

Secure Sharing & Data Deduplication over Cloud: A Survey

Shivani Sengar

Computer Science

Adina institute of science & technology

Sagar ,MP

shivanisengar18@gmail.com

Asst. prof. Ruchika Mishra

Computer Science

Adina institute of science & technology

Sagar ,MP

Ruchikamishra.adina@gmail.com

Abstract—Over the precedent few years, cloud computing has develop into more and more admired. It brings revolutionary innovation with regards to cost, resource management and utilization. Cloud computing offer nearly unlimited resources, highly reliable on-demand services with minimal infrastructure and operational cost and storage system. In the cloud storage system with data de-duplication, un-trusted entities including a cloud server and users may cause security threats to the storage system. These services offers to end-users quite than a creation by allocation possessions, software and other information, hence economic benefit and data de-duplication is the key for Cloud in terms of capital and operational expenditure.

Index Terms—Cloud Computing, Data De-duplication, style, styling, insert.

I. INTRODUCTION

Cloud computing (CC) is a promising and emerging technology for the next generation of IT applications. The difficulty and problems in the direction of the quick development of cloud computing are data refuge and solitude issues. Cloud computing is a capable tool that cost-effectively allows data outsourcing as an examination using Internet tools with elastic provisioning and usage-based pricing [1]. Cloud computing provides a low-cost, scalable, location independent infrastructure for data management and storage that is available anyplace and anytime over the Internet application on cloud storage services such as Drop-Box, Mozy and Memopal are increasing recognition.

Cloud computing has raised the delivery of IT services to a novel stage that carries the console of conventional utilities such as water and electricity to its users. The advantages of Cloud computing, such as cost effectiveness, scalability, and ease of management, encourage more and more friendship and service providers to become accustomed it and present their explanations passing through Cloud computing models. According to a modern review of IT decision makers of huge companies, 68% of the respondents expect that by 2014, more than 50% of their company's IT services will be migrated to Cloud platforms [2]. Cloud computing has become a scalable service consumption and delivery platform. Figure 1.1 shows

the system architecture in cloud computing. In a cloud environment, the cloud provider grips a huge number of distributed examines (e.g. databases, servers, Web services, etc.), which can be offered to expensive for increasing a range of cloud applications. Expensive of cloud applications can prefer from an extensive collection of distributed services when creating cloud applications. These examinations are frequently bring into play distantly through communication links and are enthusiastically put together into the applications. The cloud application designers are located in different geographic and network environments. Since the users invoke services via different communication links, the quality of services they observed are diverse.

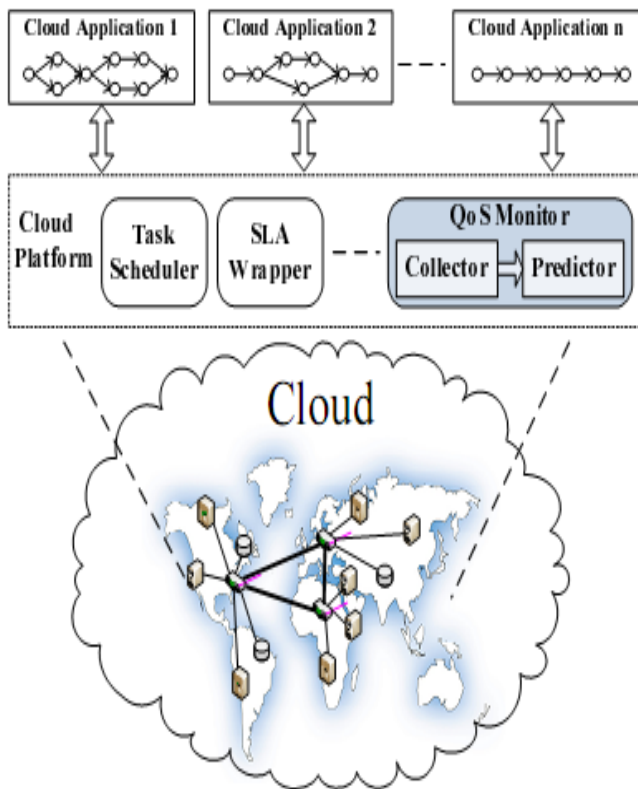


Figure 1.1: System Architecture of cloud computing.

The rapid adoption of Cloud services is going together with ever-increasing quantities of data stored at remote servers, so methods for economy disk space and network bandwidth are required. A central up and coming idea in this circumstance is deduplication where the server accumulates only a single copy of each categorizer, despite of how many clients inquired to accumulate that file. All clients that store the file just employ links to the single copy of the file accumulated at the server. Additionally, if the server previously has a copy of the file then clients do not yet require to upload it once more to the server, thus economy bandwidth with storage space i.e. client-side deduplication.

