ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

An Efficient Deduplication and Data Sharing with Dynamic Ownership Management for Cloud

Sukanya Gunjal[1], Prof. B. S. Kurhe[2]

M.E. Student, Department of Computer Engineering, SPCOE, Otur, India[1]

Assistant Professor, Department of Computer Engineering, SPCOE, Otur, India[2]

ABSTRACT: Deduplication may be a storage saving technique that has been adopted by several cloud storage suppliers like Dropbox. In cloud storage services, deduplication technology are commonly accustomed cut back the world and knowledge and data live necessities of services by eliminating redundant knowledge and storing entirely one copy of them. Deduplication is best once multiple users supply an identical data to the cloud storage, but it raises issues concerning security and possession. Issues over information security still forestall several users from migrating information to remote storage. The standard resolution is to write in code the info before it leaves the owner's premises. Client-side information deduplication specifically ensures that multiple transfers of constant contentsolely consume network information measure and space for storing of one upload. We will use server facet information deduplication. During this paper we have a tendency to planned novel server-side deduplication theme for encrypted information.

KEYWORDS: Access Control, Cloud Computing, Data Deduplication, Security.

I. INTRODUCTION

With the infinite storage space that is offered by cloud service providers, users tend to use as much space as they can and vendors constantly look for technique to minimize redundant data and maximize space savings. A technique which has been widely adopted is cross-user deduplication. The simple idea behind deduplication is to store duplicate data (either files/blocks) only one time. Therefore, if a user wants to upload a file (or block) which is already stored, the cloud provider will add the user to the owner list of that file (or block).

Deduplication has proved to obtain high space and money saving. And many cloud storage providers are currently adopting it. Deduplication has proved to achieve high cost reduction, reducing up to 90-95 percent storage needs for backup applications and up to 68% in standard systems. Clearly, the savings, which can be passed back directly or indirectly to cloud users, are significant to the economics of cloud business. Being data deduplication applied in cloud technology can reduce the data storage ,size and save network bandwidth, the dynamicity of data in cloud storage systems are different from backup and archive systems, which brings challenges for implementation of data deduplication in cloud storage systems.

Cloud computing provides an enormous resource pool by linking network resources along. Its fascinating properties, like measurability, elasticity, fault-tolerance, and pay-per-use. Thus, it's become a promising service platform. The most vital and widespread cloud service is information storage service. Cloud users transfer personal or confidential information to the information center of a Cloud Service supplier (CSP) and allow it to take care of these information. Since intrusions and attacks towards sensitive information at CSP, it's prudent to assume that CSP can't be totally sure by cloud users. Moreover, the loss of management over their own personal information, results in high information security risks, particularly information privacy leakages. Thanks to the fast development of knowledge mining and different analysis technologies, the privacy issue becomes serious. Hence, a decent observe is to solely source encrypted information to the cloud so as to confirm information security and user privacy. Deduplication has proven to realize high cost savings, e.g., reducing up to 90-95 p.cstorage wants for backup applications and up to sixty eightp.c in customary file systems. Obviously, the savings, which maybe passed back directly or indirectly to cloud users, square measure important to the political economy of cloud business. How to the approach to away to manage encrypted information storage with deduplication in Associate in Nursing economical way could be a sensible issue. However, current industrial deduplication solutions cannot handle encrypted information. Existing solutions for deduplication suffer from brute-force attacks. They can't flexibly support information access management and revocation at an equivalent time, Most existing solutions cannot guarantee responsible, security and privacy with sound performance.

The rest of this paper is organized as follows. Section II gives research background detail and the related work. In Section III, we describe proposed work of our system. In Section IV described the evaluation results on a real system. Finally some conclusions are given in Section V.

