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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

Hybrid Cloud Mechanism for Secure Authorized Deduplication

Anuvrat Kulkarni[1], Akshay Pawar[2], Akshay Jadhav[3], Pratik Asarkar[4,] Jyoti. J. Malhotra[5]

Student, Dept. of IT, MAEER' s MIT College of Engineering , Kothrud Pune, Savitribai Phule Pune University Pune , India[1,2,3,4]

Asst. Prof, Dept. of IT, MAEER' s MIT College of Engineering , Kothrud Pune, Savitribai Phule Pune University Pune , India[5]

ABSTRACT: In personal computing devices that rely on a cloud storage environment for data backup, an imminent challenge facing source de-duplication for cloud backup services is the low de-duplication efficiency due to a combination of the resource intensive nature and the limited system resources.Data de-duplication is one of important data compression techniques for 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 confidentiality of sensitive data while supporting de-duplication, 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 de-duplication. Different from traditional de-duplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. We also present several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct tested experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

KEYWORDS: Deduplication authorized duplicate check, confidentiality, hybrid cloud.

I. INTRODUCTIONS

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, de-duplication has been a well-known technique and has attracted more and more attention recently. Data de-duplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. 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, de-duplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. De-duplication can take place at either the file level or the block level. For file level de-duplication, it eliminates duplicate copies of the same file. De-duplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files.Cloud computing is an emerging service model that provides computation and storage resources on the Internet. One attractive functionality that cloud computing can offer is cloud storage. Individuals and enterprises are often required to remotely archive their data to avoid any information loss in case there are any hardware/software failures or unforeseen disasters. Instead of purchasing the needed storage media to keep data backups, individuals and enterprises can simply outsource their data backup services to the cloud service providers, which provide the necessary storage resources to host the data backups. While cloud storage is attractive, how to provide security guarantees for outsourced data becomes a rising concern. One major security challenge is toprovide the property of assured deletion, i.e., data files are permanently inaccessible upon requests of deletion. Keeping data backups permanently is undesirable, assensitive information may be exposed in the future because of data breach or erroneous management of cloud operators. Thus, to avoid liabilities, enterprises and government agencies usually keep their backups for a finite number of years and request to delete (or destroy) the backups afterwards. For example, the US Congress is formulating the Internet Data Retention legislation in asking ISPs to retain data for two years, while in United Kingdom, companies are required to retain wages and salary records for six years.

II. GOALS AND OBJECTIVES

Unforged ability of file token/duplicate-check token. Unauthorized users without appropriate privileges or file should be prevented from getting or generating the file tokens for duplicate check of any file stored at the S-CSP. The users are not allowed to collude with the public cloud server to break the unforged ability of file

1. tokens. In our system, the S-CSP is honest but curious and will honestly perform the duplicate check upon receiving the duplicate request from users. The duplicate check token of users should be issued from the private cloud server in our scheme.

2. In distinguishability of file token/duplicate-check token. It requires that any user without querying the private cloud server for some file token, he cannot get any useful information from the token, which includes the file information or the privilege information.

3. Data Confidentiality. Unauthorized users without appropriate privileges or files, including the S-CSP and the private cloud server, should be prevented from access to the underlying plaintext stored at S-CSP. In another word, the goal of the adversary is to retrieve and recover the files that do not belong to them. In our system, compared to the previous definition of data confidentiality based on convergent encryption, a higher level confidentiality is defined and achieved.

OBJECTIVE

1. To improved integrity
2. To increase the storage utilization
3. To remove the duplicate copies of data and improve the reliability.
4. To improve the security

III. LITERATURE SURVEY

1) Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou “A Hybrid Cloud Approach for Secure Authorized De-duplication” IEEE Transactions on Parallel and Distributed Systems: PP Year 2014

Abstract:-

Data de-duplication is a technique for reducing the amount of storage space an organization needs to save its data. In most organizations, the storage systems contain duplicate copies of many pieces of data. For example, the same file may be saved in several different places by different users, or two or more files that aren’t identical may still include much of the same data. De-duplication eliminates these extra copies by saving just one copy of the data and replacing the other copies with pointers that lead back to the original copy. Companies frequently use de-duplication in backup and disaster recovery applications, but it can be used to free up space in primary storage as well. To avoid this duplication of data and to maintain the confidentiality in the cloud we using the concept of Hybrid cloud. To protect the confidentiality of sensitive data while supporting de-duplication, 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 de-duplication.