In a classic storage organization with deduplication a client first sends to the attendant only a hash of the sleeve and the server checks if that hash charge previously subsists in its database. If the hash is not in the database then the server inquires for the complete file or else in observation of the reality that the file already subsists at the server feasible uploaded by an important person in addition, it informs the patron that there is no require to send the file itself. Moreover way the server marks the client as a possessor of that file and from that position on the client can request to return the file.

The client-side deduplication commences novel safety measures difficulties. For example, a server telling a client that it require not send the file make known's that some other client has the accurate identical file which could be responsive in order. A malicious client can utilize this information to ensure

whether definite files were uploaded by other customers, or still run a brute force attack which recognizes the substances of assured meadows in files owned by other consumers, by attempting to upload multiple alternatives of the same file which have unusual importance's for that area.

II. THEORETICAL BACKGROUND

Fundamentally, a cloud storage scheme can be measured to be a network of scattered statistics centers which characteristically employs cloud computing technologies like virtualization and presents some type of crossing point for storing cloud data or important information. To improve the user-friendliness of the cloud data it perhaps without cause accumulated at contradictory arrangements. On the entire, all of this is not proficient to be distinguishing to the cloud user. This cloud storage scheme needs secure techniques of preserving significant data with the intention of stop unrecoverable data loss at the same time as frequently care up with growing requires for hold. It is essential to frequently make additional copies of the information so as to be proficient to reinstate it to a previous description if require be. These copies of cloud additional increase the order for cloud storage space on cloud server. Extra constraints happen from the selection of devices utilized to right to use the data at the same time. Private and business clientele command an uncomplicated method to coordinate and right of entry their data independent of both device and position. The software provided that these characteristics must also be modified to require of the entity with no scientific environment.

In Cloud Computing, information owners may contribute to their outsourced data with a huge number of customers, who force want to only recover certain definite data files they are significances in during a given conference. One of the most accepted techniques to do so is all the way through keyword-based search. Such keyword investigate technique permits customers to discriminatingly recover files of interest and has been extensively useful in plaintext search circumstances. Unfortunately, data encryption which confines customer's skill to present keyword search and additional orders the protection of keyword privacy creates the conventional plaintext search techniques stop working for encrypted cloud data. Although conventional searchable encryption methods permit a customer to strongly search over encrypted data all the way through keywords without first decrypting it, these methods sustain only predictable Boolean keyword search without confining any significance of the files in the search result. Ranked investigate get betters system usability by standard matching files in a ranked order considering to definite significance criteria (e.g., keyword frequency) as directly outsourcing significance achieves will drops a lot of sensitive information next to the keyword privacy. On the other hand, earlier deduplication schemes cannot hold up discrepancy agreement duplicate make certain which is significant in many functions. In such an authorized deduplication scheme, each consumer is concerned a set of

opportunities during arrangement initialization. Each file uploaded to the cloud is also cleared by a set of opportunities to indicate which kind of consumers is permitted to complete the duplicate check and right of entry the files. Before presenting his duplicate check request for a file the client requires taking this file and his/her exacting benefits as inputs. The customer is capable to discover a duplicate for this file if and only if there is a copy of this file and a equivalence opportunity accumulated in cloud. To make information administration scalable in cloud computing, deduplication [3] has been a distinguished method and has pull towards you more and more consideration in recent times. Data deduplication is a dedicated data compression method for eradicating duplicate copies of duplicating data in storage space.

III. CLOUD PRIVACY AND REMOTE STORAGES

Cloud computing is an evolving concept more than ever with the recent advancement of emerging technologies. A number of investigators have tried to typify cloud computing however there are no agreed upon definition. For the purpose of this study we will utilize the National Institute of Standards and Technology (NIST) definition of cloud computing which is as follows: "Cloud computing is a representation for facilitating well-positioned, on-demand network access to a distributed collection of configurable computing stores that can be quickly prerequisites and make public with smallest administration attempt or service provider communication. Privacy issues regarding risks within cloud environments are unique for various stakeholders [4]. Privacy risks for individual users of cloud service exist such as exposure of personal information. These privacy issues are compounded when individuals are forced to give personal information against their will, or in a way in which they feel uncomfortable. Typically in these situations the organizations that are collecting the information fail to provide a legitimate reason to the customer as to why they need to collect their information [5].

