A 2-Factor Secure Authorized Deduplication using Hard Logarithm Cryptography in Clouds

Shweta Biswas
M.E (CSE) Scholar
Department of CSE,
Truba Institute of Engineering & Information Technology
Bhopal, India
shwetabiswascs@rediffmail.com

Asso. Prof. Amit Saxena
Department of CSE
Truba Institute of Engineering & Information Technology
Bhopal, India
amitsaxena@trubainstitute.ac.in

Prof. Manish Manoria
Truba Institute of Engg. & Information Technology College
Bhopal, India
manishmanoria@trubainstitute.ac.in

Abstract— Here in this paper 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 Authentication Token Based Authentication is used and Hard Logarithmic Cryptography such as Elliptic Curves are used for the Encryption of Data. The planned tactic implemented here is an efficient technique in comparison to the existing approach implemented for Data Deduplication. The Experimental analysis shows the performance of the planned method.

Index Terms— Cloud Storage, Data deduplication, Cloud Computing, Elliptic Curves, Token based Authentication.

I. Introduction

Cloud computing is a main region of knowledge that efficiently allows data outsourcing as a service using Internet methods with expandable provisioning and practice 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 paradigm drives the need for cost-efficient storage, specifically for reducing stowage space and system bandwidth overhead, which is straight related to the financial cost savings. In order to reduce the overheads on storage and network, profitable cloud stowage service providers utilize their resources efficiently through data deduplication, which refers to a technique that finds unnecessary data units transversely customers removes duplicate copies of them and provides links to the remaining data instead of storing the copies. By storing and transferring only a solitary copy of superfluous statistics, deduplication provides savings of both stowage freedom and system bandwidth.

Data deduplication fundamentally eliminates duplicate data copies with the intention of make possible a cost-effective storage. It is a type of information compression system (as single-instance data storage) that employed to avoid data redundancy [1]. There is no inconsistency between duplication and dispersed storage scheme because the technique has to identify a common bytes set inside or among files to allow single instance storage of each fragment in each of the server on the beginning of the replication based erasure coding-based, or network coding-based approaches. Data deduplication technique is measured to be one of the most-impactful storage technologies, and it is estimated that the ratio of applying deduplication will increase steadily among the storage overhaul providers [3].

Even though the data deduplication method is measured to be efficient and useful in storage systems, there are several challenging issues of data safety and solitude in the cloud stowage military where the information deduplication method is applied. These issues of security and privacy originate from the following facts:

- In the cloud computing environments, cloud servers are usually outer of the expectation sphere of the data owners (i.e., users). In fact, a wide range of the users are more than willing to put their data outsourcing task to a cloud storage provider.
- Cloud storage military are classically based on multi-tenant architecture, where there is no trust relationship among users. Chasing efficiency in terms of utilizing resources such as the cargo space and the network bandwidth leads to applying client-side data deduplication across multiple (untrusted) users.

In the cloud stowage classification with statistics 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, which act as a cloud server, and an outside adversary, which act as an user, will attempt to break data confidentiality, privacy and integrity on the outsourced data. More concretely, for cloud storage system with deduplication, we are concerned with several security issues that are raised by the adversaries: 1) sacrificing data security for deduplication, 2) in sequence escape during side channel, and 3) unauthorized arbitrary data access. The main aim is enterprise all the network. To set the data back up and disaster recovery applications for reduce

the storage space. We frequently go for de-duplication. Such systems are extensive and are frequently more apposite to customer file backup and harmonization application than better-off stowage abstractions.