In this paper they have proposed
In the proposed system we are achieving the data de-duplication by providing the proof of data by the data owner. This proof is used at the time of uploading of the file. 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.

From this paper we have referred
New de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture in which the duplicate-check tokens of files are generated by the private cloud server with private keys. Proposed system includes proof of data owner so it will help to implement better security issues in cloud computing.

2) Mr Vinod B Jadhav Prof Vinod S Wadne Secured Authorized De-duplication Based Hybrid Cloud Approach International Journal of Advanced Research in Computer Science and Software Engineering.

Abstract:-
Data de-duplication is one of important data compression techniques for 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 confidentiality of sensitive data while supporting de-duplication, 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 de-duplication. Different from traditional de-duplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. We also present several new de-duplication constructions supporting authorized duplicate check in a hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

In this paper they have proposed
To perform the duplicate check for some file, the user needs to get the file token from the private cloud server. The private cloud server will also check the user's identity before issuing the corresponding file token to the user. The authorized duplicate check for this file can be performed by the user with the public cloud before uploading this file. Based on the results of duplicate check, the user either uploads this file or runs POW

From this paper we have referred:
Security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

3) A. Abdul Samadhu, J. Rambabu, R. Pradeep Kumar, R. Santhya Detailed Investigation on a Hybrid Cloud Approach for Secure Authorized De-duplication International Journal for Research in Applied Science and Engineering Technology (IJRASET)

Abstract:-
Data de-duplication is one of important data compression techniques which are for eliminating duplicate copies of repeating data, and has been widely used in cloud storage in order to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting de-duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, papers makes the first attempt to formally address the problem of authorized data de-duplication. Different from traditional de-duplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. Also present several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that the proposed scheme is secure in terms of the definitions specified in the proposed security model.

In this paper they have proposed:
Convergent encryption provides data confidentiality in de-duplication. A user derives a convergent key from each original data copy and encrypts the data copy with the convergent key. In addition, the user also derives a tag for the

data copy, such that the tag will be used to detect duplicates. To detect duplicates, the user first sends the tag to the server side to check if the identical copy has been already stored. Both the convergent key and the tag are independently derived, and the tag cannot be used to deduce the convergent key and compromise data confidentiality.

We have referred:
Authorized data de-duplication was proposed to protect the data security by including differential privileges of users in the duplicate check several new de-duplication constructions that support in authorized duplicate check in hybrid cloud architecture.

4)JadapalliNandini, Rami reddyNavateja Reddy Implementation De-duplication System with Authorized Users International Research Journal of Engineering and Technology (IRJET)

Abstract:-
This paper represents that, many techniques are using for the elimination of duplicate copies of repeating data, from those techniques, one of the important data compression technique is data duplication. Many advantages with this data duplication, mainly it will reduce the amount of storage space and save the bandwidth when using in cloud storage. To protect confidentiality of the sensitive data while supporting de-duplication data is encrypted by the proposed convergent encryption technique before out sourcing. Problems authorized data duplication formally addressed by the first attempt of this paper for better protection of data security. This is different from the traditional duplication systems. The differential privileges of users are further considered in duplicate check besides the data itself. In hybrid cloud architecture authorized duplicate check supported by several new duplication constructions. Based on the definitions specified in the proposed security model, our scheme is secure. Proof of the concept implemented in this paper by conducting test-bed experiments.

In this paper they have proposed:
A Client program is used to model the data users to carry out the file upload process. A Private Server program is used to model the private cloud which manages the private key and handles the file token computation. A Storage Server program is used to store and de-duplicates files. The Client provides the function calls to support token generation and de-duplication along the file upload process.

From this paper we have referred:
We observed that the information to Check de-duplication and upload the files, Fetching the Signs using Hashing Algorithm, Checking for Duplication, file uploading, file downloading and attacker trying to attack(block) the cloud.

5) Sharma Bharat, Mandre B.R. A Secured and Authorized Data De-duplication with Public Auditing International Journal of Computer Applications (09758887)

Abstract:-
The popularity and widespread use of Cloud have brought great convenience for data sharing and data storage. The data sharing with a large number of participants take into account issuers like data integrity, efficiency and privacy of the owner for data. In cloud storage services one critical challenge is to manage ever increasing volume of data storage in cloud. To make data management more scalable in cloud computing field, de-duplication a well-known technique of data compression to eliminating duplicate copies of repeating data in storage over a cloud. Even if data de-duplication brings a lot of benefits in security and privacy concerns arise as user's sensitive data are susceptible to both attacks insider and outsider. A convergent encryption method enforces data confidentiality while making de-duplication feasible. Traditional de-duplication systems based on convergent encryption even though provide confidentiality but do not support the duplicate check on basis of differential privileges. This paper presents, the idea of authorized data de-duplication proposed to protect data security by including differential privileges of users in the duplicate
Check.

