A Review over Cloud based Data Deduplication and Related Security Constraints

Laxmikant Malphedwar
Asst. Professor
Computer Engineering
DYPatil school of Engg. Academy, Ambi, Pune

Seema V. Karale
BE Scholar
Computer Engineering
D Y Patil school of Engg. Academy, Ambi,

Sagar A. Suvarnakar
BE Scholar
Computer Engineering
D Y Patil school of Engg. Academy, Ambi, Pune

Harshada Patil
BE Scholar
Computer Engineering
D Y Patil school of Engg. Academy, Ambi, Pune

Mitali Aher
BE Scholar
Computer Engineering
D Y Patil school of Engg. Academy, Ambi, Pune

ABSTRACT
The information change and exchange is specifically or in a roundabout way used by the greater part of the persons. This information should be put away at secure remote area and this can be accomplished by utilizing cloud computing. While concealing stage and usage points of interest. Cloud computing is presently crucial for quick and Remote access of information conjointly some significant issues are likewise should be looked into. One discriminating test of cloud storage administrations is the administration of the steadily expanding volume of information. A usage of better cloud dependably needs a strong stage and foundation fulfillment. A critical issue over Cloud storage is Scalability which is testing in light of element handling in put away information by the client. To make data administration versatile in cloud computing, deduplication has been a surely understood procedure. Despite the fact that information deduplication brings a ton of advantages, security and protection concerns emerge as clients delicate information are defenseless to both inside and pariah assaults on the grounds that conventional encryption, while giving data classifiedness, is inconsistent with information deduplication likewise identical information duplicates of distinctive clients will prompt diverse cipher texts, making deduplication inconceivable. In information deduplication document level check can likewise disregarded with piece level confirmation however it is additionally includes extreme test of similarity. In this paper the most part concentrating on different difficulties over usage on cloud information for most proficient, solid, powerful and quick data access.

Keywords
Cloud computing, Hybrid cloud, Deduplication, FleCS, Constraints, Challenges and Solutions.

1. INTRODUCTION

1.1 Cloud
Cloud storage is the outsider information stockroom as well as utilization for putting away colossal information. Cloud computing [1] and stockpiling arrangements give clients and endeavors different capacities to store and procedure their information in outsider server farms. At the establishment of cloud computing is the more extensive idea of united base and shared administrations. Distinctive sorts of information for various clients are put away on cloud with security [2][3]. It is extremely difficult to deal with the redesigned information which is upgraded by client.Cloud computing, additionally concentrates on boosting the viability of the common assets. This can work for designating assets to clients. Numerous organizations, for example, Amazon, Google, Microsoft are, quicken their paces in creating Cloud Computing frameworks and improving their administrations to accommodate a bigger measure of clients. In any case, security and protection issues exhibit an in number hindrance for clients to adjust into Cloud Computing frameworks.

1.2 The cloud Computing is diverse with following factors

1.2.1 Overseen Functionality
In particular, the administration you utilize is given by another person and oversaw for your sake. In case you're utilizing Google Documents, you don't need to stress over purchasing umpteen licenses for word-handling programming or staying up with the latest. One fundamental standard of cloud computing is that you no more need to stress how the administration you're purchasing is given.

1.2.2 User wish Functionality
Cloud administrations are accessible on-interest and frequently purchased on membership premise. In some cases cloud computing is free or paid-for in different ways. Much the same as power, you can purchase as much or as meager of a cloud computing administration as you need starting with one day then onto the next. It implies you don't need to purchase your own particular massive PC framework and dangers make them stay there doing nothing.

1.2.3 User Privacy Functionality
It comes in two essential flavors, open and private, which are the cloud counterparts of the Internet and Intranets. Email and free administrations like the ones Google gives are the most well-known samples of open mists. Private cloud computing works similarly however you get to the assets you use through secure system associations, much like an Intranet. Organizations, for example, Amazon additionally give you a chance to utilize their freely available cloud to make your own

particular secure private cloud, known as a Virtual Private Cloud (VPC), utilizing virtual private system (VPN) association.