II. Related work

A. Hybrid Cloud Approach for Secure Authorized De-Duplication:

De-duplication of data has many forms. Typically, there is no one best way to implement data de-duplication across an whole an organization. Instead, to maximize the benefits, organizations may deploy more than one deduplication strategy. Cloud data storage services mostly refer de-duplication, which removing redundant data by storing only single copy of every file or block. It is very essential to know the backup and backup challenges, when selecting deduplication as a solution.

Advantages: This De-duplication technique reduces the space and bandwidth requirements of data storage services, and is most effective when applied with multiple users, a common practice by cloud storage offerings.

Limitations: Data deduplication does not work with traditional encryption techniques. While using data deduplication technique it should not reduce fault tolerance mechanism. Types of data de-duplication are described below: File-level de-duplication: This de-duplication technique is commonly called as single-instance storage, file-level data de-duplication compares a file that has to be archived or backup that has already been stored by checking all its attributes against the index. The index is updated and stored only if the file is unique, if not than only a pointer to the existing file that is stored references. Only the single instance of file is saved in the result and relevant copies are replaced by"stub" which points to the original file.

B. Content Addressable Storage:

Eliminating multiple copies of any file is a form of the de-duplication. Single instance storage (SIS) environments can detect and eliminate redundant copies of identical files. After a file is stored in a single-instance storage system than, all the other references to same file, will refer to the original, single copy. Single instance storage systems compare the content of files to detect if the incoming file is identical to an existing file in the storage system. Content-addressed storage is typically combined with single-instance storage functionality. While file level de-duplication avoids storing files that are a duplicate of another file, many files that are considered unique by single-instance storage measurement may have a huge amount of redundancy within the files or between files. For example, it would take only one small element (e.g., a new date inserted into the title slide of a presentation) for single-instance storage to through two large files as being different and requiring them to be stored without further de-duplication.

Advantages: CAS system provides higher searching speed for documents.

Limitations: This system only provides performance benefits when there are more read operations than update operations.

Zheng Yan proposed Encrypted Data Management with Deduplication in Cloud Computing. Cloud computing offers a new way to deliver services by rearranging resources over the Internet and providing them to users on demand. It plays an important role in supporting data storage, processing, and management in the Internet of Things (IoT). Various cloud service providers(CSPs) offer huge volumes of storage. Chuan-Mu Tseng present, A cluster-based data deduplication technology. The proposed method excludes bloom filter's false positives that do not need to check all the index tables. It only needs to query one index table, effectively reducing the time to exclude the false positives.

Our approach:-

In this paper, we propose a scheme based on data ownership challenge, data sharing and data security. We present an advanced scheme to support stronger security by encrypting the file with differential privilege keys and to manage encrypted data storage with deduplication. We aim to solve the issue of deduplication in the situation where the data holder is not available or difficult to get involved. Meanwhile, the performance of data deduplication in our scheme is not influenced by the size of data, thus applicable for big data.

III. PROPOSED SYSTEM

We motivate to save cloud storage and preserve the privacy of data holders by proposing a scheme to manage encrypted data storage with deduplication. Our scheme can flexibly support data sharing with deduplication even when the data holder is offline, and it does not intrude the privacy of data holders. We propose an effective approach to verify data ownership and check duplicate storage with secure challenge and big data support. We integrate cloud data deduplication with data access control in a simple way, thus reconciling data deduplication and encryption. We prove the security and assess the performance of the proposed scheme through analysis and simulation. The results show its efficiency, effectiveness and applicability.