For organizations using cloud services there is a privacy risk associated with non-compliance with internal privacy policies or external privacy rules that cloud consequence in failure of standing and reliability with their customers. For cloud service providers the privacy risk is with the exposure of sensitive information being stored on their servers, non-compliance, legal liability, loss of reputation, and customer's trust. As a means to reduce cloud privacy risks, [6] recommend that organizations use a combination of privacy policies and contractual terms to create accountability in the form of visible, enforceable promises to dependable data handling conditions. Transparency with respect to information handling practices allows individuals to be informed about how their data is handled within the cloud and defines the responsibility of people and the organization behavior their personal information. Establishing accountability with privacy performs in the cloud assists to make sure agreement with cloud rules and construct confidence.

Remote storage is the method to extend users' disk space without adding more hard disks on local computers. The remote storages usually are services running on other machines (servers) which automatically copies files from user's local disks. In some cases, after copying, the files physical size on user disks are reduced to minimum (a few KBs) while the logical sizes displayed remain the same. Similarly, cloud remote storages are remote storages which use cloud services. Cloud services copy files from user's machine and store them in logical pools. The physical storages can be distributed, replicated among multiple servers or locations. Cloud services providers are hosting companies who own these physical structures. They also provide services, applications to interact with user's devices and logical structures. Cloud remote storages allow users accessing their stored data from other supported devices as long as they have internet connection. They also enable users to share data with others. Among the most popular cloud remote storages are OneDrive such as SkyDrive, Dropbox and Google Drive. Since the encryption process is deterministic and is obtained from the data content, identical data copies will generate the same convergent key and consequently the identical ciphertext. To avoid unauthorized right to use, a protected proof of rights protocol [7] is also required to provide the proof that the client certainly be in possession of the identical file when a duplicate is originated. After the proof, following customers with the identical file will be making available a pointer from the server exclusive of requiring uploading the similar file. A consumer can download the encoded file or data with the indicator from the server, which can only be decrypted by the matching information owners with their convergent keys. Consequently, convergent encryption permits the cloud to execute deduplication on the ciphertexts and the proof of rights avoids the unauthorized consumer to right of entry the file.

However, since files are stored and distributed over the internet, security problems emerge. One major problem is data exfiltration. By uploading sensitive information to cloud storages, users have to put their trust on the providers to protect those data. Nonetheless, there are many situations that can affect this trust. Cloud providers can peek into data or transfer data to other parties in some circumstance. Cloud servers can be hacked, resulting in exposing user's sensitive data. It also can be the consequence of poor security design decision coming from the provider architects.

IV. DATA DEDUPLICATION

Data deduplication is the procedure of recognizing and removing unnecessary copies of repeating data in the cloud or storage. This technique is currently widely used to improve the storage utilization. Data deduplication can be done at file level and block level. File level deduplication identifies and removes same files. This method is also called as Single instance storage. Once the file is stored other same copy of file point to this pointer. Block level detects superfluous information within and crossways documentation. This method is also called as

sub file deduplication For the most part in cloud remote storage context, deduplication fulfills one of the most common requirement as reducing the overall costs of providing the same service that client could do themselves in their own data centers. Not only the process utilizes storage hardware costs it also decreases backup and recovery costs while improving network efficiency. Cloud services like Dropbox employs deduplication across multiple user client, as the user base grows, service cost per user decreases subsequently.

From end-users' point of view, one of the most notable advantage of deduplication is the storage service cost. Basically, they pay less for the same amount of storage. For example, Dropbox offers 1TB storage for only \$9.99 while an external hard drive would cost more than a few hundred dollars for the same capacity. Another noteworthy gain is the upload time/bandwidth for duplicated files. Large files such as .iso or movies could be uploaded almost instantly if they are duplicated, providing significant boost to user experience. In general, deduplication process involves 4 steps:

1. Split data into small units, (ex: files, blocks).
2. Calculate unique hash value for each unit.
3. Detect if there is any file on the storage having the same hash value.
4. Put reference to the duplicated file.

Among deduplication schemes, the first and the second steps vary the most. Data can be compared at file level where the scheme generates hashes of files and compares. Meanwhile, block/byte level units allow deduplication in a finer grained manner with the cost of more operations. Hashing mechanisms are also varying among schemes but the universal idea is to produce a unique, mathematical representation of unit that can be located and compared. Data Deduplication is a method to prevent duplication of repeating data. During deduplication processes, unique chunks of data such as files or block of bytes are analyzed followed by discarding the duplication. The result is only one instance of repeat data are stored/transferred hence greatly improve storage utilization and transfer time/bandwidth. Initially, deduplication is heavily used on duplication prone situations such as backups or virtual desktops [8].