In this paper they have proposed:
The security will be analysed in authorization of duplicate check and confidentiality of data. For security of Duplicates check by considering antagonist for both internal and external, will try to break the system and by accessing cloud data or will illegal entrance to system.

From this paper we have referred:

Cloud User: -

A cloud user is which who wants to outsource data on public storage which acts as a public cloud in cloud computing.

Public Storage:-

Public Storage is an storage disk which allow to store the users data on it's with include of authorized and not allow to upload the duplicate data.

IV. EXISTING SYSTEM APPROACH

From the above literature survey we have concluded that existing data de-duplication systems, the private cloud are involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges. Such architecture is practical and has attracted much attention from researchers. The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud. We present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. In this way, the users without corresponding privileges cannot perform the duplicate check. Furthermore, such unauthorized users cannot decrypt the cipher text even collude with the S-CSP.

V. ALGORITHM USED

1) Convergent Encryption

Convergent encryption provides data confidentiality in de-duplication. A user (or data owner) derives aconvergent key from each original data copy and encrypts the data copy with the convergent key. In addition, the user also 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. To detect duplicates, the user first sends the tag to the server side to check if the identical copy has been already stored. Note that both the convergent key and the tag are independently derived and the tag cannot be used to deduce the convergent key and compromise data confidentiality. Both the encrypted data copy and its corresponding tag will be stored on the server side.

A convergent encryption scheme can be defined with four primitive functions:

1. **KeyGenCE**(M)!K is the key generation algorithm that maps a data copy M to a convergent key K.
2. **EncCE(K, M)!**C is the symmetric encryption algorithm that takes both the convergent key K and the data copy M as inputs and then outputs a ciphertextC;
3. **DecCE(K, C)!**M is the decryption algorithm that takes both the ciphertextC and the convergent key K as inputs and then outputs the original data copy M; and
4. **TagGen(M)!T (M**) is the tag generation algorithm that maps the original data copy M and outputs a tag T (M).

2) Proof Of Ownership

The notion of proof of ownership(POW) enables users to prove their ownership of data copies to the storage server.Specifically, POW is implemented as an interactive algorithm (denoted by POW). The verifier derives a short value ϕ(M) from a data copy M. To prove the ownership of the data copy M, the properneeds to send ϕto the verifier such that ϕ= ϕ(M).

PSEUDO CODE

```
Step1:Calculate the two convergent key values
Step2: Compare the two keys and files get accessed.
Step3: Apply de-duplication to eradicate the duplicate values.
Step4: Ifany other than the duplicates it will be checked once again and make the data unique.
Step5: That data will be unique and also more confidential the authorized can access and data is stored.
```

VI. SYSTEM ARCHITECTURE

1. Secret Sharing Scheme:

Secret sharing scheme performs two operations namely Share and Recover. The secret is divided and shared by using Share. With enough shares, the secret can be extracted and recovered with the algorithm of Recover. The input to this module is file. It performs dividing of file into fixed size blocks or shares. These blocks are then encoded and allocated on cloud server at different nodes. When user request for file these blocks are decrypted and by combining these blocks file is given to user.

2. Tag Generation:

In this tag similarity is considered a kind of semantic relationship between tags, measured by means of relative co-occurrence between tags, known as J. coefficient. The input to this block is file blocks. This module assigns tags to each block for duplication check. The output of this module is blocks with tag assigned.

3. Convergent Encryption Module

Traditional encryption, while providing data confidentiality, is incompatible with data de-duplication. Specifically, traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different cipher texts, making de-duplication impossible. Convergent encryption has been proposed to enforce data confidentiality while making de-duplication 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. Since the encryption operation is deterministic and is derived from the data content, identical data copies will generate the same convergent key and hence the same cipher text.