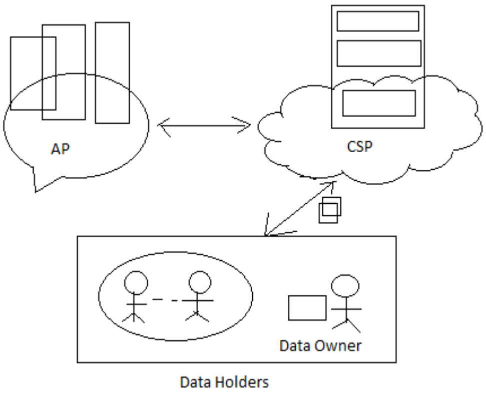


Fig. 1. Block Diagram

IV. EVALUATION

Our scheme works in a system containing three types of entities:

1. A CSP that offers a storage service. A cloud service provider is a company that offers service component of cloud computing - typically Infrastructure as a Service (IaaS), Software as a Service (SaaS) or Platform as a Service (PaaS) - to other business. Cloud service providers are sometimes referred to as cloud providers or CSPs.
2. A data owner that stores its data at the CSP (assume as only one data owner for one data M); and
3. Data holders (ui, i = 1 . . . n) that are eligible data users and could save the same data as the data owner at the CSP.

Our proposed system design data deduplication framework is composed of three steps: In the first step, the data and its metadata is indexed in such a way as to ensure complete data privacy against a semi-honest cloud service provider. The second steps consists in performing the multi-user private keyword searchable encryption on the encrypted data in cloud, keeping the searches and the resulting files secret from the cloud service provider. Step 3 makes use of a strategy to support data sharing between users, by utilizing the existing metadata, the indexing structures, and the searchable encryption scheme. This ensures a data confidentiality in the duplication.

DOI:10.15680/IJIRCCE.2017. 0505257

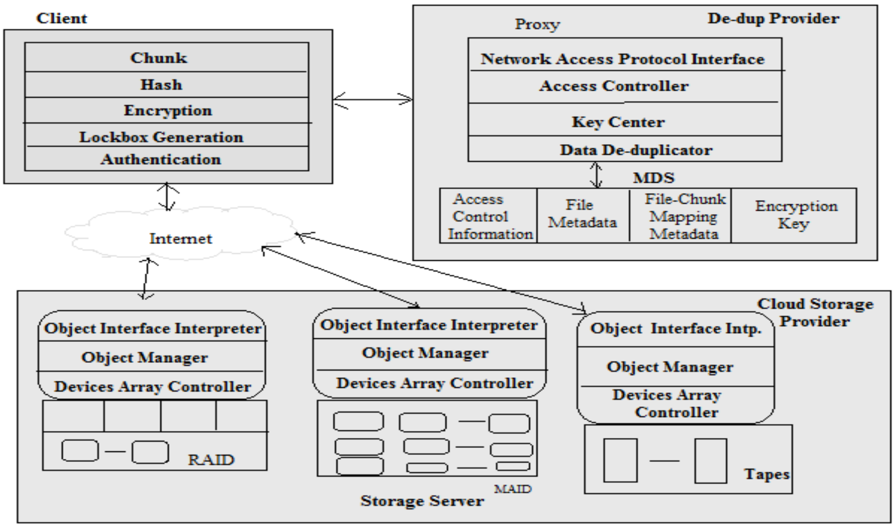


Fig. 2. System Flow

(A) Primitive Function System:

KeyGen(F) : The key generation algorithm takes a file content F as input and outputs the convergent key ckFof F;
Encrypt (ckF;F) : The encryption algorithm takes the convergent key ckFand file content F as input and outputs the ciphertextctF;
Decrypt (ckF; ctF): The decryption algorithm takes the convergent key ckFandciphertextctFas input and outputs the plain file F;
TagGen(F) : The tag generation algorithm takes a file content F as input and outputs the tag tagFof F.

(B) File Uploading:

Phase 1 (cloud client ’! cloud server): client performs the redundancy check with the cloud server to confirm if such a file is stored in cloud storage or not before uploading a file. If there is a duplicate, another protocol called Proof of Ownership will be run between the clients and the cloud storage server. Otherwise, the following protocols (including phase 2 and phase 3) are run between these two entities.
Phase 2 (cloud client ’! auditor): client uploads files to the auditor, and receives a receipt from auditor.
Phase 3 (auditor ’! cloud server): auditor helps generate a set of tags for the uploading file, and send them along with this file to cloud server.