V. DATA REPLICATION AND STORAGE ON CLOUD COMPUTING

A Data Grid is a geographically-distributed teamwork in which all participants necessitate admission to the datasets yield within the association. Replication of the datasets is consequently a key requirement to ensure scalability of the cooperation, dependability of data access and to reservation bandwidth. Replication is constrained by the size of storage existing at altered positions inside the Data Grid and the bandwidth amongst these sites. A imitation organization arrangement therefore confirms admittance to the necessitated data while management the essential storage.

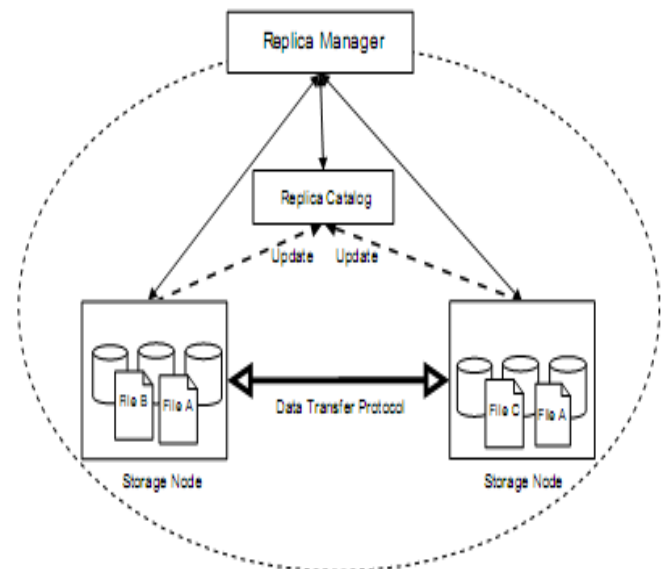


Figure 1.2: Replica Management Architecture.

A replica management system, shown in Figure 1.2, consists of cargo space nodes which are connected to each other through high-performance data transport protocols. The replica manager uninterrupted the formation and supervision of reproductions allowing to the requests of the customers and the accessibility of storage, and a collection or a directory keeps pathway of the duplications and their sites. The collection can be demanded by presentations to determine the number and the positions of existing duplications of a specific dataset. In some systems, the manager and the directory are combined into one unit. Client-side software usually consists of a library that can be incorporated into applications and a set of commands or GUI utilities that are built on peak of the libraries. The client libraries permit querying of the catalog to determine datasets and to apply for replication of a exacting dataset.

VI. SECURITY ISSUES ON CLOUD STORAGE WITH DATA DEDUPLICATION

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

- In the cloud computing environments, cloud servers are usually outside of the trust domain 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 services are typically based on multi-tenant architecture, where there is no trust relationship among users. Chasing efficiency in terms of utilizing resources such as the storage space and the network bandwidth leads to applying client-side data deduplication across multiple (untrusted) users.

In the cloud storage system 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, privacy and integrity on the outsourced data. More concretely, for cloud cargo space arrangement with deduplication, we are concerned with several security issues that are raised by the adversaries: 1) sacrificing data security for deduplication, 2) information leakage through side channel, and 3) unauthorized arbitrary data access.

Sacrificing Data Confidentiality for Deduplication: Cloud storage service providers that are using deduplication, however, are typically reluctant to apply encryption on the stored information, since encrypting on information impedes executing information deduplication [9][11]. They may be unlikely to stop using deduplication due to the high cost savings offered by the technique. This eventually incurs the loss of confidentiality for the data stored on the cloud storage.

Information Leakage through Side Channel: In a storage system using client-side data deduplication, losing confidentiality of the outsourced data is not the only security problem. As shown in [9], client-side deduplication commonly incurs a side channel through which a malicious user (i.e., an adversary) may get sensitive information about the other user's data. The side channel is caused by two inherent properties; 1) data transmission over the network is visible to an adversary, and 2) a small-sized hashed value of a file is used to settle on the continuation of the same folder on the obscure server.

Using the side channel, an adversary can effortlessly recognize the continuation of a file by uploading the file and monitoring its network activity. If the whole file is not transmitted over the network, the adversary learns that the file already exists on the server. Furthermore, the adversary is even able to learn the content of the file by mounting an online-guessing attack. That is, an adversary trying to figure out the content of the file will build a dictionary that consists of guessed versions of that file and repeat uploading each guess to the server until finding out that a deduplication event occurs.

