MultiCopy Dynamic Data Storage in Cloud Computing: A Review

Pooja Mishra¹, Sumit Dhariwal²
¹M.E (CSE) Scholar, Department of CSE, SIRT-E, Bhopal, India
²Asst. Professor, Department of CSE, SIRT-E, Bhopal, India

Abstract—As the rapid data expansion of data most current years, cloud computing has turn out to be one of the furthermost accepted area in IT industry with considers to cost, proficient use of computing possessions for example storage space and system bandwidth in the cloud computing communications. Here in this paper survey of all the existing techniques are analyzed and discussed which are based on the criteria of storing Multiple Copies of Dynamic Data in Cloud Computing. By Analyzing the various advantages and limitations of the existing techniques a new and efficient is implemented in future.

IndexTerms—Cloud Storage, Data deduplication, Cloud Computing, Multicopy Dynamic Data.

I. INTRODUCTION

In recent years, the popularity of cloud storage services has increased dramatically. Most business processes have been digitalized, i.e., information such as communication data, accounts, contracts, advertising material, construction or business plans only exists in digital form. For a company, the loss of data could ruin the basis for business. Additionally, companies are legally obliged to preserve tax records for a certain period (3-7 years), and to leave them available to the fiscal authorities. This requires secure methods of preserving important data in order to prevent unrecoverable data loss, whilst constantly keeping up with increasing demands for storage space. It is necessary to regularly make extra copies of the information, so as to be able to restore it to an earlier version if need be. These copies further escalate the demand for storage space. Additional requirements arise from the variety of devices used to access the data simultaneously [1]. Private and business users demand an easy way to synchronize and access their data independent of both device and location. The software providing these features must also be tailored to the needs of the individual with no technical background. In order to meet these demands, companies make large investments into their IT infrastructure.

Additional hardware and software is required, as well as staff for its operation and maintenance. Larger companies might have to consider building a dedicated data center. These expenses conflict with the continuing need to reduce costs in order to stay competitive. Cloud storage services offer user-friendly, easily accessible and money-saving ways of storing and automatically backing up arbitrary data. These services are available on-demand on the Internet [2]. A customer simply accesses the website of a cloud storage provider and rents storage space as necessary by selecting one of the provider's packages.

If the use of cloud storage services carries such great advantages, why are individuals and companies alike still hesitant to entrust their data to the cloud? Usage of a cloud storage provider basically means entrusting data to a third party where no prior relationship based on trust has been established. Individuals who upload personal information to the cloud want to be sure that only certain people are able to access it.
This should also exclude the provider, since there is no justify able reason for access the data. Companies may entrust containing sensitive business data and valuable intellectual property which may be of great interest for industrial reasoning [5]. The unauthorized disclosure of customer information, business secrets or research data poses a serious threat to a company’s business. In addition, compliance requirements with both internal security guidelines and legal regulations have to be met. The cloud storage provider may be subject to different legal regulations than the user.

A cloud service has three distinct characteristics that differentiate it from traditional hosting. It is sold on demand, typically by the minute or the hour; it is elastic a user can have as much or as little of a service as they want at any given time and the service is fully managed by the cloud service provider the consumer needs nothing but a personal computer and Internet access. The advantage of cloud is cost savings. The prime disadvantage is security on cloud computing is used by many software industries nowadays. Since the security is not provided in cloud, many companies adopt their unique security structure. For e.g. Amazon has its own security structure. Introducing a new and uniform security structure for all types of cloud is the problem we are going to tackle in this idea. Since the data placed in the cloud is accessible to everyone, security is not guaranteed.

Figure 1: Architecture of cloud storage.

In addition, the large amount of cloud data and owner's confined computing competence additional makes the job of data correctness auditing in a cloud environment exclusive and even difficult for entity cloud customers. Consequently, facilitating public audit ability [10], [13], [14] for cloud storage is of critical importance so that owners can alternative to a concentrated third party auditor (TPA) to audit cloud storage services and sustain strong storage accuracy assurance, while saving their own valuable computing resources.

As data in Cloud is dynamic, static auditing is not an adequate amount of cloud environment. A dynamic auditing is required to authenticate the data integrity of the dynamic data. But as data are self-motivated in cloud, it is not uncomplicated to have an auditing competently. Server can put into effect replay attack and counterfeit attack to fail the auditing procedure. The dynamic procedures consist of alteration, insertion and deletion. Whenever you like dynamic operation is achieved the owner sends to bring up to date message to the auditor characterizing the index number of that message. The auditor updates the table. The message m and the tag are reinstated by the new
message and tag in message modification. The new message m and new tag are inserted in insertion operation. The message m and tag are deleted from the index table and all the entries below the deleted message move upwards. After performing updates in the table, the auditor conducts the data integrity test for the keep informed data. Auditor sends the consequence to the owner and he deletes the local copy of keep informed data.