(C) Encryption method for deduplication:

Firstly, this ensures data confidentiality through Attribute based encryption, symmetric key encryption, and PKC. The original user data is encrypted using symmetric encryption with DEK, which is then encrypted using the Encrypt Key algorithm under access policy AP. Assuming that the symmetric key algorithm is secure (for example, using a standard algorithm such as AES), the scheme’s data confidentiality merely relies on the security of the Encrypt Key algorithm.
Secondly, in Convergent encryption users check the Convergent keys from each data set or original data and encrypt the data copy with the generated convergent key. Users also add the tag for the data so that the tag will helps to detect the duplicate data. By using converged key generation algorithm to encrypt the user data. This will ensure the security, ownership and authority of the data. Convergent encryption provides a viable option to enforce data confidentiality while realizing deduplication. It encrypts/decrypts a data copy with a convergent key, which is derived by computing the cryptographic hash value of the content of the data copy itself. After key generation and data encryption, users retain the keys and send the cipher data to the cloud. Since encryption is deterministic, identical data copies will generate the same convergent key and the same cipher data. This makes the cloud to perform deduplication

on the cipher texts. The cipher texts can only be decrypted by the corresponding data owners using their convergent keys.

1. Thirdly, Message-Locked Encryption, A standard message-locked encryption scheme consists of five algorithms, Setup, KeyGen, Enc, Dec, and TagGen.
2. Setup: takes 1�, returns a public parameter P;
3. KeyGen: takes P and a message M, returns a messagederivedkey K;
4. Enc: takes P, message-derived key K and message M, returns a ciphertext C;
5. Dec: takes P, message-derived key K and ciphertextC,returns a message M;
6. TagGen: takes P and ciphertext C, returns a tag T.

