An Efficient Hybrid Data Deduplication on Clouds using CHAP and Signature based Algorithms

Shivani Sengar
Computer Science
Adina Institute of science and Technology
Sagar (MP)

Ruchika Mishra
Assistant Professor
Computer Science
Adina Institute of science and Technology
Sagar (MP)

Swati Jain
Assistant Professor (HOD)
Computer Science
Adina Institute of science and Technology
Sagar (MP)

ABSTRACT
Data Deduplication refers to a method that diminishes the unnecessary facts on the storage and transmitting on the network, and is think about to be solitary of the most-enabling storage technologies that presents well-organized resource exploitation in the haze computing. Here in this daily a new approach for Data Deduplication is proposed which is founded on the notion of applying Challenge Handshake Authentication Protocol and then applying Elliptic Curve based Cryptography. The Planned procedure applied here provides better Security in Comparison with the Existing Protocols and also provides better Data Deduplication. The Proposed methodology also provides better Security Prevention from various attacks low Computational Cost and Time.

Keywords
Cloud Computing, Data Deduplication, Challenge Handshake Authentication Protocol, Elliptic Curve Cryptography, Authorization.

1. INTRODUCTION
In, today's digital environment currently evolving and securing multimedia content internet over communication channel is a significant challenges day-by-day for cryptographic point-of-view. Mainly these content has developed progressively more significant need to get protection during data transmission over internet know how to exist illusory. It is challenging to solve the problems of data security and privacy on data deduplication but positively essential for contribution an established and established cloud storage service. Cloud computing has garnered much interest in recent years in the computing industry, the media, and academia. It is a form of pay-per-use distributed computing consisting of data centers providing commodity resources for massively scalable units of computing and storage for commercial enterprise applications as well as scientific computing; these facilities are delivered as a service to a global population of users over the Internet and wireless data networks. Cloud computing is a major area of knowledge that efficiently allows data outsourcing as a service using Internet methods with expandable provisioning and usage-based pricing [1]. Many cloud examination vendors present remote data outsourcing and encouragement repairs by utilizing storage space and network resources on cloud storage infrastructures. As the fast growth of data volumes increases demand for data outsourcing on cloud storage services, pay-as-you-use cloud pattern drives the need for cost-efficient storage, specifically for reducing storage space and network bandwidth overhead, which is directly related to the financial cost savings.

Data deduplication fundamentally eliminates duplicate data copies with the intention of make possible a cost-effective storage. It is a type of information solidity system (as single-instance data storage) that employed to avoid data redundancy [1]. There is no inconsistency between duplication and distributed storage system because the technique has to find a common bytes set inside or among files to allow single instance storage of each fragment in the server on the beginning of the replication based erasure coding-based, or network coding-based approaches. Data deduplication technique is considered to be one of the most forceful storage technologies, and it is estimated that the ratio of applying deduplication will increase steadily among the storage service providers [2]. Using cloud servers the remote data auditing service comprises a set of protocols designed to prove the undamaged of the remote data exist in cloud storage more consistently and proficiently, devoid of downloading the complete data. Additionally, the outsourced data is also issued to manage by unpredictable third party cloud providers [3].

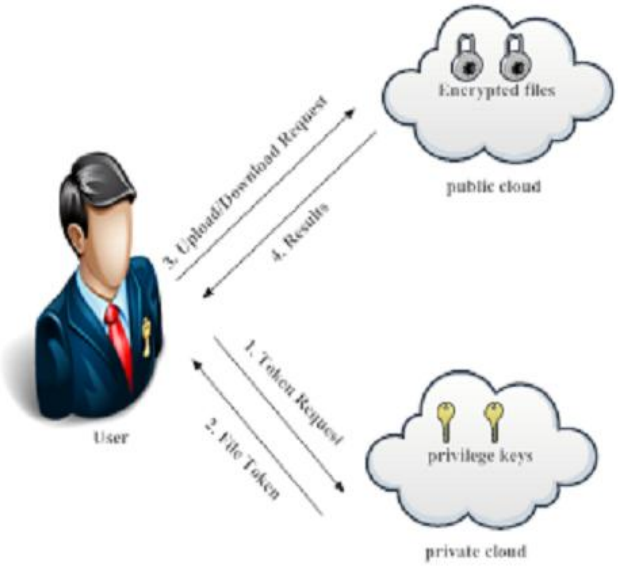

Figure 1: Working of authorized de-duplication [5]

The RDA frameworks use spot checking method to authenticate the outsourced data in which a small section of entire data only is necessitated to be right to use by the assessor. This method presents either a probabilistic or deterministic declaration for the data undamaged [4]. This idea of securing data in cloud stowage amenities has attracted many researchers to work in this field with the aim of constructing a trusted control model of cloud storage. In the cloud stowage organization with data deduplication, untrusted entities including a cloud server and users may cause security threats to the storage system.