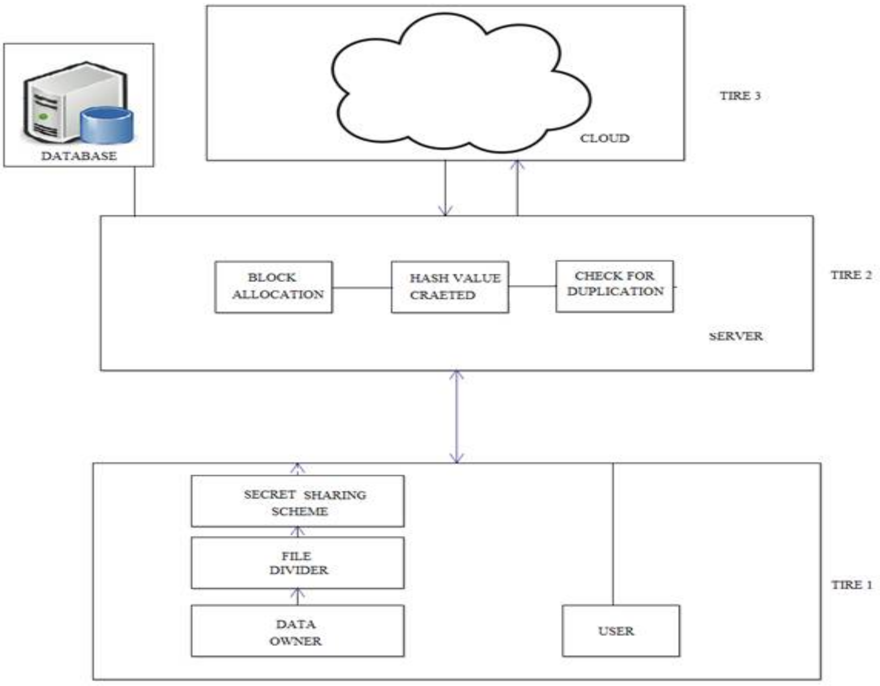

Fig 01. System Architecture.

VII. IMPLEMENTATION RESULT

We implement a prototype of the proposed authorized deduplication system, in which we model three entities as separate C++programs. A Client program is used to model the data users to carry out the file upload process. A Private Server program is used to model the private cloud which manages the private keys and handles the file token computation. A Storage Server program is used to model the S-CSP which stores and deduplicated files. Our implementation of the Client provides the following function calls to support token generation and deduplication along the file upload process.

1. FileTag(File) - It computes SHA-1 hash of the File as File Tag;
2. TokenReq(Tag, UserID) - It requests the Private Server for File Token generation with the File Tag and User ID;
3. DupCheckReq(Token) - It requests the Storage Server for Duplicate Check of the File by sending the file token receivedfrom private server;
4. ShareTokenReq(Tag, {Priv.}) - It requests the Private Server to generate the Share File Token with the File Tag andTarget Sharing Privilege Set;
5. FileEncrypt(File) - It encrypts the File with Convergent Encryption using 256-bit AES algorithm in cipher block chaining
6. (CBC) mode, where the convergent key is from SHA-256 Hashing of the file;
7. FileUploadReq(FileID, File, Token) – It uploads the File Data to the Storage Server if the file is Unique and updates the
8. File Token stored. Our implementation of the Private Server includes corresponding request handlers for the tokengeneration and maintains a key storage with Hash Map.
9. TokenGen(Tag, UserID) - It loads the associated privilege keys of the user and generate the token with HMAC-SHA-1