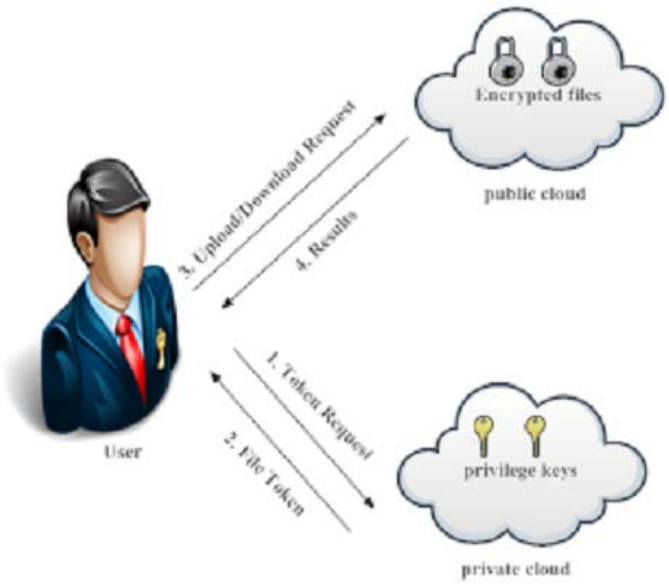

Figure-1: operational of official de-duplication [4]

There are three entity describe in our scheme as shown in figure 1, those are,

- Consumers
- Private mist
- S-CSP in public mist

De-duplication performed [4] by S-CSP by inspection if the filling of two annals are the same and provisions only one of them. Based on the set of benefits the right to use right of a file is described. The precise definition of a benefit diverges transversely applications. Even though data de-duplication brings a lot of reward, security and privacy apprehensions happen as users sensitive data are susceptible to both the insider and outsider attacks. When evaluates the conventional encryption with information repetition. It will present data privacy. In the conventional encryption needs dissimilar users to encrypt data with their own keys. Thus the same copies of different users will show the way to dissimilar cipher texts, making de-duplication impracticable

II. Literature Survey

Jin Li et. al's proposed and implemented a new advance for protected endorsed Deduplication over Hybrid Cloud Approach [5]. Since Data Deduplication provides eliminating of spare copies of repeating Data and is used to diminish the quantity of storeroom Space. Here in this document duplicate-check token of records are generated by the private obscure Server with Private Keys.

DuPLESS [6] is a real deduplication system, which is built on commercial cloud storage services, that provides security against brute-force attacks launched by malicious clients or an untrusted server. In order to achieve the desired goals while satisfying the required security, an Oblivious Pseudo Random Function (OPRF) protocol and message-locked encryption (MLE) [7] were utilized for their construction. OPRF is a randomized protocol between clients and the key server, which ensures that the key server learns nothing about the inputs and the resulting outputs, and the clients learn nothing about the key. MLE is a generalized version of convergent encryption designed.

In this paper author Jin Li,et.al [8] give explanation for deduplication to save from harm the privacy of susceptible data while sustaining deduplication, the convergent encryption method has been offered to encrypt the data earlier than outsourcing. To improved defend data protection; this paper creates the initial effort to properly concentrate on the difficulty of authorized data deduplication. Different from conventional deduplication methods, the discrepancy benefits of customers are additional well thought-out in duplicate check alongside the data itself. They also present quite a lot of novel deduplication buildings secondary official matching checkered in hybrid cloud construction. As a resistant of notion, they realize a example of our proposed authorized duplicate verify method and accomplish test bed experiments. They show that their proposed authorized duplicate confirm method acquires smallest operating cost evaluated to standard process.

N.O. Agrawal et. al's provides a new way of providing Secure Deduplication and Data Security with Efficient and Reliable CEKM [9]. Here in this paper Add security features insider attacker on De-duplication and outsider attacker by using the detection of masquerade activity by risk-averse attackers. The problem of injury of pinched data if we diminution the value of that pinched evidence to the aggressor to achieving effectual and consistent key organization in protected de-duplication

Bhushan Choudhary et. al's provides analysis and Survey of various Data Deduplication Techniques in Cloud [10]. Security by counting differential benefits of clients in the duplicate copy check. Some endangered primitives applied as a part of our harmless de-duplication i.e. Symmetric encryption, Convergent Encryption, Proof of Ownership, Identification Protocol. The security issue is to appraise the effectual operation of cloud band width and disk usage.