1.3 Points of interest

In the event that your business is offering online Library or repairing Electronics, why get included in the low down of purchasing and keeping up an intricate PC framework? On the off chance that you run a protection office, do you truly need your business specialists squandering time running hostile to infection programming, updating word-processors, or agonizing over hard-commute crashes? As we truly don't need them jumbling your lavish PCs with their own messages, illicitly shared MP3 records when you could leave that obligation to another person? Distributed computing permits you to purchase in just the administrations you need, when you need them, cutting the forthright capital expenses of PCs and peripherals. You evade hardware going outdated and other well-known IT issues like guaranteeing framework security and dependability. You can include additional administrations or take them away immediately as your business needs change. It's truly brisk and simple to add new applications or administrations to your business without sitting tight weeks or months for the new PC and its product.

1.4 Capacity strategies in cloud computing

Distributed storage for the most part uses the virtualization innovation for putting away information. In this segment different capacity procedures are presented[4].

1.4.1 Certain Storage Security to Data in Online

In this plan information is divided in such way that every segment is verifiability secure and does not to be scrambled. These segments are put away on diverse servers on the system which are known just to the client.

1.4.2 Open Auditing with Complete Data Dynamic Support

Confirmation of information respectability at problematic servers is the significant concern in cloud storage with open review capacity trusted element with aptitude and abilities.Information proprietors don't have can be designated as an outer review gathering to get to the danger of outsourced information when required.

1.4.3 Productive Third Party Auditing

Cloud customers spare information in cloud server so that security and additionally information stockpiling rightness is essential concern. The information proprietors having tremendous measure of outsourced information and the take of inspecting the information rightness in a cloud situation can be troublesome and extravagant for information proprietors. To bolster outsider inspecting where client securely assign in trustworthiness checking undertakings to outsider inspectors this plan can practically ensure the concurrent limitation of information lapse implies the ID of making trouble servers.

1.4.4 Powerful and Secure Storage Protocol

Current pattern is clients outsourcing information into administration supplier who have enough range for capacity with lower stockpiling expense. A safe and effective stockpiling convention is recommended that ensures the information stockpiling classifiedness and uprightness. Information and programming procedure convention step executed by cloud clients to add the protection requirement structure to the product and information before exchanging them to the cloud.

1.4.5 Capacity Security of Data

The information is secured in server taking into account client's decision of security strategy with the goal that information is given high secure need assets are being shared crosswise over server inconvenience to information security in cloud. Transmitting information over web is unsafe because of the interloper assaults information encryption assumes an essential part in cloud environment.

1.4.6 Secure and Dependable Storage Service

Capacity administration of licenses shoppers to the information in cloud and also permitted to use the accessible all around qualified application with no stress information stockpiling upkeep. In spite of the fact that cloud supplier's advantages, such an administration surrenders the poise of client's information that acquainted crisp variability dangers with cloud information rightness.

1.4.7 Ideal distributed storage frameworks

Cloud information stockpiling which obliges no exertion is gaining more ubiquity for individual, venture and establishment's information reinforcement and synchronization. At its center, the construction modeling comprises of three parts, an information processor that procedures information before it is sent to the cloud an information verifier that weighs whether the information in the cloud has been messes around with, and a token generator that generator tokens which empowers the distributed storage suppliers to recover fragments of buyer information.

1.4.8 Procedure of access and store little records with capacity

Hadoop conveyed record framework server reasons are inspected for little document inconvenience of local Hadoop circulated document framework. HADOOP disseminated document framework is upheld by extensive measure of little records, for information arrangement revision are not viewed as prefetching system is not likewise exhibited. Hadoop appropriated document framework, in a substantial group, a huge number of servers both host specifically joined capacity and execute client application errand.

1.5 Hybrid Cloud

Hybrid cloud [5] is a distributed computing environment which utilizes a blend of on-premises, private cloud and open cloud administrations with organization between the two stages. By permitting workloads to move in the middle of private and open mists as processing needs and expenses change, cross breed cloud gives organizations more prominent adaptability [6]and more information sending choices.

Open cloud's adaptability and versatility wipes out the requirement for an organization to make atrocious capital consumptions to oblige fleeting spikes popular[1]. General society cloud supplier supplies process assets, and the organization pays for the assets it devours.