Unauthorized Arbitrary Data Access: Besides an in sequence escape from side to side the side conduit, a cloud storage system that applies the client-side deduplication technique is also vulnerable to another type of attack [10]. This new security threat originates from the fact that client-side deduplication systems typically adopt a hashing based strategy, in which a small-sized hash value is calculated for each file (or block) and is used to find photocopy of the same file. In such a storage system, by accepting the hash value for the file, the cloud server allows anyone who possesses the hash value to download the entire file.

An adversary, who does not have a whole file except its hash value, will exploit this vulnerability to get the ownership

of that file. The adversary can convince the cloud storage server that it owns that file by presenting just the hash value, hence can download the entire file from the server.

VII. CLOUD SECURITY ISSUES

The Cloud security is besides the focus of this effort. Unlike earlier investigations of cloud security concerns, our vital objective is to make available a much more absolute and methodical reporting of the research literature shared to this topic. We give a wide general idea of publications in the areas of cloud computing security and security of remote storage and computation [12]. In particular, the topics covered in this work include:

- **Client authentication and authorization:** We cover the existing unit of work on techniques for troublesome and developing the border between a cloud provider and its customers more often than not carried out by means of a web browser.
- **Security limitations of hardware virtualization:** We explain the difficulties that have faced along with the enormous exploit of hardware virtualization by cloud providers. We point out how virtualization can be used to acquire illegal information from susceptible clients and also point out improvement methods that can be utilized. As well, we also concentrate on vulnerabilities associated to the procedure and sharing of virtual machine (VM) images.
- **Flooding attacks and denial of service (DoS):** Because cloud computing schemes are planned to level according to the demand for stores, an attacker may utilize that feature to maliciously concentrate huge segments of the cloud's computing power, infuriating the superiority of service that the cloud provides to other simultaneous customers. We talk about dissimilar types of attacks on cloud ease of use and their possible results.
- **Cloud dependability, or its capability to confine and representation illegal action:** We discuss competence that a detained liable scheme should have and explanations for accomplishing this competence most cloud providers incriminate their customers according to the authentic procedure of their infrastructure during a pre-established time slice. In the case of a service that is being flooded this procedure will be understandable high which in its twist will most probable interpret to bills that are much advanced than anticipated.
- **Tests and results for remote storage security:** We describe numerous methods that can be utilized by cloud clients to authenticate reliability of their outsourced data.
- **Security of outsourced calculation:** As a final point, we offer a general idea of existing techniques for promising confidentiality and reliability of outsourced calculations.

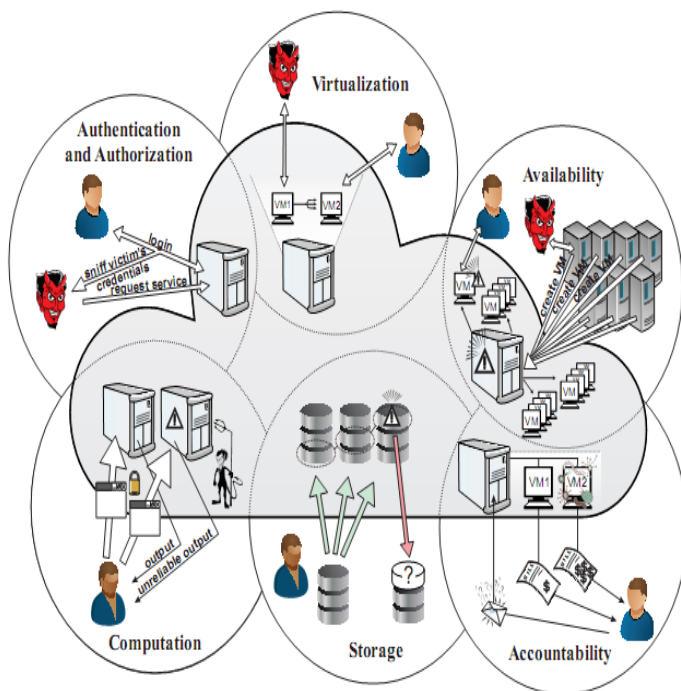


Figure: Overview [12] of cloud security issues.

VIII. BENEFITS OF A CRYPTOGRAPHIC STORAGE SERVICE

The foundation assets of a cryptographic storage service are that (1) control of the data is maintained by the customer and (2) the security properties are derived from cryptography, as opposed to legal methods, substantial protection or right to use control. Consequently, such a service provides numerous forceful benefits over other storage space examines based on public cloud communications. In this segment, we remember some of the most important apprehensions with cloud computing as summarized in the Cloud Security Alliance's recent report [13] and highlight how these concerns can be mitigated by such an architecture.