![Dynamic TPA System](image)

**Figure 2: Dynamic TPA System**

TPA checks the integrity of data on the cloud on the behalf of the users, and it provides the reasonable way for the users to check the validity of data in the cloud. On the whole, enabling public auditing services plays a vital role in establishing cloud economy, where by users need way to assess to risk and gain faith in the cloud [3]. Public auditing in addition to user provides the external party to verify the correctness of stored data against the external attacks. However these schemes [4], [5], [6] don’t involve the privacy protection of the data. It is a main disadvantage which affect the security of the protocols in cloud computing.

**THIRD PARTY AUDITOR**

The third party auditor (TPA), who has expertise and capabilities that cloud users do not have and is trusted to assess the cloud storage service security on behalf of the user upon request. Users rely on the CS for cloud data storage and maintenance. They may also dynamically interact with the CS to access and update their stored data for various application purposes [7]. The users may resort to TPA for ensuring the storage security of their outsourced data, while hoping to keep their data private from TPA.

We consider the existence of a semi-trusted CS as does. Namely, in most of time it behaves properly and does not deviate from the prescribed protocol execution [8] [9]. However, during providing the cloud data storage based services, for their own benefits the CS might neglect to keep or deliberately delete rarely accessed data files which belong to ordinary cloud users.

**ENCRYPTION TECHNIQUES**

Some amount of data security can be achieved through the encryption of the data. A Searchable Symmetric Encryption (SSE) is used to enable users to securely retrieve the ciphertext. The problems with these techniques are they support only Boolean keyword search. That means whether a keyword
exists in a file or not, without considering the difference of relevance with the queried keyword of these files in the result. Some encryption techniques are:

**ECIES:** The Elliptic Curve Integrated Encryption Scheme is a public-key encryption scheme based on ECC. It is designed to be semantically secure in the presence of an adversary capable of launching chosen-plaintext and chosen-cipher text attacks.

The advantage of the ECIES is that on one hand it is meanwhile quite well investigated and thus considered secure while on the other hand just a very short bit length is needed as compared to other asymmetric systems.

**ECC:** Elliptical curve cryptography is a public key encryption technique based on elliptic curve theory that can be used to make faster, less significant and more competent cryptographic keys. ECC produces keys during the properties of the elliptic curve equation as an alternative of the conventional method of creation as the produce of very huge prime numbers. The technology can be used in coincidence with most public key encryption methods, such as RSA, and Diffie-Hellman. According to some examiners, ECC can give way a stage of security with a 164-bit key that other schemes require a 1,024-bit key to accomplish.

**AES:** It is stand for Advanced Encryption Standard. It is a specification of the electronic data encryption. The Advanced Encryption Standard comprises three block ciphers, AES-128, AES-192 and AES-256. AES as a fixed block size of 128 bits and a key size of 128, 192, or 256 bits. The block-size has a maximum of 56 bits, but the key-size has no theoretical maximum. The cipher uses number of encryption rounds which converts plain text to cipher text.

**DES:** It stands for Data encryption standard. It is a widely-used method of data encryption with the help of private or secrete key. DES uses 56-bit key to each 64-bit block of data. It can run in various modes and involves 16 rounds or operations. Although this is considered "strong" encryption, many companies used ‘triple DES’ that uses three keys in succession.

### II. LITERATURE SURVEY

In this paper [8], here author has initiate the ideas of resemblance significance and method strength to prepare the confidentiality concern in searchable encryption methods and then explain the lack of confidence difficulty by proposing a two-round searchable encryption (TRSE) method. New knowledge in the cryptography group of people and information retrieval (IR) community are employed, counting homomorphic encryption and vector space model. In the anticipated method, the common of calculating effort is completed on the cloud while the customer takes part in position which assurances top-k multi-keyword recovery over encrypted cloud data with high protection and convenient competence. Our involvements can be reviewed as follows:

We suggest the ideas of relationship significance and method strength. We consequently complete the primary effort to prepare the privacy concern in searchable encryption and they give you an idea about server-side ranking based on order-preserving encryption (OPE) unavoidably abuses data confidentiality.

We recommend a TRSE scheme, which fulfills the protected multi-keyword top-k retrieval over encrypted cloud information. Specifically, for the primary instance they employ relevance score to support multi-keyword top-k retrieval.
Thorough analysis on security demonstrates the suggested method assurances high data privacy. Furthermore, performance analysis and experimental results show that our scheme is efficient for practical utilization. By security analysis, they demonstrate that the recommended method assurances data privacy. According to the efficiency evaluation of the proposed method over an actual dataset, general investigational effects show that our method ensures practical efficiency.