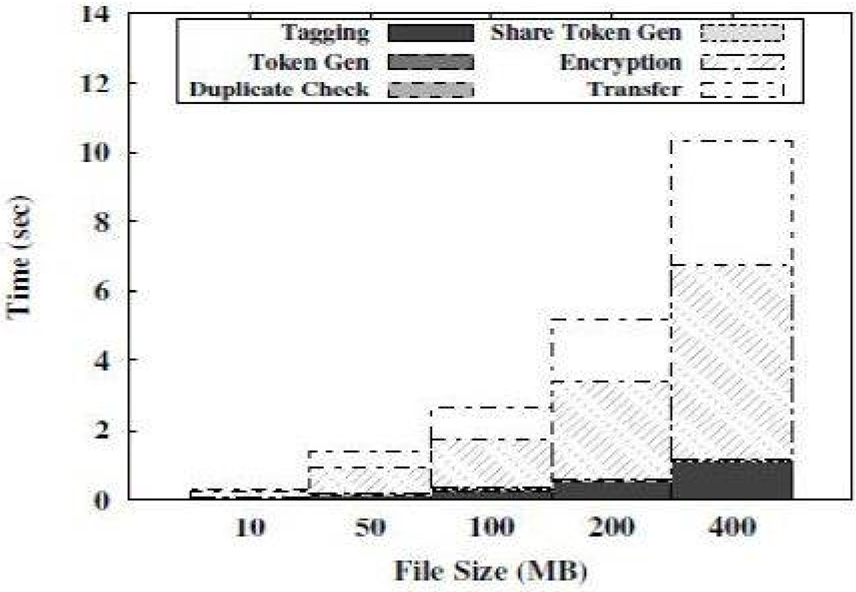

Fig 02 Time Breakdown for Different File Size

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

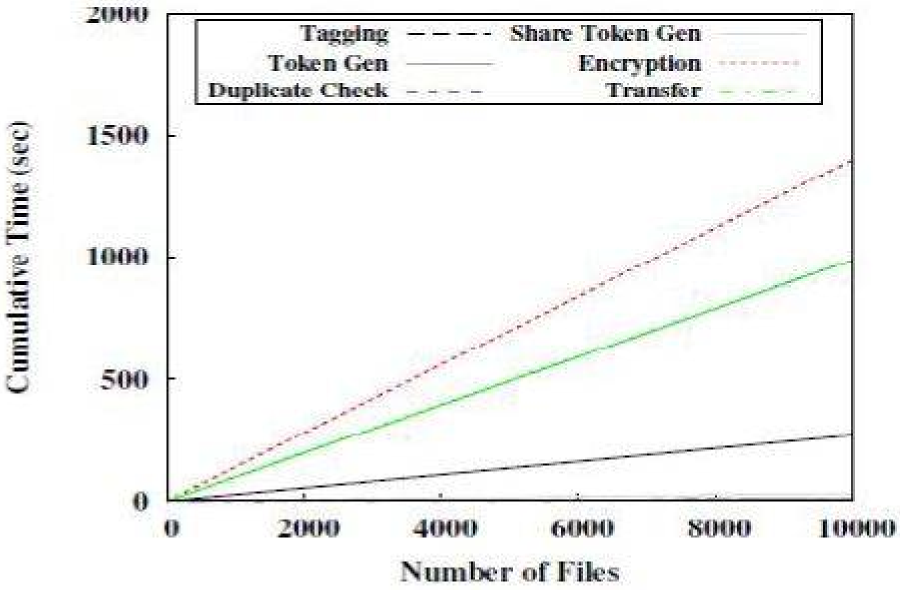

Fig 02 Time Breakdown for Different Number of Stored File Size

CONCLUSION AND FUTURE WORK

Several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture, in whichthe duplicate-check tokens of files are generated by the private cloud server with private keys. Security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct testbed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

REFERENCES

1. Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou “A Hybrid Cloud Approach for Secure Authorized De-duplication” IEEE Transactions on Parallel and Distributed Systems: PP Year 2014
2. Mr Vinod B Jadhav Prof Vinod S Wadne Secured Authorized De-duplication Based Hybrid Cloud Approach International Journal of Advanced Research in Computer Science and Software Engineering
3. Abdul Samadhu, J. Rambabu, R. Pradeep Kumar, R. Santhya Detailed Investigation on a Hybrid Cloud Approach for Secure Authorized De-duplication International Journal for Research in Applied Science and Engineering Technology (IJRASET)
4. JadapalliNandini, Rami reddyNavateja Reddy Implementation De-duplication System with Authorized Users International Research Journal of Engineering and Technology (IRJET)
5. Sharma Bharat, Mandre B.R. A Secured and Authorized Data De-duplication with Public Auditing International Journal of Computer Applications (09758887)
6. Wee Keong Ng SCE, NTU Yonggang Wen SCE, NTU Huafei Zhu Private Data De-duplication Protocols in Cloud Storage SAC12 March 2529, 2012, Riva del Garda, Italy. Copyright 2011 ACM 9781450308571/12/03
7. Shweta D. Pochhi, Prof. Pradnya V. Kasture Encrypted Data Storage with De-duplication Approach on Twin Cloud International Journal of Innovative Research in Computer and Communication Engineering
8. Backialakshmi. N Manikandan. M SECURED AUTHORIZED DE-DUPLICATION IN DISTRIBUTED SYSTEM IJIRST International Journal for Innovative Research in Science and Technology— Volume 1 — Issue 9 — February 2015
9. BhushanChoudhary, AmitDravid A Study On Secure Deduplication Techniques In Cloud Computing International Journal of Advanced Research in Computer Engineering and Technology (IJARCET) Volume 3, Issue 12, April 2014
10. James S. Plank LihaoXu Optimizing Cauchy Reed-Solomon Codes for Fault-Tolerant Network Storage Applications The 5th IEEE International Symposium on Network Computing and Applications (IEEE NCA06), Cambridge, MA, July, 2006.