1.6 Security and Privacy in Cloud Computing

Five most illustrative security and protection traits classifiedness, honesty, accessibility, responsibility and security resolvability have been distinguished here [1]. Starting with these properties, the connections are exhibited among them, the vulnerabilities that may be misused by aggressors, the danger models, and also existing guard techniques in a cloud situation. Future exploration bearings are already decided for every trait [8]. From the point of view

of cloud storage administrations, information trustworthiness relies on upon the security of operations while away notwithstanding the security of the transferring and downloading sessions. Classifiedness can be accomplished by embracing powerful encryption plans. Then again, the uprightness and denial issues are not took care of well on the present cloud administration stage. Restricted SSL session just ensures one-way uprightness. There is no instrument for the client or administration supplier to check whether the record has been altered in the cloud storage.

1.7 Data Deduplication

Information Deduplication is a strategy for decreasing the measure of storage room an association needs to spare its information. This strategy is for the most part used to minimize the storage room & it spares data transfer capacity. The same record may be spared in a few better places by distinctive clients, or two or more documents that aren't indistinguishable may even now incorporate a great part of the same information. Deduplication [9] wipes out these additional duplicates by sparing only one duplicate of the information and supplanting alternate duplicates with pointers that lead back to the first duplicate. Organizations every now and again utilize Deduplication in reinforcement and catastrophe recuperation applications, yet it can be utilized to free up space in essential stockpiling too.

The Deduplication is of two sorts

Document level Deduplication

In Document-level Deduplication, the same document names get supplanted by the put away duplicate. On the off chance that there leave one document which is as of now put away then that record get supplanted.

Piece level Deduplication

In piece level Deduplication, the substance of the record is checked. It checks by pieces and uproot repetition.

1.7.1 Deduplication Benefits

The essential advantage of information Deduplication is that it decreases the measure of storage that associations need to purchase, which thus diminishes costs. Taking out additional duplicates of information spares cash on direct circle equipment costs, as well as on related expenses, similar to power, cooling, support, floor space, and so forth. Deduplication can likewise diminish the measure of system transfer speed needed for reinforcement forms, and now and again, it can accelerate the reinforcement and recuperation process.

1.7.2 Deduplication versus Compression

Deduplication is once in a while mistaken for compression, another procedure for decreasing stockpiling necessities. While Deduplication takes out excess information, compression utilizes calculations to spare information all the more succinctly. By differentiation, Deduplication just kills additional duplicates of information; none of the first information is lost. Likewise, pressure doesn't dispose of copied information in this way; the capacity framework could at present contain various duplicates of packed documents. Regularly undertakings use Deduplication and compression together with a specific end goal to boost their reserve fund.

2. THE VARIOUS CHALLENGES OVER INTRODUCED DEDUPLICATION

It is not ready to oversee by the cloud director arrange that prompts the system movement and multifaceted nature of equipment. To beat this there are numerous document lumping and information pressure strategies are utilized.

One thought is that the remarkable issues connected with distributed computing security have not been perceived. Another thought is that the particular security necessities for distributed computing have not been all around characterized inside of the group.

One worry is that the clients would prefer not to uncover their information to the cloud administration supplier [7]. Another concern is that the clients are uncertain about the uprightness of the information they get from the cloud. Here, more than customary security components will be needed for information security.

One of the fundamental difficulties that keep end clients from embracing cloud storage administrations is the trepidation of losing information or information defilement.

2.1 One of kind prerequisites for cloud computing information security from a couple of alternate points of view.

1. Storing information into and bringing information from gadgets and machines behind a cloud are basically a novel type of database outsourcing.

2. Once the dataset is put away remotely; a Web program is a standout amongst the most advantageous methodologies that end clients can use to get to their information on remote administrations. In the time of cloud computing, Web security assumes a more essential part than any time in recent memory.

3. With the advancement of fast system innovations and extensive transfer speed associations, more mixed media information are being put away and partook in the internet.

2.2 Information deduplication restrictions

2.2.1 Rendition support

Server-side information deduplication is just relying on specific renditions of capacity chiefs or most recent servers [6]. Henceforth for ideal proficiency when utilizing server-side information deduplication, move up to the upheld check adaptation. Additionally, Client-side information Deduplication is likewise should be redesigned which is critical.