By exploiting some vulnerabilities in data deduplication, both an inside adversary, who act as a cloud server, and an outside adversary, who act as a user, will attempt to break data confidentiality, confidentiality and truth on the outsourced data. More concretely, for cloud stowage organization with deduplication, we are concerned with several security issues that are raised by the adversaries: 1) sacrificing data security for deduplication, 2) evidence escape finished side channel, and 3) unauthorized arbitrary data access. The main aim is enterprise all the network. To positioned the data backup and tragedy salvage applications for decrease the storage space. We often go for deduplication. Such systems are widespread and are often more proper to user file backup and bringing together applications than comfortable storage ideas.

2. LITERATURE SURVEY

In this paper [6], novelist has to cracking competently the unruly of deduplication with differential privileges in cloud computing, here they think about a mixture cloud construction consisting of a public cloud and a private cloud. As using presented move toward for data deduplication the private cloud is entailed as a proxy to permit data owner/users to strongly achieve duplicate check with differential benefits. Such construction is expedient and has anxious much consciousness from make investigations from statistics landlords only subcontract their information stowage by exploiting public cloud while the data process is arrangement with in private cloud. A novel technique maintaining discrepancy duplicate ensure is planned below this combination cloud edifice where the S-CSP exist in in the public mist. The operator is only allowable to implement the matching checked for files noticeable with the equivalent rights.

The main goal of this paper is to provide stronger security by encrypting the file with degree of difference license keys. In this method, the users deprived of equivalent freedoms cannot realize the matching checkered. In totaling, such illegal users cannot decode the ciphertext even join composed with the S-CSP.

As their planned technique has to approve matching check and behavior test-bed trials to analyze the upstairs of the sample. Security examination displays that their organization is protected in terms of the descriptions precise in the planned refuge classical. Here they demonstration that the upstairs is negligible associated to the standard convergent encryption and file upload processes.

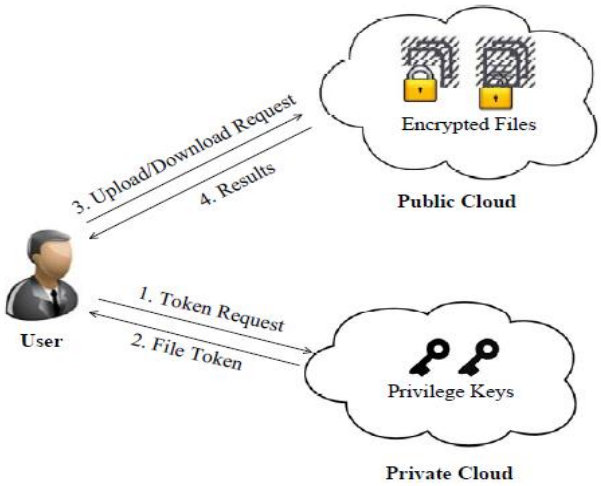

Figure 2: Planning for sanctioned deduplication[6]

To guard the confidentiality novelist has been planned [7] to encode the information before subcontracting. To improved defend data refuge this newspaper makes the initial effort to legitimately essence on the struggle of lawful statistics deduplication. Unusual from conventional deduplication systems the degree of difference rights of workers are additional measured in matching checked as well the data itself. Here they also present common new deduplication structures sustaining legal matching checked in mixture cloud construction. Security examination demonstrations that their technique is protected in languages of the imageries precise in the anticipated security representation. As a proof of idea, they put into practice a prototype of our proposed approved duplicate check method and behavior test bed experiments using our prototype. We validate that our recommended lawful replica checkered technique bring upon by hand in significant photograph appraised to standard processes. It keeps the memory by deduplicating the data and thus makes available us with enough memory. It provides authorization to the private firms and protects the confidentiality of the significant data.