Wee Keong et. al's also propose a new and efficient technique for the Private Data Deduplication protocol in Cloud Data Storage [11]. It pompous in the framework of two-party computations using private data de-duplication protocol. Algorithm is best for de-duplication protocol for private data storage. How to define the security of private data de-duplication protocols how to formalize the functionality of private data de-duplication protocols, and how to construct private data de-duplication protocols if exist.

Jorge Blasco et. al's improved Data Deduplication using a Tunable Proof of Ownership using Bloom Filters [12]. Their computational competence in expressions of bandwidth and I/O, for both legal clients and the server. In addition, PoW methods should not entail the server to load the large portion from its back-end storage at each execution of PoW. The main issue is cause root of these risks lies in the precision that proof of ownership only relies on the knowledge of a static, small portion of information.

Sharma Bharat et. al's proposed and implemented a Secure and Authorized statistics Deduplication in mixture Cloud with Public audit [13]. Specific clients are only allowable to achieve the matching check and storage provider for distinct files with the equivalent privileges to access. Issue is that duplicate check do not support differential privileges from convergent encryption even though provide confidentiality.

John R. Douceur et. al's provides Reclaiming gap from Duplicate library in a Serverless scattered File structure.[14]. That method is shows that the duplicate-file coalescing system is scalable, highly efficient, and fault-tolerant. The main issue is enabling the identification, and to retrieve planetary from this accompanying replication to type it accessible for skillful file duplication.

III. Proposed Methodology

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

Step 1: User drive a demand to scheme for challenge worth.
Step 2: Organization take challenge worth.
Step 3: Organization analyze timestamp T_1.
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: Consumer adds broadcast time + t1 to tot_time.
Step 10: consumer take code word.
Step 11: Users decide MD5 hashing purpose on challenge worth + pwd + tot_time.
Step 12: consumer compute MD5 hashing on this information.
Step 13: consumer propel this data to scheme.
Step 14: scheme conventional data D1.
Step 15: scheme compute timestamp T3.
Step 16 scheme determines (challenge worth + password + T3).
Step 17: System determines MD5 hashing on (challenge value + password + 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.

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, $4a$

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

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

Client Sends {F,Tm,Sg,) 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.

FLOW CHART

The figure shown below is the first factor authentication between client and server. Here for the first factor authentication one time private key is used where the server generates a random number for the client and destroys as soon as the authentication gets finish.