2.2.2 Qualified capacity pools

Information on irregular access storage can't be deduplicated. Just information away pools that are connected with consecutive get to i.e. record can be deduplicated [4]. You must empower document stockpiling pools for information deduplication. Customer records must be sure to an administration class that determines a deduplication-empowered capacity pool.

2.2.3 Encrypted documents

As a security safety measure, you can take one or a greater amount of the accompanying steps:

1. Enable capacity gadget encryption together with customer side information deduplication.
2. Use customer side information deduplication just for hubs that are secure.
3. If you are unverifiable about system security, empower Secure Sockets Layer (SSL).

4. If you don't need certain items to be handled by customer side information deduplication, you can prohibit them on the customer.
5. Use fundamental summon to recognize conceivable security assaults on the server amid customer side information deduplication.

2.2.4 Document size

Just documents that are more than 2 KB are deduplicated. Records that are 2 KB or less are not deduplicated.

2.2.5 Operations that appropriate client side information deduplication

The accompanying operations overshadow customer side information deduplication:

1. LAN information development
2. Sub record reinforcement operations
3. Simultaneous-compose operations
4. Server-started sessions

We can't plan or empower any of those operations amid customer side information deduplication. On the off chance that any of those operations happen amid customer side information deduplication, customer side information deduplication is killed, and a message is issued to the lapse log.

2.3 Issues with document level deduplications and piece level deduplications

Document level deduplication will spare a generally little measure of space on your storage. Piece level deduplication will spare more space on your disk/tape and variable square level deduplication will spare considerably more space on your disk/tape file. Then again, as with some other information stockpiling innovation or method, be exhorted that your mileage will fluctuate contingent upon the measure of duplicated information you have in your document frameworks. Block level is for the most part better; however it does rely on upon situation. Record level has less overhead. Piece level is not that vastly different than customary pressure.

Here there is a difficulty: Buy dedupe or get more capacity at the same cost. Here is the logic:

1. Dedupe is questionable. This means that in one situation you may get a half stockpiling lessening, however in the following you may get 0%. That is only a hazard that you need to take. What's more, it could change after some time contingent upon numerous components. So your vigorously deduped capacity could all of a sudden inflatable on you without much cautioning, improbable, however dedupe is dependably a bet around its value.
2. Dedupe dependably has an execution hit. How much changes yet for the most part it is somewhat high so you from time to time need it on a superior stockpiling framework.
3. Dedupe permitting is by and large really extravagant. The "dedupe expense" is ordinarily so high that you can purchase a lot of extra circles at the same cost.

2.4 Hybrid Cloud Storage Challenges

Information security and agreeability concerns are serving to drive endeavor enthusiasm for hybrid mists as opposed to all-open or all-private cloud arrangements. In this hybrid model, a percentage of the organization's information lived in the private cloud, some in people in general and quite a bit of it in both. On the other hand, the half sortscloud storage methodology has issues. While it does help address consistence concerns, it doesn't eradicate them. There still are consistence issues that farthest point what information can go to people in general cloud, or be in both mists. In the meantime, replicating information to people in general cloud briefly takes an excessive amount of time, and a considerable measure of transfer speed, so the thought of "cloud-blasting" to handle burden spikes keeps running into the truth that duplicating the information expected to nourish new cases in the cloud may take longer than the spike length of time.

Another issue is information synchronization. At the point when changes are made, which is the prime duplicate, and how would you keep open cases from "using so as to go" on the private information or making out.

3. VARIOUS PROPOSED SYSTEMS AS SOLUTIONS TO CHALLENGES

3.1 Live Deduplication Stockpiling Of Virtual Machine Pictures In An Open-Source Cloud

Deduplication is a methodology of abstaining from putting away information obstructs with indistinguishable substance, and has been indicated to successfully lessen the circle space for putting away multi-gigabyte virtual machine (VM) pictures. In any case, it stays testing to send deduplication in a genuine framework, for example, a cloud stage, where VM pictures are frequently embedded and recovered. One proposed system called LiveDFS, a live deduplication record framework that empowers deduplication stockpiling of VM pictures in an open-source cloud that is conveyed under ease merchandise equipment settings with constrained memory foot shaped impressions. LiveDFS has a few unmistakable components, including spatial area, prefetching of metadata, and journaling. LiveDFS is POSIX-consistent and is actualized as a Linux piece space record framework. Contrasted with a common record framework without deduplication, It is demonstrated that LiveDFS can spare no less than 40% of space for putting away VM pictures, while accomplishing sensible execution in importing and recovering VM pictures.