To achieve a protected and reliable cloud storage service, a secure multi-owner data sharing method is proposed [8] rendering to any operator in the collection so that they can steadily portion information with others users by the un-trusted cloud. The Group executive is used for decrease of the execution time of the key generation at the user end or data owner side. Public-key cryptosystem construct constant-size ciphertext as well-organized designation of decryption privileges for any set of ciphertexts are attainable. Anyone can comprehensive any set of secret keys and make them as compressed as a single key. The private key proprietor can generate a continuous size collective important of ciphertext set in cloud, but another encoded files external stay behind secret. The aggregate key strongly sent to users or keep in a smart card with limited storage. We characterize recognized investigation of security in the average model.

In particular, their approach [8] is more elastic than graded key project which can only except chairs if all key-holders portion a comparable set of privileges methods give the first community important patient measured encryption for elastic grading, which was pending now to be recognized. The difficult trouble is how to efficiently share encrypted data. Obviously workers can transfer the encoded statistics from the stowage, decrypt them then direct them to others for allocation, but it misplaces the worth of cloud stowage. Users should be able to delegate the access rights of the sharing data to others so that they can access

these data from the server directly. However, finding an efficient and make safe way to share unfinished data in cloud storage is not insignificant. An inadequacy of their exertion is the predefined bounce of the quantity of supreme ciphertext classes. In mist stowage, the quantity of ciphertexts more often than not produces quickly. So we have to hold back an adequate amount of ciphertext classes for the upcoming expansion.

In this paper, author [9] presents a new discretion preserving refuge explanation for cloud amenities. Here in this technique transaction with user indefinite admittance to cloud amenities and communal stowage wait persons using non-bilinear collection autographs to guarantee unidentified confirmation of cloud provision client's user. Users use tamper resistant devices during the generation and storing of user keys to protect against collusion attacks. On the other pointer, if operators break provider's rules, their access rights are withdrawn. Here we analyze modern privacy preserving solutions for cloud services and summarize our explanation based on advanced cryptographic components it also offers anonymous access, unconnected ability and the confidentiality of transmitted data. Due to this fact, cloud provision earners using our solution can authenticate more clients in the same time. Additionally, there method gives output the untried results and degree up to the performance with related solutions.

Proposed Methodology

The Planned Procedure implemented here is based on the ASymmetric Encryption that uses a Secrete Key 'K1' & 'K2'.

Step 1: User direct appeal to organization for challenge charge.

Step 2: Scheme take challenge assessment.

Step 3: System calculate timestamp T1.

Step 4: System take password value.

Step 5: System send challenge value + T1.

Step 6: User received challenge value + T1.

Step 7: User calculate current timestamp T2.

Step 8: User calculates total transmission time = 2 * (T2 - T1) + processing time.

Step 9: User enhances broadcast time + t1 to tot_time.

Step 10: User take keyword.

Step 11: Worker sregulate MD5 hashing purpose on contest worth + pwd + tot_time.

Step 12: User compute MD5 chopping on this statistics.

Step 13: User direct this statistics to organization.

Step 14: scheme conventional statistics D1.

Step 15: organization estimate timestamp T3.

Step 16 Organization regulates (contest value + password + T3).

Step 17: Scheme governs MD5 mincing on (encounter value + keyword + T3).

Step 18: If it matches then session is valid. Cheek whether the password valid or not

if valid send allowed

else send not allowed

else session expires.

Step 19: User will show whether session expires or not.

If not expired then whether password valid or not.

3. PROPOSED ALGORITHM

Setup: Here in this phase first of all the Elliptic Curve Parameters are set and public and private key pairs are generated using KeyGen(.). Suppose the General Elliptic Curve Equation is defined by:

$$y^2 = ax^3 + bx + c$$

Where, $4a^3 + 27b^2 \neq 0$

Client chooses any random point over elliptic Curve E(F) that would be the chosen Secrete key of the client sk using secrete key and Common Base Point B public key is generated.

Pk=Sk.P

SigGen: The Shared Data File F={m1,m2…mn}, first of all choose a random integer 'u' and hence generate Tag for the Shared Data File F using

$$T_m = name||Nd||u||Sig$$

Client Starts generating Signatures Sg for each of the block mi,

$$S_g = (H(mi).u^{mi)^{\alpha}}$$

The Client Generating of Linked List based on the signatures and create a First Node of the Linked List and the other Nodes are constructed using H(mi).