V. Snapshots

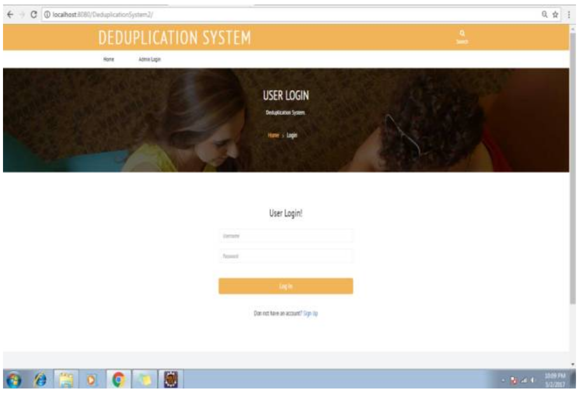


Fig. 1. User Login

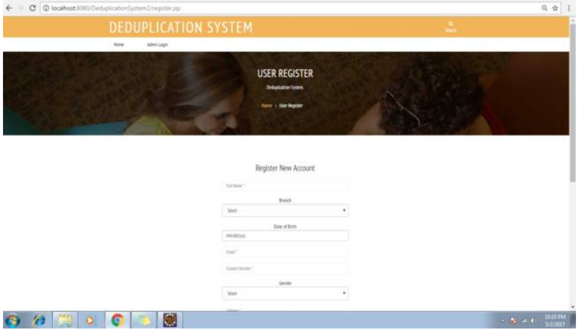


Fig. 2. User Registration

DOI:10.15680/IJIRCCE.2017. 0505257

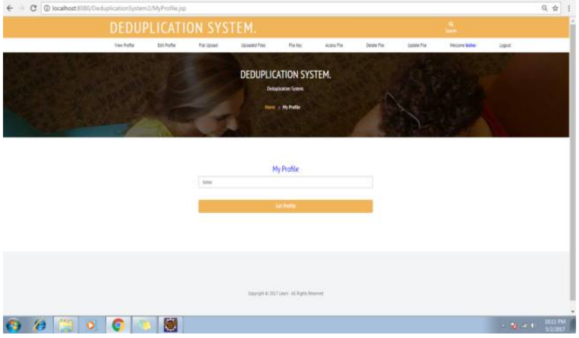
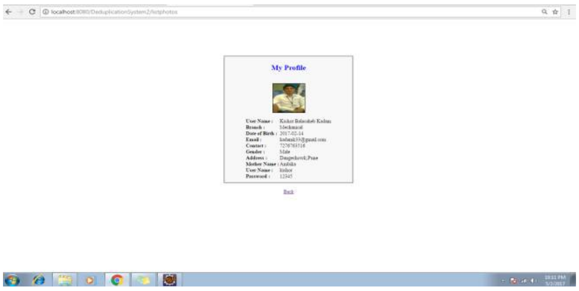
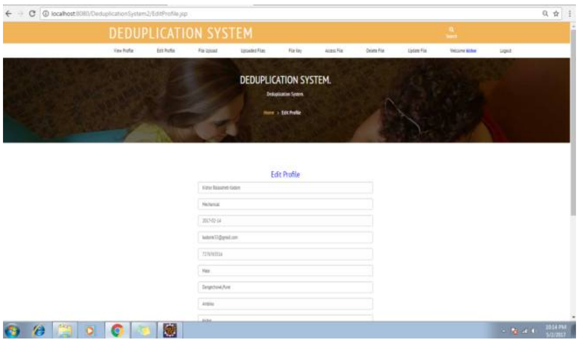


Fig. 3. User Profile


My Profile

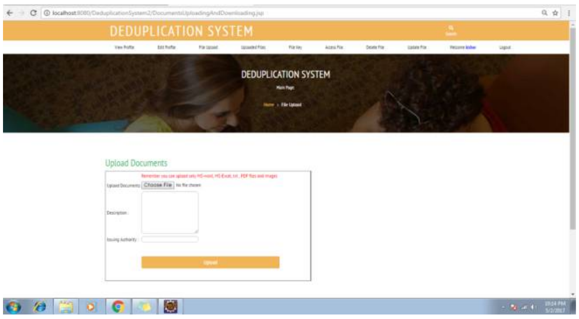
User Name : Kishor Balasaheb Kadam
Branch : Mechanical
Date of Birth : 2017-02-14
Email : kadamk33@gmail.com
Contact : 7276763516
Gender : Male
Address : Dangechowk,Pune
Mother Name : Ambika
User Name : kishor
Password : 12345

Fig. 4. Profile



Kishor Balasaheb Kadam
Mechanical
2017-02-14
Male
Dangechowk,Pune
Ambika
kishor

Fig. 5. Edit profile



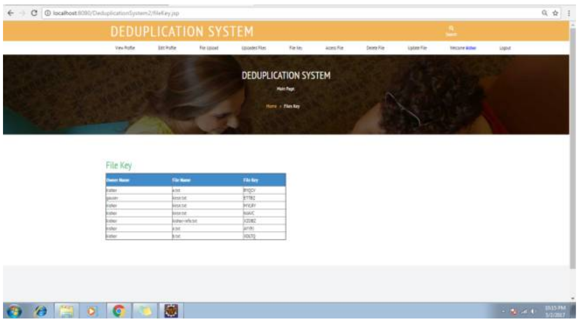
Remember you can upload only MS-word, MS-Excel, txt , PDF files and images

Upload Documents : Choose File No file chosen

Description :

Issuing Authority :