- A. *Regulatory compliance:* Most countries have laws in place that make organizations answerable for the security of the data that is assigned to them. This is particularly so for the case of personally identifiable information, medical records and financial records. And since organizations are often held responsible for the achievements of their suppliers the use of public cloud storage examine can occupy important legal hazards. In a cryptographic storage service the data is encrypted on idea by the data processor(s). This approach client can be promised that the privacy of their data is protected irrespective of the accomplishments of the cloud storage supplier. This significantly diminishes any authorized introduction for both the client and the supplier.
- B. *Geographic constraints:* Data that is accumulated in definite legal jurisdictions may be area under discussion to guidelines even if it was not composed there. Because it can be easier said than done to establish accurately where

one's data is being accumulated once it is sent to the cloud (i.e., many service providers have data centers arranged all the way through the globe) some clients may be unenthusiastic to utilize a public cloud for apprehension of raising their legal experience. In a cryptographic storage space service data is only accumulated in encrypted form so any law that is appropriates stored data has small to no consequence on the client. This decreases legal exposure for the client and permits the cloud storage supplier to make most favorable utilize of its storage communications, in this manner dropping charges.

- C. *Subpoenas:* If an organization becomes the subject of an investigation, law enforcement agencies may request right to use to its data. If the data is stored in a public cloud the demand may be made to the cloud provider and the concluding could even be avoided from informing the client. This can have strict effects for clients. First, it anticipates the client from demanding the appeal. Second, it can show the way to law enforcement having right to use to data from customers that are not under examination [25]. In such a circumstance can occur due to the fact that service providers often store multiple customers' data on the same disks. In a cryptographic storage service, since data is stored in encrypted form and in view of the fact that the client maintains ownership of all the keys, any request for the data must be made directly to the customer.
- D. *Security breaches:* Even if a cloud storage provider implements strong security practices there is always the possibility of a safety measures violate. If this happens the purchaser may be legally answerable. In a cryptographic storage examination data in encrypted and data integrity can be confirmed at any time. Therefore, a security breach poses little to no risk for the customer.
- E. *Electronic discovery:* Digital information the stage an significant position in legal actions and regularly relations are need to look after and manufacture confirmations for proceedings. Organizations with high stages of litigation may require keeping a copy of huge amounts of data on-premise with the aim of promise its reliability. This can perceptibly contradict the advantages of using a cloud storage service.
- F. *Data retention and destruction:* In many cases a customer may be responsible for the retention and destruction of data it has accumulated. If this data is stored in the cloud, on the other hand, it can be complicated for a client to determine the reliability of the data or to authenticate whether it was appropriately removed. A cryptographic storage service eases these apprehensions since data integrity can be demonstrated and because the information essential to decrypt data (i.e., the master key) is reserved on-premise. Secure data removal can be efficiently accomplished by just erasing the master key.

IX. LITERATURE SURVEY

In this paper [14], author has to solving efficiently the problem of deduplication with discrepancy rights in cloud computing, here they think about a hybrid cloud architecture consisting of a public cloud and a private cloud. As using existing approach for data deduplication the private cloud is involved as a proxy to permit data owner/users to strongly achieve duplicate check with differential benefits. Such architecture is convenient and has concerned much awareness from make inquiries from data owners only outsource their data storage by utilizing public cloud while the data process is deal with in private cloud. A new method sustaining differential duplicate ensure is planned below this mixture cloud planning where the S-CSP resides in the municipal cloud. The user is only permitted to execute the duplicate check for files marked with the parallel privileges. The main goal of this paper is to provide stronger security by encrypting the file with differential privilege keys.

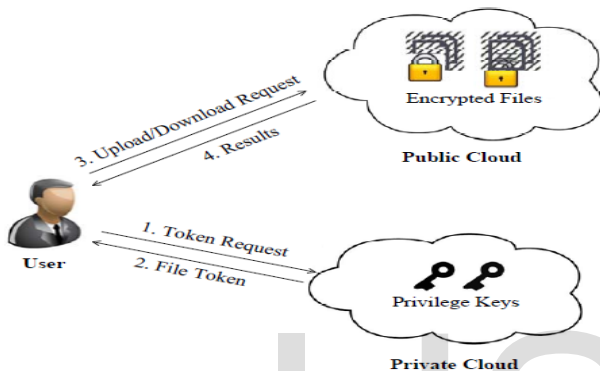


Fig. 1. Architecture for authorized deduplication[14].