Client Signs the Generated Started Linked List Root Node using secrete key sk

$$sig_{sk}\big(H(R)\big) \leftarrow (H(R))^{\alpha}$$

Client Sends {F,Tm,Sg,$sig_{sk}\big(H(R)\big)$) to Third Party Auditor (TPA).

Data Deduplication: When the Block is received to the TTP will checks the Data is Already stored to the Storage Panel or not. If already Stored then Discarded, otherwise stores in Storage Panel.

The Figure exposed below is the flow diagram of the planned procedure. The Data Owner who wants to share Data Over multiple receivers needs to be authenticated to the Trusted Third Party. If Data Owner is trusted Party then only Data is Shared and During the Sharing of Data TTP/CA Checks Data Deduplicity and stored in the Storage Panel.

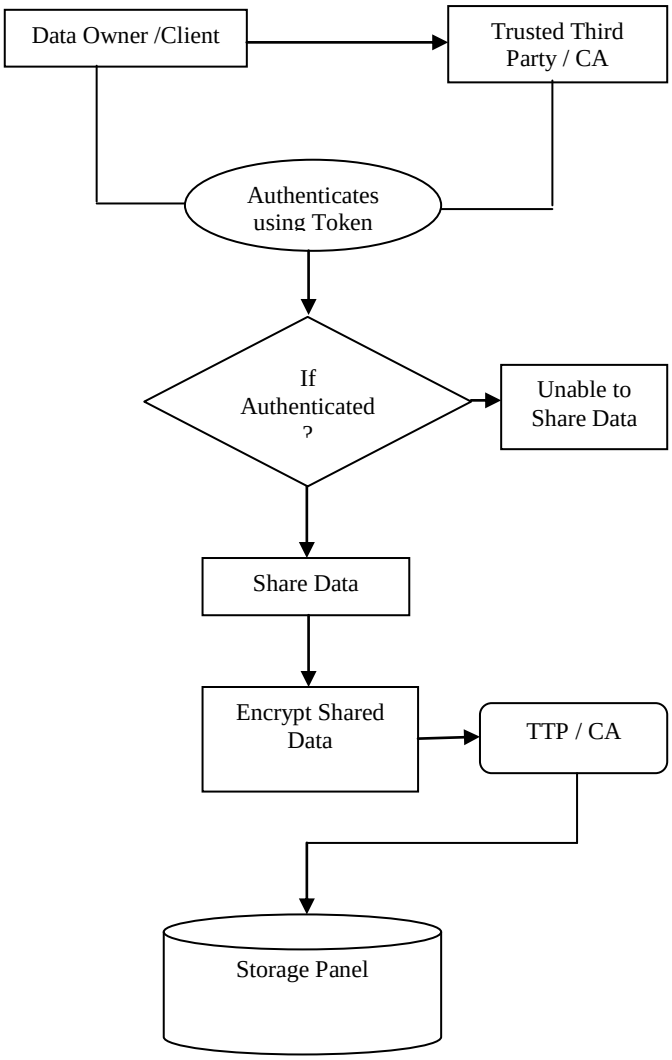

Figure 3: Flow Chart of the Planned Procedure

4. RESULT ANALYSIS

The table shown below is the analysis and comparison of Breakdown Time on the basis of File Size. Here the comparison is done on File Size of 10, 50, 100, 200, 400 MB and hence Breakdown Time is computed. Here the breakdown Time Computed is the total Breakdown time including all the steps involved in Data Deduplication. The Untried results show that the planned practice implemented takes less Time as compared to the existing methodology.

Table 1. Time BreakDown for Different File Sizes

	Time Breakdown (Sec)	
File Size (MB)	**Existing Work**	**Proposed Work**
10	1	0.53
50	1.54	0.78
100	2.23	1.1
200	5.32	3.52
400	10.17	8.35

The table shown below is the analysis and comparison of Cumulative Time on the basis of Various Number of Files. Here the comparison is done on Various Files of 2000, 4000, 6000, 8000, 10000 and hence Cumulative Time is computed. Here the Cumulative Time Computed is the total Cumulative time including all the steps involved in Data Deduplication. The Untried consequences show that the planned procedure implemented takes less Time as compared to the existing methodology.