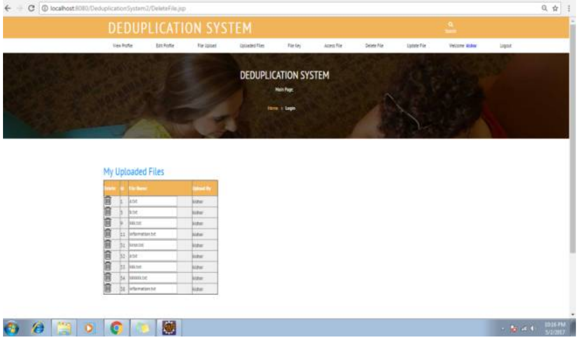
Fig. 6. Upload Documents



File Key

Owner Name	File Name	File Key
kishor	a.txt	BYQCV
gaurav	kesar.txt	ETTBZ
kishor	kesar.txt	MYLRY
kishor	kesar.txt	NIAVC
kishor	kishor-refs.txt	XZDBZ
kishor	a.txt	AYYPI
kishor	b.txt	XDLTQ

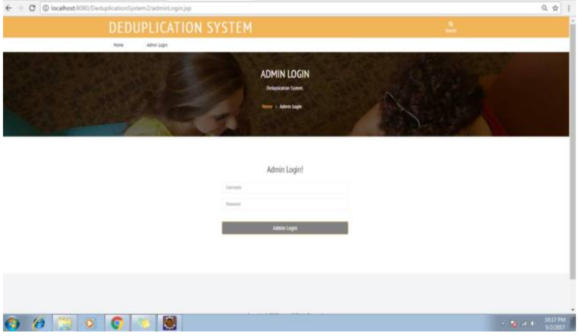
Fig. 7. File Key



My Uploaded Files

Delete	Id	File Name	Upload By
🗑	1	a.txt	kishor
🗑	5	b.txt	kishor
🗑	9	kkk.txt	kishor
🗑	11	information.txt	kishor
🗑	31	lorem.txt	kishor
🗑	32	a.txt	kishor
🗑	33	kkk.txt	kishor
🗑	34	kkkkkk.txt	kishor
🗑	38	information.txt	kishor

Fig. 8. My Uploaded File



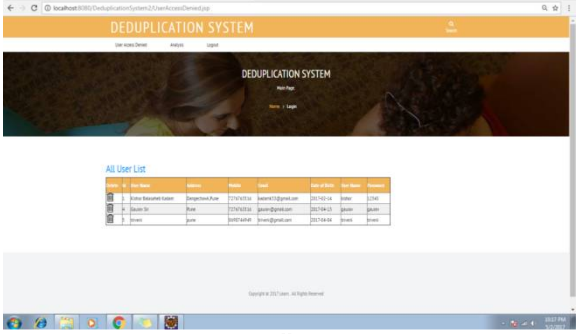
Admin Login!

Username

Password

Admin Login

Fig. 9. Admin Login



All User List

Delete	Id	User Name	Address	Mobile	Email	Date of Birth	User Name	Password
🗑	1	Kishor Balasaheb Kadam	Dangechowk,Pune	7276765516	kadamk13@gmail.com	2017-02-14	kishor	12345
🗑	4	Gaurav Sir	Pune	7276765516	gaurav@gmail.com	2017-04-13	gaurav	gaurav
🗑	5	triveni	pune	8698744949	triveni@gmail.com	2017-04-04	triveni	triveni

Copyright © 2017 Learn . All Rights Reserved

Fig. 10. All User List

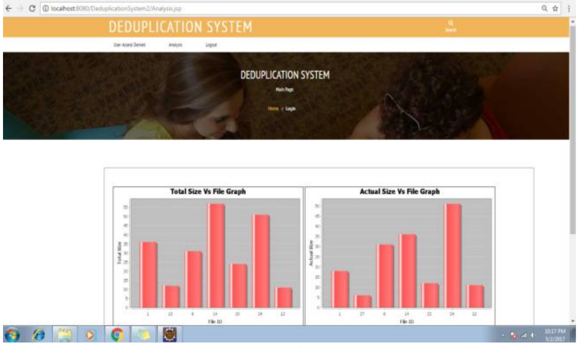


Fig. 11. Result Analysis

X. Conclusion

Managing data with deduplication is an important practice for achieving a successful cloud service, especially for large data storage. In this paper, a practical scheme that manages the encrypted large volume data in cloud with deduplication based on ownership challenge has been discussed. This survey will provide researcher and developer, an idea on privacy in deduplication, hype and challenges which intern facilitates them to evaluate and improve the existing and new deduplication techniques. The real time scenarios with their algorithm implementation have been dealt as a future work of the paper. We will try to save storage space on cloud in order to avoid traffic on server.