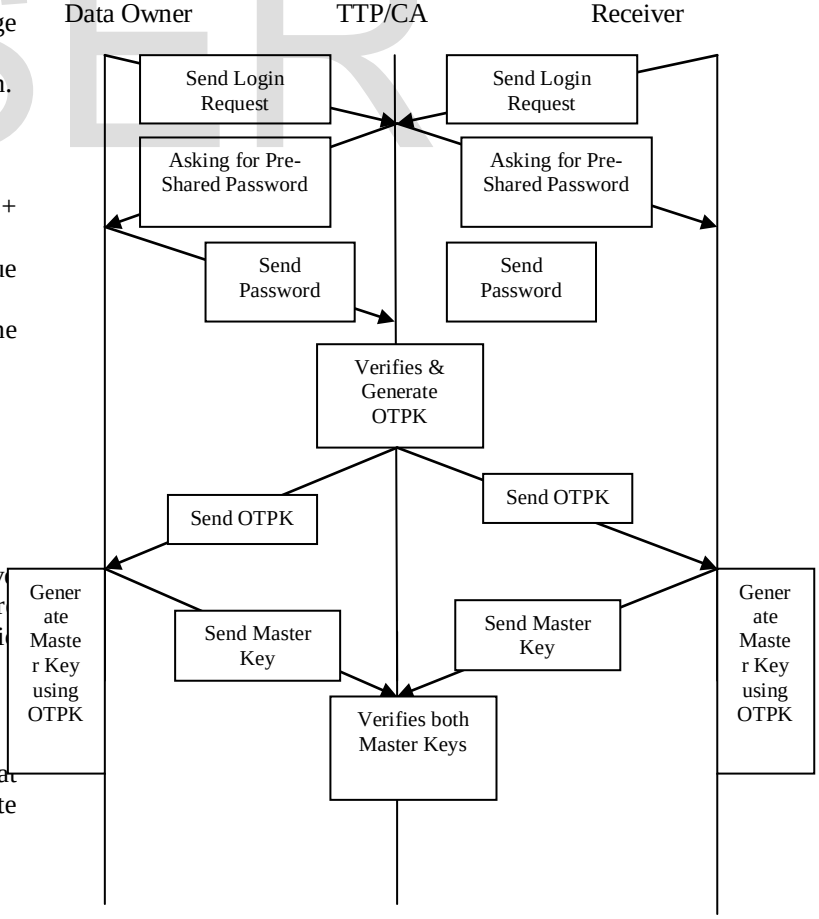

Figure 2. Internal Flow of Authentication between Data Owner & Central Authority

The Figure shown below is the pour chart of the planned method. 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.

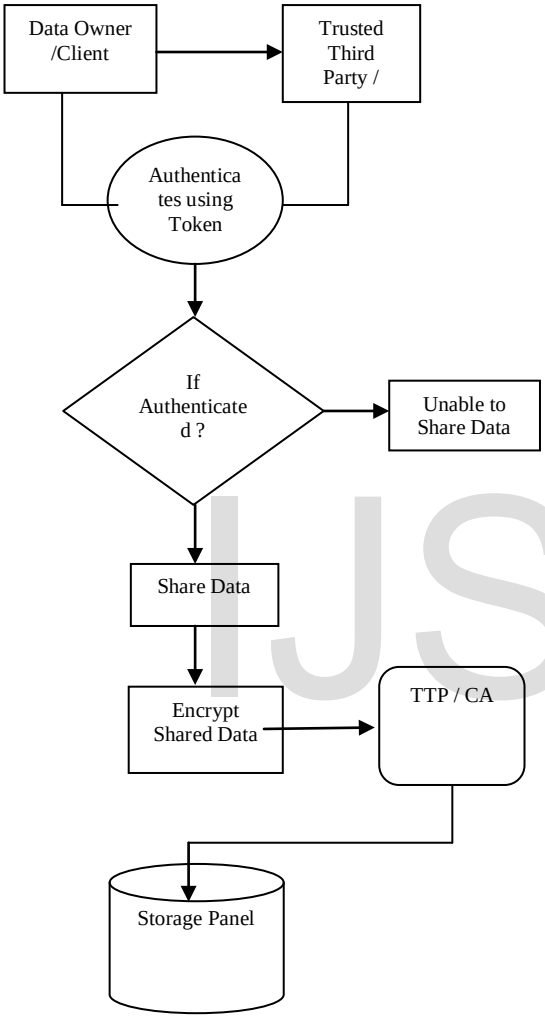

Figure 3. Flow Chart of Proposed Methodology

\

IV. Result Analysis

The table shown below is the examination and contrast 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 consequences show that the planned method implemented takes less Time as compared to the existing methodology.

	Time Breakdown (Sec)	
File Size (MB)	**Jin Et. al's Work**	**Proposed Work**
10	1	0.56
50	1.75	0.82
100	2.45	1.2
200	5.8	3.37
400	10.3	7.82

Table 1. Time BreakDown for Different File Sizes

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 method implemented takes less Time as compared to the existing methodology.

	Cumulative Time (Sec)	
No. of Files	**Jin Et. al's Work**	**Proposed Work**
0	0	0
2000	479	412
4000	846	791
6000	1358	1200
8000	1500	1360
10000	1850	1650

Table 2. Time BreakDown for Different Number 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 Deduplication Rtio 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 method

implemented takes less Time as compared to the existing methodology.