Table 2. Time BreakDown for Dissimilar Quantity of Stored Files

	Cumulative Time (Sec)	
No. of Files	**Existing Work**	**Proposed Work**
0	0	0
2000	468	448
4000	826	785
6000	1328	1264
8000	1487	1383
10000	1748	1673

The Table shown below is the analysis and comparison of Breakdown Time on the basis of Deduplication Ratio. Here the comparison is done on Deduplication Ratio of 20, 40, 60, 80, 100 % and hence Breakdown Time is computed. Here the breakdown Time Computed is the total Breakdown time including all the steps involved in Data Deduplication. The Untried consequences show that the planned practice implemented takes less Time as compared to the existing methodology.

Table 3: Comparison of Time Breakdown for Dissimilar Deduplication Ratio

	Time (Sec)	
Deduplication Ratio (%)	**Existing Work**	**Proposed Work**
0	3.5	3.2
20	2.8	2.4
40	2.6	2.3
60	1.8	1.6
80	1.5	1.2
100	1.1	0.7

The Figure shown below is the analysis and comparison of Breakdown Time on the basis of File Size. Here the comparison is done on File Size of 10, 50, 100, 200, 400 MB and hence Breakdown Time is computed. Here the breakdown Time Computed is the total Breakdown time including all the steps

involved in Data Deduplication. The Untried results show that the planned practice implemented takes less Time as compared to the existing methodology.

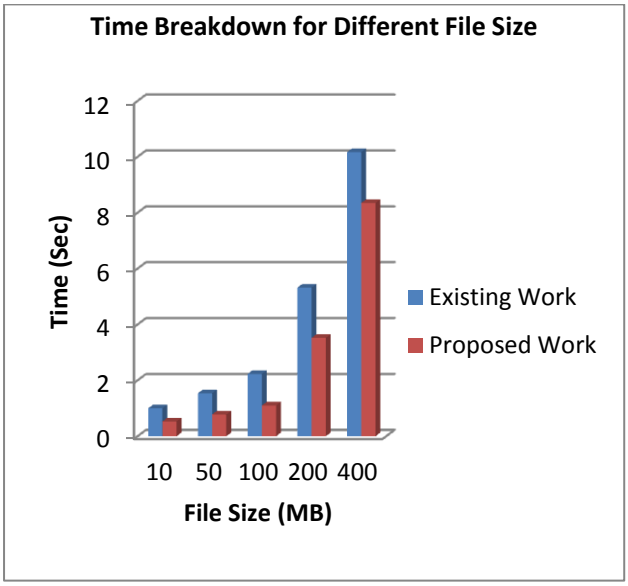

Figure 4.Comparison of Time Breakdown for Different File Sized

The Figure shown below is the analysis and comparison of Cumulative Time on the basis of Various Number of Files. Here the comparison is done on Various Files of 2000, 4000, 6000, 8000, 10000 and hence Cumulative Time is computed. Here the Cumulative Time Computed is the total Cumulative time including all the steps involved in Data Deduplication. The Untried consequences show that the planned procedure implemented takes less Time as compared to the existing methodology.

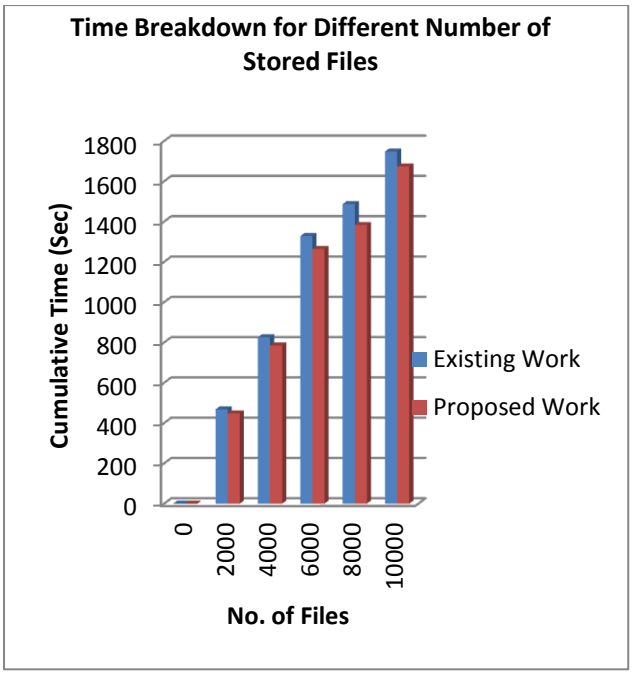

Figure 5.Comparison of Time Breakdown for DissimilarQuantity of Stored Files

The Figure shown below is the analysis and comparison of Breakdown Time on the basis of Deduplication Ratio. Here the comparison is done on DeduplicationRtio of 20, 40, 60, 80, 100 % and hence Breakdown Time is computed. Here the breakdown Time Computed is the total Breakdown time including all the steps involved in Data Deduplication. The Untried consequences show that the planned practice implemented takes less Time as compared to the existing methodology.

