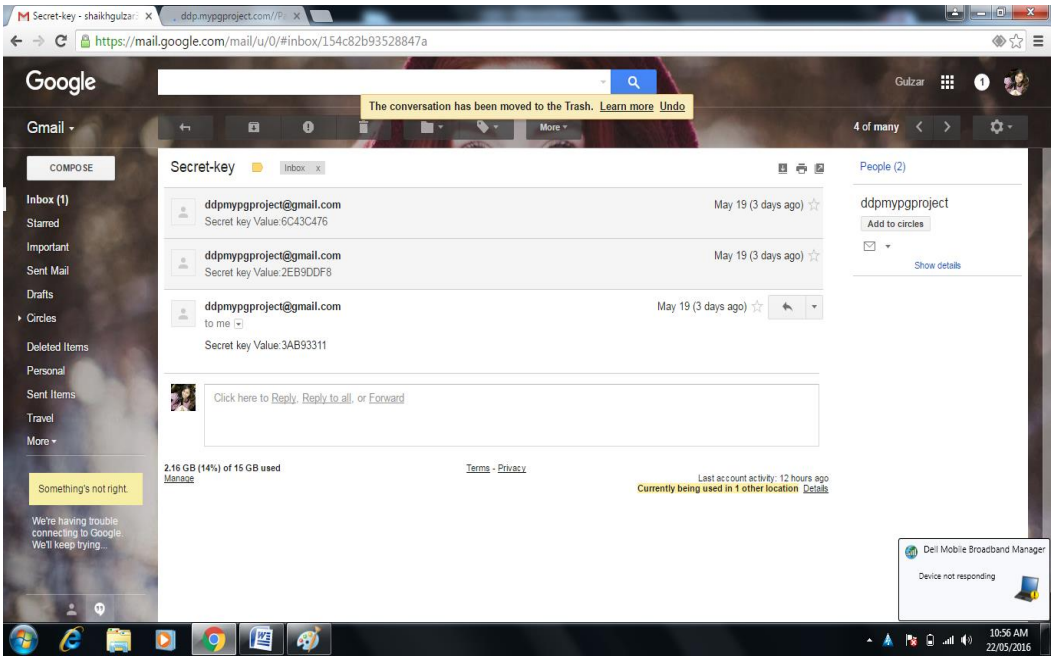
i. Size of the File - before encryption
ii. Size of file - after encryption.



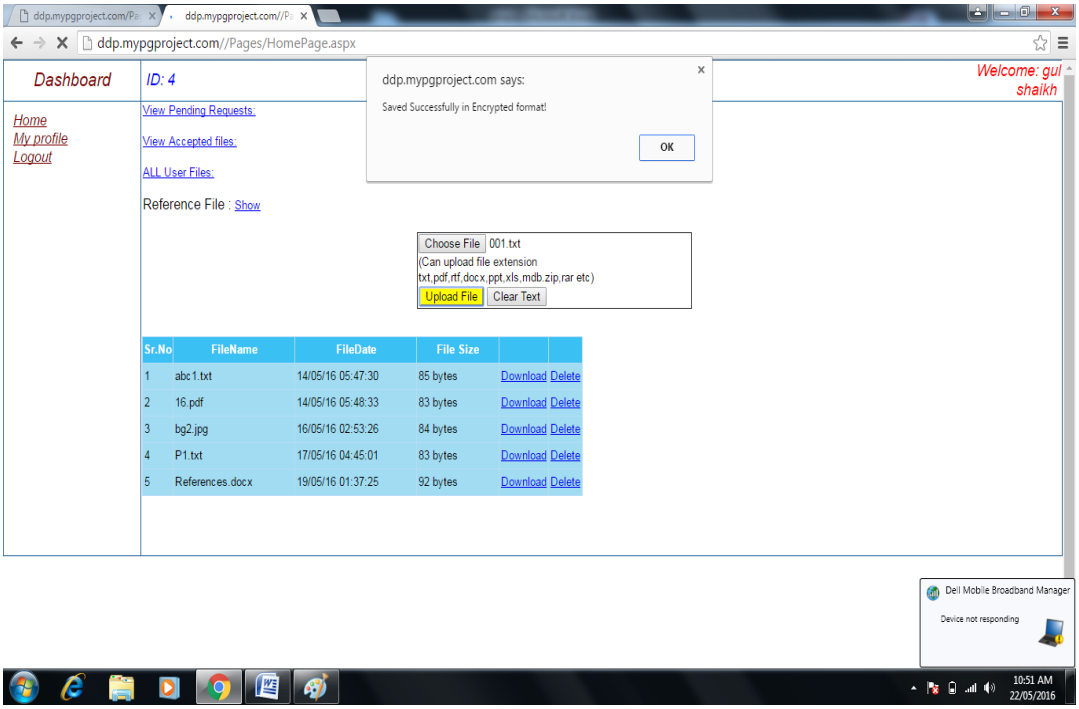
D. *Security*:

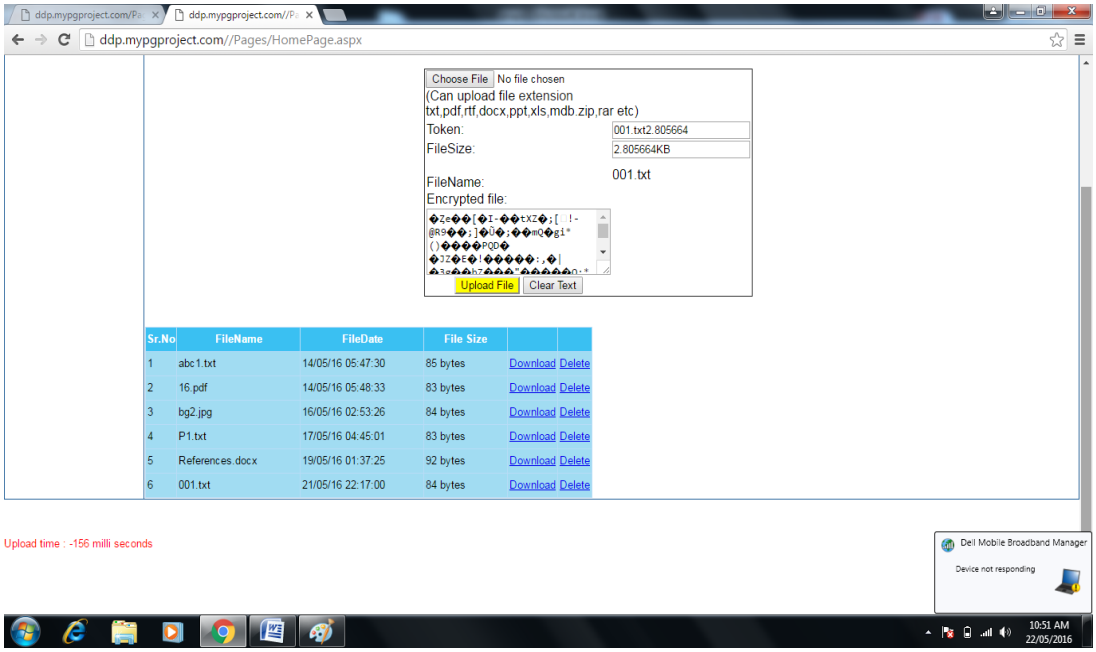
For enhanced verification and authorization we have designed a two-step authentication:

i. By offering vibrant OTP/Secret key on to the user's registered Email address.



ii. By Encrypting File





V. CONCLUSION

Here in the security aspect we have tried to expose that our scheme is secure in terms of the attacks either may by the insider or by the outsider. By encrypting file and sending OTP, we have tried to give more security to our files by two step authentication. So we can say that data Deduplication method makes less use of storage space as well as minimum usage of bandwidth.

VI. FUTURE WORK

We have designed system in consideration of single cloud for storage, in future we can also do duplication on Email or on Phone details. Also we can try for new technology like Hadoop and map-reduce for checking performance.

REFERENCES

[1] Guljar P. Shaikh,S. D. Chaudhary, Priyanka Paygude and Debnath Bhattacharyya, "Achieving Secure Deduplication by Using Private Cloud and Public Cloud", In International Journal of Security and Its Applications Vol. 10, No. 5, 2016, pp.17-26.

[2] Jin Li,yan kit li,xiaofengchen, patrickP.C.lee and wenjing lou,"A Hybrid Cloud Approach for Secure Authorized Deduplication", In IEEE Transactions on Parallel and Distributed Systems, April 7,2015,pp 1206 – 1216, DOI:10.1109/TPDS.2014.2318320.

[3] MihirBellare, SriramKeelveedhi, and Thomas Ristenpart, "Dupless: Serveraided encryption for deduplicated storage", In USENIX Security Symposium, Washington DC, August 14-16, 2013, pp 179-194.

[4] MihirBellare, SriramKeelveedhi, and Thomas Ristenpart, "Message-locked encryption and secure deduplication", in proceedings of EUROCRYPT, Athens Greece, March 3, 2013, pp 296–312.

[5] ShaiHalevi, Danny Harnik, Benny Pinkas, and Alexandra Shulman-Peleg, "Proofs of ownership in remote storage systems". In Proceedings of the 18thACM Conference on Computer and Communications Security, Hangzhou, China, 2011,pp 491–500.

[6] Guljar P. Shaikh, "De-duplication with Authorization in Hybrid Cloud Approach for Security", in International Journal of Computer Sciences and Engineering(IJCSE), Volume-4, Special Issue-4, June 2016.

[7] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. "Twin clouds: An architecture for secure cloud computing". In Workshopon Cryptography and Security in Clouds (WCSC 2011), 2011.

[8] Jin Li, XiaofengChen,MingqiangLi,Jingwei Li and Patrick P. C. Lee, "Secure Deduplication with Efficient and Reliable Convergent Key Management", In IEEE Transactions on Parallel and Distributed Systems, May 12, 2014.pp 1615 –1625, DOI: 10.1109/TPDS.2013.284.

[9] J. Yuan and S. Yu. "Secure and constant cost public cloud storage auditing with deduplication". IACR Cryptology ePrint Archive, 2013:149, 2013.

[10] M. Bellare, C. Namprempre, and G. Neven, "Security proofs foridentity-based identification and signature schemes," J.Cryptol.,vol. 22, no. 1, pp. 1–61, 2009.

[11] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted de-duplication," in Proc. 24th International Conference Large Installation System Admin., 2010, pp. 29–40.

[12] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage," in Proc. 27th Annual ACM Symp. Appl. Computer, 2012, pp. 441–446.

[13] Qinlu He, Zhanhuai Li, Xiao Zhang, "Data Deduplication Techniques", in International Conference on Future Information Technology and Management Engineering China, 2010, pp.43-433.

[14] D. Meister and A. Brinkmann, "Multi-Level Comparison of Data Deduplication in a Backup Scenario", SYSTOR '09 Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference, ISBN: 978-1-60558-623-6, Article No. 8, ACM New York, NY, USA ©2009.