3.2 CABdedupe: A Causality-Based Deduplication Performance Booster for Cloud Backup Services

Existing arrangements that utilize the deduplication innovation for cloud reinforcement benefits just concentrate on expelling repetitive information from transmission amid reinforcement operations to decrease the reinforcement time, while giving careful consideration to the restore time that contend is an imperative perspective and influences the general nature of administration of the cloud reinforcement administrations. A causality based deduplication execution promoter for both cloud reinforcement and restore operations, called CABdedupe [7], which catches the causal relationship among sequential adaptations of datasets that are prepared in different reinforcements/restores, to expel the unmodified information from transmission amid reinforcement operations

as well as restore operations, along these lines to enhance both the reinforcement and restore exhibitions. CABdedupe is a middleware that is orthogonal to and can be incorporated into any current reinforcement framework. In the broad examinations, CABdedupe is incorporated into two current reinforcement frameworks and encourage true datasets, demonstrate that both the reinforcement time and restore time are altogether diminished.

3.3 Private Information Deduplication Conventions In Distributed Storage

A deduplication procedure for private information stockpiling is presented and formalized. Naturally, a private information deduplication convention permits a customer who holds private information demonstrates to a server who holds a rundown string of the information that he/she is the proprietor of that information without uncovering additional data to the server. The security of private information deduplication conventions is formalized in the recreation based structure in the setting of two-gathering calculations [9]. A development of private deduplication conventions in view of the standard cryptographic suspicions is then introduced and broke down. The proposed private information deduplication convention is provably secure expecting that the hidden hash capacity is crash strong, the discrete logarithm is hard and the deletion coding calculation can eradication up to α-portion of the bits in the vicinity of pernicious foes in the vicinity of malevolent enemies. To the best insight this is the first deduplication convention for private information stockpiling.

3.4 AA-Dedupe: An Application-Aware Source Deduplication Approach for Cloud Backup Services in the Personal Computing Environment

We reviewed AA-Dedupe [16], an application-mindful source deduplication plan, to altogether lessen the computational overhead, build the deduplication throughput and enhance the information exchange proficiency. The AA-Dedupe methodology is inspired by key perceptions of the considerable contrasts among applications in information excess and deduplication qualities, and along these lines is in light of an application-mindful record structure that successfully abuses this application mindfulness. The exploratory assessments, in light of an AA-Dedupe model usage, demonstrate that the plan can enhance deduplication proficiency over the condition of-craftsmanship source-deduplication techniques by a variable of 2-7, subsequent in abbreviated reinforcement window, expanded force effectiveness and diminished expense for cloud reinforcement administrations.

3.5 Post-Process Deduplication

With post-process deduplication, new information is initially put away on the capacity gadget and after that a procedure at a later time will break down the information searching for duplication [16]. The advantage is that there is no compelling reason to sit tight for the hash counts and lookup to be finished before putting away the information along these lines guaranteeing that store execution is not debased. Usage offering arrangement based operation can give clients the capacity to concede streamlining on dynamic records, or to process documents taking into account sort and area. One potential downside is that you might superfluously store copy information for a brief while which is an issue if the capacity framework is close full limit.

3.6 In-Line Deduplication

This is the procedure where the deduplication hash figuring's are made on the objective gadget as the information enters the gadget continuously. On the off chance that the gadget spots a piece that it officially put away on the framework it doesn't store the new square, just references to the current piece. The advantage of in-line deduplication [5] over post-process deduplication is that it requires less capacity as information is not copied. On the negative side, it is much of the time contended that in light of the fact that hash figuring's and lookups takes so long, it can imply that the information ingestion can be slower consequently diminishing the reinforcement throughput of the gadget. Be that as it may, certain sellers with in-line deduplication have shown hardware with comparative execution to their post-process deduplication partners. Post-process and in-line deduplication strategies are frequently vigorously faced off regarding.