As their proposed method has to authorize duplicate check and conduct test-bed experiments to calculate the overhead of the prototype.

To protect the confidentiality author has been proposed [15] to encrypt the data before outsourcing. To enhanced protect data security this paper makes the initial attempt to officially concentrate on the difficulty of authorized data deduplication. Unusual from conventional deduplication systems the degree of difference privileges of users are additional considered in duplicate check as well the data itself. Here they also present common new deduplication constructions sustaining authorized duplicate check in hybrid cloud architecture. Security analysis shows that their method is secure in expressions of the descriptions particular 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 testbed experiments using our prototype. We demonstrate that our suggested authorized replica check method bring upon yourself negligible transparency evaluated to normal operations. 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 secure and dependable cloud storage service, a secure multi-owner data sharing method is proposed

[16] according to any user in the group so that they can steadily share data with others users by the un-trusted cloud. The Group executive is used for reduce 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 human rights for any set of ciphertexts are realizable. Anyone can comprehensive any set of secret keys and make them as compressed as a single key. The private key proprietor can generate a constant-size collective enter of ciphertext set in cloud, but another encrypted files outside 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.

The difficult [16] trouble is how to efficiently share encrypted data. Obviously users can download the encrypted data from the storage, decrypt them then send them to others for sharing, but it loses the value of cloud storage. 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 work is the predefined bounce of the number of maximum ciphertext classes. In cloud storage, the number 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 [17] presents a new privacy-preserving security solution for cloud services. Here in this method deal with user unspecified access to cloud services and shared storage servers using non-bilinear group signatures to ensure anonymous authentication of cloud service client's user. Here the solution provides registered users with anonymous access to cloud services and also offers anonymous authentication. This signifies that user's personal attributes (age, valid registration, successful payment) can be proven without make knowing user's identity. Consequently, users can use services without any threat of profiling their performance. On the other hand, if users 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, unlink ability and the confidentiality of transmitted data. Due to this fact, cloud service providers using our solution can authenticate more clients in the same time. Additionally, there method gives output the experimental results and measure up to the performance with related solutions.

In this paper author [18] has try to assess how can cloud providers earn their customer's trust and provide the security, privacy and reliability, when a third party is meting out sensitive data in a remote machine established in various countries. A thought of utility cloud has been characterized to provide a variety of services to the users. Various technologies can help to concentrate on the challenges of security, privacy

and trust in cloud computing. Unfortunately, the implementation of cloud computing came before the suitable technologies become visible to deal with the supplementary confronts of trust. This opening between implementation and improvement is so extensive that cloud computing consumers don't fully expectation this innovative way of computing. To close this opening, we require identifying with the trust issues join together with cloud computing from both a technology and business perception. Then we'll be able to establish which up-and-coming technologies could best address these problems. Here the author [18] has analyzed the trusted computing in the cloud computing environment and the function of trusted computing platform in cloud computing. The advantages of this move toward are to make bigger the trusted computing technology into the cloud computing environment to accomplish the trusted computing prerequisites for the cloud computing and then accomplish the trusted cloud computing. The significance of

trust varies from organization to organization, depending on the data's value. Additionally, the less expectation an endeavor has in the cloud provider, the more it wants to be in charge of its data smooth the technology. On the other hand, it's fundamental that consumers and providers change their way of thinking's. Trusting cloud computing might differ from trusting other systems, but the objective stay behinds the same to improve business and continue aggressive by take advantage of the advantages of a new technology. Any new technology must progressively build its standing for good presentation and security, earning user's trust over time. We will make more protocol to make available high security for security management, Business continuity management, Identity & access management, Privacy & data protection and application Integrity in the future.

S. No.	Paper	Author	Advantages	Issues
1.	Secure De-duplication And Data Security With Efficient And Reliable CEKM[19]	N.O.AGRAWAL, S.S.KULKARNI	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 scratch of stolen numbers if we decrease the value of that stolen information to the attacker to achieving efficient and reliable key management in secure de-duplication.
2.	A Study on Authorized De-duplication Techniques in Cloud Computing.[20]	Bhushan Choudhary, Amit Dravid	Security by counting differential benefits of clients in the duplicate copy check. Some protected primitives utilized as a part of our safe de-duplication i.e. Symmetric encryption, Convergent Encryption, Proof of Ownership, Identification Protocol.	The security issue is to evaluate the efficient utilization of cloud band width and disk usage.
3.	Private Data De-duplication Protocols in Cloud Storage.[21]	Wee Keong Ng, Yonggang Wen, Huafei Zhu	It formalized in the context 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
4.	A Tunable Proof of Ownership Scheme for Deduplication Using Bloom Filters.[22]	Jorge Blasco, Agustin Orla	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.
5.	A Secured and Authorized Data Deduplication in Hybrid Cloud with Public Auditing.[23]	Sharma Bharat, Mandre B.R.	Specific clients are only permitted to perform the duplicate 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.
6.	Reclaiming Space from Duplicate Files in a Serverless Distributed File System.[24]	John R. Douceur, Atul Adya, William J. Bolosky, Dan Simon, Marvin Theimer	That method is shows that the duplicate-file coalescing system is scalable, highly efficient, and fault-tolerant. <small>IJSER © 2017 http://www.ijser.org</small>	The main issue is enabling the identification, and to reclaim space from this incidental duplication to make it available for controlled file replication.