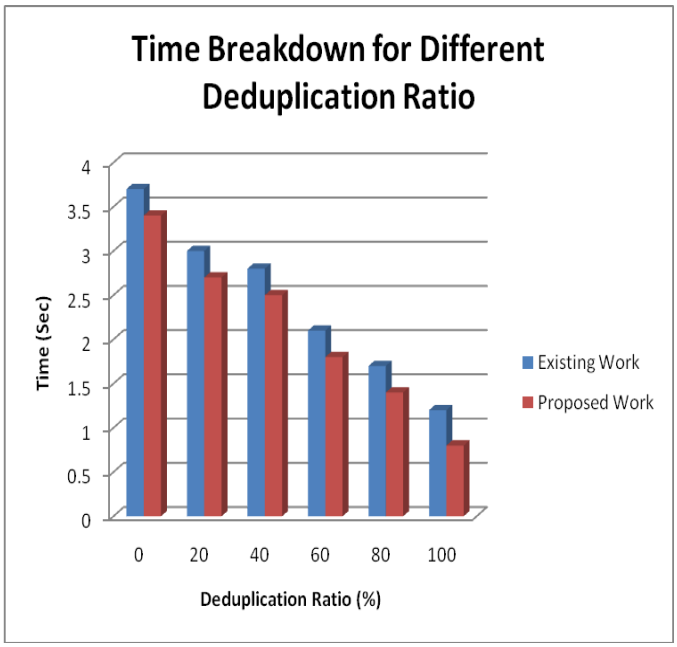

Figure 4. Time Breakdown for Different Deduplication Ratio

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 consequences show that the planned method implemented takes less Time as compared to the existing methodology.

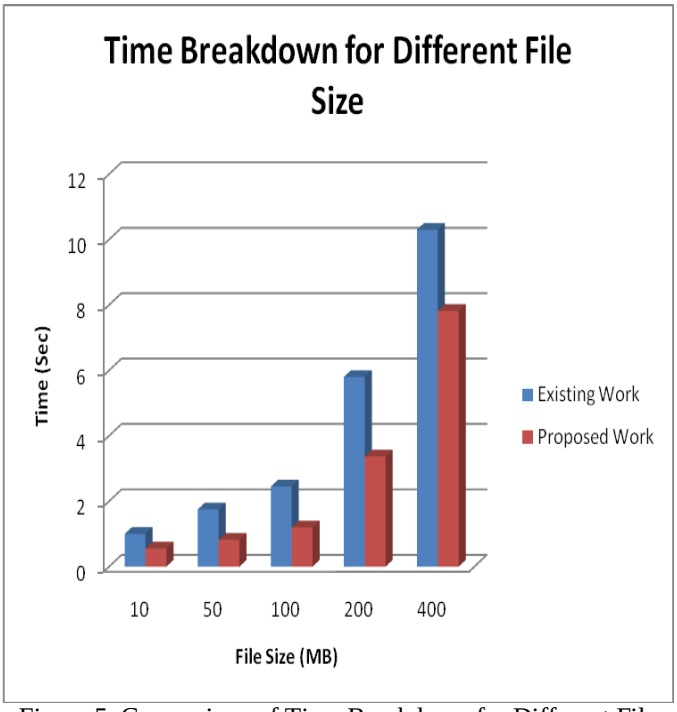

Figure 5. Comparison of Time Breakdown for Different File Sized

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 Rtio 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 method implemented takes less Time as compared to the existing methodology.

	Time (Sec)	
Deduplication Ratio (%)	**Jin Et. al's Work**	**Proposed Work**
0	3.7	3.4
20	3	2.7
40	2.8	2.5
60	2.1	1.8
80	1.7	1.4
100	1.2	0.8

Table 3. Comparison of Time Breakdown for Different Deduplication Ratio

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 method implemented takes less Time as compared to the existing methodology.