3.7 Source versus Target Deduplication

Another approach to consider information deduplication is by where it happens. At the point when the deduplication happens near where information is made, it is regularly alluded to as source deduplication. When it happens close where the information is put away, it is ordinarily called target deduplication. Source deduplication guarantees that information on the information source is deduplicated. The deduplication procedure is straightforward to the clients and reinforcement applications. Going down a deduplicated record framework will regularly bring about duplication to happen bringing about the reinforcements being greater than the source information. Target deduplication is the procedure of uprooting copies of information in the auxiliary store. For the most part this will be a reinforcement store, for example, an information storehouse or a virtual tape library.

3.8 Encryption of Files

This is utilization the regular mystery key k to scramble and in addition unscramble information. This will use to change over the plain content to figure content and again figure content to plain content [8]. Here three essential capacities are utilized:

KeyGenSE: k is the key era calculation that produces κ utilizing security parameter 1.

EncSE (k, M): C is the symmetric encryption calculation that takes the mystery κ and message M and after that yields the ciphertext C;

DecSE (k, C): M is the symmetric decoding calculation that takes the mystery κ and cipher text C and afterward yields the first message M.

Private encryption

A client gets a united key from every unique information and scrambles the information duplicate with the concurrent key. The client additionally infers a tag for the information duplicate, such that the tag will be utilized to recognize copies.

3.9 ID Protocol

A recognizable proof convention can be portrayed with two stages: Proof and Verify. In the phase of Proof, a client C can show his character to a verifier by performing some ID verification identified with his personality. The data of the prover/client is his private key skU that is delicate data.

3.10 Hybrid Architecture for Secure Deduplication

At an abnormal state, setting of hobby is a venture system, comprising of a gathering of subsidiary customers (for instance, workers of an organization) who will utilize the S-CSP and store information with deduplication procedure. The S-CSP performs deduplication by checking if the substances of two records are the same and stores stand out of them [7]. Every benefit is spoken into the type of a short message called token. Every record is connected with some document tokens, which indicate the tag with determined benefits. Part of the private cloud server will be clarified in the paper. Piece level deduplication can be effortlessly derived from record level deduplication. In particular, to transfer a document, a client first performs the record level copy check. If the record is a copy, then every one of its pieces must be copies also.

3.11 Convergent encryption

Convergent encryption [3] is utilized to encode and unscramble file.User can get the united key from every unique information copy, then utilizing that key encrypt information file. Also client infers tag for information duplicate to check copy data. If tags are same then both records are same. Both joined key and tag are autonomously infers. United encryption, otherwise called substance hash keying, is utilized to create indistinguishable cipher text from indistinguishable plaintext documents. The least complex usage of joined encryption can be characterized as: Alice gets the encryption key from her document F such that K = H(F), where H is a cryptographic hash function.Convergent encryption plan can be characterized with four primitive capacities:

1. KeyGenCE(M) - > K is the key era calculation that maps an information duplicate M to a merged key K;
2. EncCE(K, M) - >C is the symmetric encryption calculation that takes both the merged key K and the information duplicate M as inputs and afterward yields a ciphertext C;
3. DecCE (K, C) - >M is the unscrambling calculation that takes both the ciphertext C and the concurrent key K as inputs and afterward yields the first information duplicate M ; and
4. TagGen(M) - > T (M) is the label era calculation that maps the first information duplicate M and yields a label T (M).