X. CONCLUSION

Data privacy has been a major concern in cloud storage since users have to trust the cloud service providers for security and privacy. Here we also review on various existing method for new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture to analyzed almost every security threat various de-duplication method for both the cloud models, in which the duplicate-check tokens of files are generated by the private cloud server with private keys.

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 Berkeley, 2009.
- [2] B. Narasimhan and R. Nichols, "State of cloud applications and platforms: The cloud adopters' view," *Computer*, vol. 44, no. 3, pp. 24–28, 2011.
- [3] S. Quinlan and S. Dorward, "Venti: A new approach to archival storage," in Proc. 1st USENIX Conf. File Storage Technol., Jan. 2002, p.7.
- [4] Pearson, S., & Benameur, A. Privacy, security and trust issues arising from cloud computing. In *Cloud Computing Technology and Science (CloudCom)*, 2010 IEEE Second International Conference on (pp. 693-702).
- [5] Tweney, A., & Crane, S. (2007). *Trustguide2: An exploration of privacy preferences in an online world. Expanding the Knowledge Economy: Issues, Applications, Case Studies.*
- [6] Pearson, S., & Charlesworth, A. (2009). Accountability as a way forward for privacy protection in the cloud. *Cloud Computing*, 131-144.
- [7] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, *ACM Conference on Computer and Communications Security*, pages 491–500. ACM, 2011.
- [8] Meyer, D. T., and Bolosky, W. J. A study of practical deduplication. *ACM Transactions on Storage (TOS)*, 2012.
- [9] D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side channels in cloud services: Deduplication in cloud storage," *IEEE Security and Privacy Magazine*, vol. 8, pp. 40-47, 2010.
- [10] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, *ACM Conference on Computer and Communications Security*, pages 491–500. ACM, 2011.
- [11] M. Mulazzani, S. Schrittwieser, M. Leithner, M. Huber, and E. Weippl, "Dark clouds on the horizon: using cloud storage as attack vector and online slack space," in Proc. USENIX Security Symposium (SEC'11), 2011.
- [12] Everaldo Aguiar, Yihua Zhang, and Marina Blanton, "An Overview of Issues and Recent Developments in Cloud Computing and Storage Security" 2012.
- [13] Cloud Security Alliance. «Security Guidance for Critical Areas of Focus in Cloud Computing. April 2009.
- [14] 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.
- [15] 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.
- [16] Gade Swati, Prof. Prashant Kumbharkar, "Cryptosystem For Secure Data Sharing In Cloud Storage" *IJIRT Volume 1 Issue 6* 2014.
- [17] Lukas Malina and Jan Hajny, "Efficient Security Solution for Privacy-Preserving Cloud Services" 6th International Conference On Telecommunications Signal Processing Year 2013.
- [18] Pardeep Kumar, Vivek Kumar Sehgal, Durg Singh Chauhan, P. K. Gupta and Manoj Diwakar, "Effective Ways of Secure, Private and Trusted Cloud Computing" *JCSI International Journal of Computer Science Issues*, Vol. 8, Issue 3, No. 2, May 2011.
- [19] 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 (IIAEM)*, November 2014.
- [20] 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.*
- [21] Wee Keong Ng, Yonggang Wen, Huafei Zhu, "Private Data Deduplication Protocols in Cloud Storage" *ACM 978-1-4503-0857-1/12/03*, 2011.
- [22] Jorge Blasco, Agustin Orfila, "A Tunable Proof of Ownership Scheme for Deduplication Using Bloom Filters" June 18, 2014.
- [23] 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*
- [24] 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.
- [25] Wired. «Company Caught in Texas Data Center Raid Loses Suit Against FBI. » 8 April 2009.