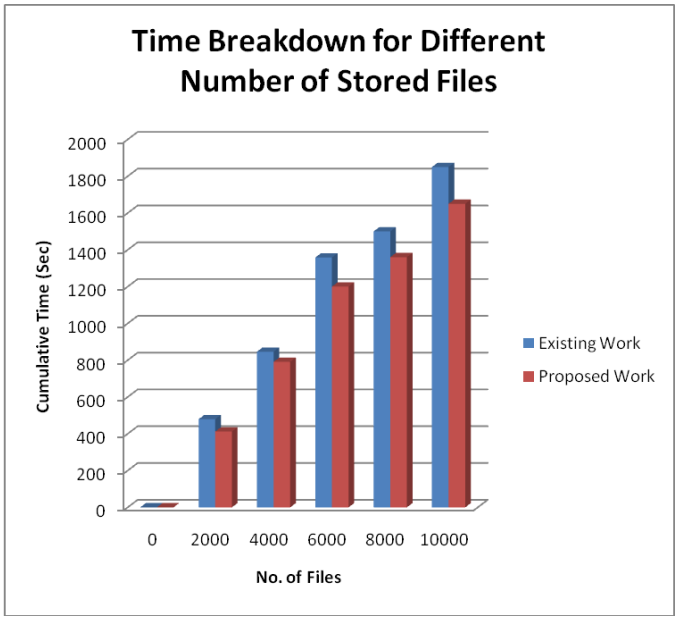

Figure 6. Comparison of Time Breakdown for Different Number of Stored Files

V. Conclusion

The Methodology Applied here for the Data Deduplication using Hard Logarithmic based Cryptography provides efficient Data Deduplication and better Storage Space as compared to the existing methodology for Data Deduplication. The Proposed Methodology provides better Deduplication Ratio a well as Computational Time for Privilege Set Size and Time Breakdown for VM Datasets. The Algorithm also provides Less Time for Different Number of Stored Files.

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 Authentication Token Based Authentication is used and Hard Logarithmic Cryptography such as Elliptic Curves are used for the Encryption of Data. The Proposed Methodology implemented here is an efficient technique in comparison to the existing approach implemented for Data Deduplication. The Experimental analysis shows the performance of the proposed methodology.

References

[1] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, et al., “Above the clouds: A berkeley view of cloud computing," Technical Report UCB/EECS-2009-28, Dept. EECS, UC Berkerely, 2009.

[2] Dutch T. Meyer and William J. Bolosky. 2012. A study of practical deduplication. Trans. Storage 7, 4 (2012), 1–20.

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

[4] Jadapalli Nandini, Ramireddy Navateja Reddy, “Implementation of Hybrid Cloud Approach for Secure Authorized Deduplication” International Research Journal of Engineering and Technology (IRJET) Volume: 02 Issue: 03, June-2015.

[5] 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, IEEE 2014.

[6] M. Bellare, S. Keelveedhi, and T. Ristenpart, “Dupless: Server-aided encryption for deduplicated storage," in Proc. USENIX Security Symposium (SEC'13), pp. 179-194, and 2013.

[7] M. Bellare, S. Keelveedhi, and T. Ristenpart, “Message-locked encryption and secure deduplication," Advances in Cryptology - EUROCRYPT'13, LNCS 7881, pp. 296-312, 2013.

[8] 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 System VOL: PP NO:99 YEAR 2013.

[9] N.O.AGRAWAL1, S.S.KULKARNI, “Secure Deduplication and Data Security with Efficient and Reliable CEKM” International Journal of Application or Innovation in Engineering & Management (IJAIEM), November 2014.

[10] Bhushan Choudhary, Amit Dravid , “A Study on Authorized Deduplication Techniques in Cloud Computing” International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, Issue 12, April 2014.

[11] Wee Keong Ng, Yonggang Wen, Huafei Zhu, “Private Data De-duplication Protocols in Cloud Storage” ACM 978-1-4503-0857-1/12/03, 2011.

[12] Jorge Blasco, Agustin Orfila, “A Tunable Proof of Ownership Scheme for Deduplication Using Bloom Filters” June 18, 2014.

[13] Sharma Bharat, Mandre B.R. “A Secured and Authorized Data Deduplication in Hybrid Cloud with Public Auditing” International Journal of Computer Applications (0975 – 8887) Volume 120 – No.16, June 2015

[14] John R. Douceur, Atul Adya, William J. Bolosky, Dan Simon, Marvin Theimer, “Reclaiming Space from Duplicate Files in a Serverless Distributed File System” July 2002.