4. CONCLUSION

This paper makes the essential choose to formally address the matter of affirmed data deduplication. Very surprising from antiquated deduplication frameworks, the differential benefits of client's are extra pondered in copy check other than the data itself. Numerous new deduplication developments supporting endorsed copy sign up a cross breed cloud plan. Security investigation shows that this topic is secure as far as the definitions laid out in the arranged security model. As an image of thought, actualizing a picture of an arranged endorsed copy check subject and behavior work environment tests exploitation of the picture. Arranged sanction copy check subject brings about littlest overhead contrasted with conventional operations is demonstrated here. A model for provable data possession is introduced, in which it is desirable to minimize the file block accesses, the computation on the server, and the client–server communication. Key components are the support for spot checking, which ensures that the schemes remain light weight, and the homomorphic verifiable tags, which allow to verify data possession without having access to the actual data file. The notion of robust auditing, which integrates remote data checking with forward error-correcting codes to mitigate arbitrarily small file corruptions and propose a generic transformation for adding robustness is also reviewed.

5. ACKNOWLEDGMENTS

We are thankful to all our faculties of Computer Engineering department who really encouraged and guided us well to prepare this paper.

6. REFERENCES

[1] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. ofUSENIX LISA, 2010.

[2] Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang and Yang Xiang, Mohammad Mehedi Hassan and AbdulhameedAlelaiwi Member, IEEE .2015 “Secure Distributed Deduplication Systems with Improved Reliability.”

[3] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, “Secure deduplication with efficient and reliable convergent key management,” in IEEE Transactions on Parallel and Distributed Systems, 2014, pp. vol.25(6).

[4] Mr. S. Venkatraman, Dr. D. Srinath, Mr. C. Prasanth A Secure Authorized NLP Based Duplicate Check Scheme For Data Deduplication ISSN 2348 – 9928 IJAICT Volume 1, Issue 12, April 2015

[5] Yinjin Fu, Hong Jian, NongXiao,LeiTian “AA-Dedupe: An Application-Aware Source Deduplication Approach for Cloud Backup Services in the Personal Computing Environment” Cluster Computing (CLUSTER), 2011 IEEE International Conference

[6] Tharam Dillon, Chen Wu and Elizabeth Chang “Cloud Computing: Issues and Challenges” 2010 24th IEEE International Conference on Advanced Information Networking and Applications.

[7] Yujuan Tan; Hong Jiang; Dan Feng; Lei Tian; Zhichao Yan, "CABdedupe: A Causality-Based Deduplication Performance Booster for Cloud Backup Services," in *Parallel & Distributed Processing Symposium (IPDPS), 2011 IEEE International* , vol., no., pp.1266-1277, 16-20 May 2011

[8] Iuon –Chang Lin, Po-chingChien ,”Data Deduplication Scheme for Cloud Storage” International Journal of Computer and Control(IJ3C),Vol1,No.2(2012)

[9] TanupriyaChaudhari , Himanshushrivastav, VasudhaVashisht, “A Secure Decentralized Cloud Computing Environment over Peer to Peer”,IJCSMC,April,2013

[10] T.Sivashakthi, Dr. N Prabakaran “A Survey on Storage Techniques in Cloud Computing”International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 12, December 2013.

[11] Mr. S. Venkatraman, Dr. D. Srinath, Mr. C. Prasanth A Secure Authorized NLP Based Duplicate Check Scheme For Data Deduplication ISSN 2348 – 9928 IJAICT Volume 1, Issue 12, April 2015

[12] Min Fu, Dan Feng, Yu Hua, , Xubin He, Zuoning Chen, Jingning Liu, Wen Xia, Fangting Huang, and Qing Liu

"Reducing Fragmentation for In-line Deduplication Backup Storage via Exploiting Backup Historyand Cache Knowledge." IEEE Transactions Volume: PP 2015

[13] BogaVenkatesh,Anamika Sharma,Gaurav Desai,DadaramJadhavSecure Authorised Deduplication by Using Hybrid Cloud Approach in International Journal of Innovative Research in Advanced Engineering (IJIRAE)

[14] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical Report, 2013.

[15] A Hybrid Cloud Approach for Secure Authorized Deduplication Jin Li; Yan Kit Li; Xiaofeng Chen; Lee, P.P.C.; Wenjing Lou Parallel and Distributed Systems, IEEE Transactions on Year: 2015, Volume: 26, Issue: 5

[16] RajashreeShivshankarWalunj, Deepali Anil Lande,NilamShrikrushnaPansare "Secured Authorized Deduplication Based Hybrid Cloud" The International Journal Of Engineering And Science (IJES) Volume 3

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