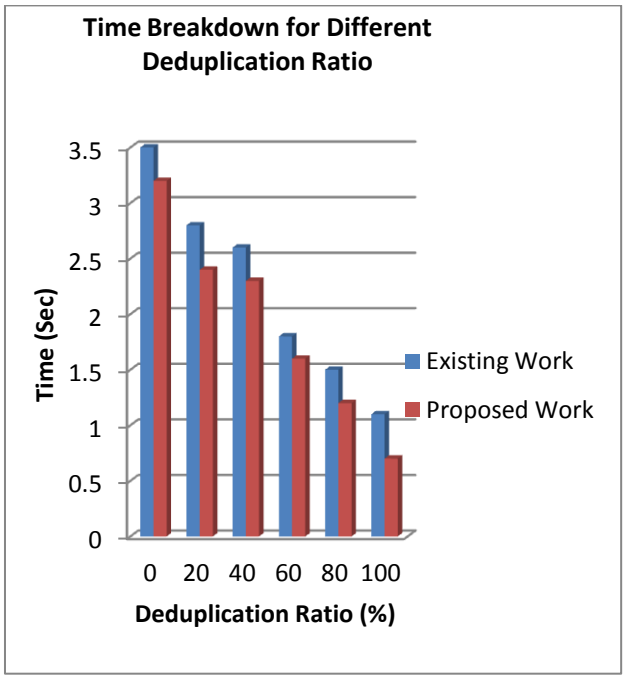

Figure 6.Time Breakdown for Dissimilar Deduplication Ratio

5. CONCLUSION

A new and efficient technique for the Data Deduplication over Hybrid Cloud is proposed using 2-Factor Authentication between Data Owner and Trusted Third Party. Here for 2-Factor Confirmation Perfunctory Based Confirmation is used and Solid Logarithmic Cryptography such as Elliptic Curves are used for the Encryption of Data. The Planned Procedure realized here is an efficient technique in comparison to the prevailing method implemented for Data Deduplication. The Experimental analysis shows the performance of the planned procedure.

6. REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the clouds: A berkeley view of cloud computing," University of California, Berkeley, Tech. Rep. USB-EECS-2009-28, Feb 2009.

[2] D. Russell, "Data deduplication will be even bigger in 2010," Gartner, 2010.

[3] Giuseppe Ateniese, Randal Burns, Reza Curtmola, Joseph Herring, Osama Khan, Lea Kissner, Zachary Peterson, and Dawn Song. 2011. Remote data checking using provable data possession. ACM Trans. Inf. Syst. Secur. 14, 1 (2011), 1–34.

[4] Bo Chen, Reza Curtmola, Giuseppe Ateniese, and Randal Burns. 2010. Remote data checking for network coding-based distributed storage systems, 2010.

[5] JadapalliNandini, RamireddyNavateja Reddy, "Implementation of Hybrid Cloud Approach for Secure

Authorized Deduplication" International Research Journal of Engineering and Technology (IRJET) Volume: 02 Issue: 03, June-2015.

[6] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication", IEEE Transactions on Parallel and Distributed Systems, 2014.

[7] N.B. Kadu, Mr. Amit Tickoo, Mr.Saurabh I. Patil, Mr. Nilesh B. Bhagat , Mr. Ganesh B. Divte, "A Hybrid Cloud Approach for Secure Authorized Deduplication" International Journal of Scientific and Research Publications, Volume 5, Issue 4, April 2015.

[8] GadeSwati,Prof.PrashantKumbharkar, "Cryptosystem For Secure Data Sharing In Cloud Storage"IJIRT Volume 1 Issue 6 2014.

[9] Lukas Malina and Jan Hajny, "Efficient Security Solution for Privacy-Preserving Cloud Services" 6thInternational Conference On Telecommunications Signal Processing Year 2013.