XI. Acknowledgement

I express my sincere thanks to my project guide Prof. B. S. Kurhe who always being with presence & constant, constructive criticism to made this paper. I would also like to thank all the staff of computer department for their valuable guidance, suggestion and support through the paper work, who has given co-operation for the project with personal attention. Above all I express our deepest gratitude to all of them for their kind-hearted support which helped us a lot during paper work. At the last I thankful to my friends, colleagues for the inspirational help provided to me through a paper work.

References

1. W. K. Ng, W. Wen, and H. Zhu, “Private data deduplication protocols in cloud storage,” Proc. ACM SAC’12, 2012.

2. M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, “Secure data deduplication,” Proc. StorageSS’08, 2008.
3. N. Baracaldo, E. Androulaki, J. Glider, A. Sorniotti, “Reconciling end-to-end confidentiality and data reduction in cloud storage,” Proc. ACM Workshop on Cloud Computing Security, pp. 21-32, 2014.
4. P. S. S. Council, “PCI SSC data security standards overview,” 2013.
5. D. Harnik, B. Pinkas, and A. Shulman-Peleg, “Side channels in cloud services, the case of deduplication in cloud storage,” IEEE Security & Privacy, vol. 8, no. 6, pp. 40-47, 2010.
6. C. Wang, Z. Qin, J. Peng, and J. Wang, “A novel encryption scheme for data deduplication system,” Proc. International Conference on Communications, Circuits and Ssytems (ICCCAS), pp. 265-269, 2010.
7. Malicious insider attacks to rise, http://news.bbc.co.uk/2/hi/ 7875904.stm.
8. Kirubakaran R, Mano Prathibhan C, Karthika C., “Cloud Based Model for Deduplication of Large Data” IEEE International Conference on Engineering and Technology (ICETECH), 1-4, 2015.
9. XiaolongXu, Qun Tu., “Data deduplication mechanism for cloud storage systems” International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, 286-294, 2015.
10. N. Lakshmi Pritha, N.Velmurugan, Dr. S.GodfreyWinster, A.Vijayaraj., “Deduplication Based Storage and Retrieval of Data from Cloud Environment” International Conference on Innovation Information in Computing Technologies(ICIICT), 1-6, 2015.
11. N.Jayapandian, Dr.A.M.J.Md.Zubair Rahman, I.Nandhini., “A Novel Approach for Handling Sensitive Data with Deduplication Method in Hybrid Cloud” IEEE International Conference on Control System, Computing and Engineering, 1-6, 2015.

12. JunbeomHur, Dongyoung Koo, Youngjoo Shin, Kyungtae Kang., “Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage” IEEE Transactions on Knowledge and Data Engineering, 99, 1-6, 2016.
13. Mane VIdyaMaruti and Prof. MininathK.Nighot., “Authorized Data Deduplication Using Hybrid Cloud Technique” International Conference on Energy Systems and Applications (ICES), Pune, India,26(5), 1206-1216, 2015.
14. Zheng Yan, Senior Member, IEEE, Wenxiu Ding, Xixun Yu, Haiqi Zhu, and Robert H. Deng., Deduplication on Encrypted Big Data in Cloud” IEEE Transactions On Big Data, 2(2), 138-150, 2016.

BIOGRAPHY

Sukanya Gunjal is a Student in the Computer Engineering Department, SharadChandra Pawar College of Engineering, Savitribai Phule Pune University. Her research interests are Cloud Computing, Communication, Security etc.