Ranked search significantly improves scheme usability by recurring the matching files in a ranked arrange concerning to definite significance criteria (e.g., keyword frequency), thus creation one step earlier in the direction of sensible use of privacy-preserving data hosting services in the circumstance of Cloud Computing. To accomplish their plan objectives on both scheme safety measures and usability they suggest bringing mutually proceed of both crypto and IR community to plan the ranked searchable symmetric encryption scheme, in the spirit of “as-strong-as-possible” security guarantee. Specifically, they investigate the statistical calculate approach from IR and text-mining to embed weight information (i.e. significance score) of each file for the duration of the organization of searchable index earlier than outsourcing the encrypted file collection.

In this paper [9], for the first time we define and resolve the problem of efficient yet confined ranked keyword search over encrypted cloud information. Ranked search significantly improves scheme usability by recurring the matching files in a ranked order concerning to definite significance criteria (e.g., keyword frequency), thus making one step earlier in the direction of practical deployment of privacy-preserving data hosting services in Cloud Computing. We first give a simple however perfect construction of ranked keyword search under the modern searchable symmetric encryption (SSE) security definition, and demonstrate its inefficiency. To accomplish more convenient performance, then they recommend a definition for ranked searchable symmetric encryption and offer an efficient plan by appropriately using the offered cryptographic primitive order-preserving symmetric encryption (OPSE). Systematic study give you an idea about that our anticipated solution take pleasure in “as-strong-as-possible” safety measures promise evaluate to earlier SSE methods, while appropriately appreciating the objective of ranked keyword search. A widespread experimental outcome shows the efficiency of the anticipated solution.

In this paper [10], they recommend a secure cloud storage scheme sustaining privacy-preserving public auditing. Our work is between the initial a small amount of ones to support privacy-preserving public auditing in cloud computing with a center of attention on data storage space. In addition, with the prevalence of cloud computing, a foreseeable increase of auditing tasks from unusual customers may be delegated to TPA. We use the homomorphic linear authenticator and random masking to assurance that the TPA would not become skilled at any information about the data substance stored on the cloud server during the well-organized auditing procedure which not only removes the load of cloud consumer from the tedious and probably costly auditing job but also improves the customers’ apprehension of their outsourced data outflow. The TPA can achieve multiple auditing jobs in a consignment way for enhanced effectiveness. Widespread study demonstrates that our methods are provably protected and extremely efficient. Our preliminary experiment conducted on Amazon EC2 instance further demonstrates the fast performance on public cloud as an important future extension, which is expected to robustly cope with very huge scale information and thus encourage users to adopt cloud storage services more confidently.

In this paper [11], we present such a system—Cloud Capacity Manager (CCM)—an on-demand compute capacity management system of virtualized data centers at scales of that combines various low-overhead techniques, motivated by practical on-field observations, to achieve scalable capacity of machines. CCM achieves this scale by employing three-level hierarchical management architecture. CCM also discards light on the tradeoffs due to two inevitable concerns in large-scale commodity data
centers: 1) sustaining low equipped overhead, known unpredictable cost of performing arts management process essential to deal out resources, and 2) coping with the enhanced incidences of these operations’ failures. The capacity managers at each level continuously monitor and aggregate black-box VM CPU and memory practice information, and then use this aggregated data to make self-sufficient and localized capacity allocation choices. An experimental evaluation on a fairly large infrastructure, that to achieve better capacity multiplexing, the focus needs to not only be on the accurate prediction of workload demand and aggressive optimization of the allocation algorithms, but also on dealing with the practical limitations of real-life infrastructures.

The advent of cloud computing data holders is encouraged to subcontract their complex data management systems from local sites to profitable public cloud for huge elasticity and economic savings. But for defending data confidentiality susceptible data has to be encrypted before outsourcing which outdated conventional. Making an allowance for the huge amount of data consumers and documents in cloud it is essential for the investigate service to authorize multi-keyword query and make available result similarity ranking to meet the efficient data retrieval required. Related jobs. In this paper [12], author has describe and explain the difficult trouble of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE) and create a set of exacting privacy conditions. Among various multi-keyword semantics they make a decision the well-ordered average of “coordinate corresponding”, i.e., as many matches as feasible to confine the comparison between search query and data documents and additional utilize “inner product comparison” to quantitatively formalize such principle for similarity measurement. We primary suggest a essential MRSE method using secure inner product calculation and then considerably get better it to assemble unusual privacy conditions in two levels of threat models. Comprehensive study examining privacy and effectiveness assurances of proposed methods is known and tests on the real-world dataset additional give you an idea about recommended methods definitely initiate low transparency on calculation and communication.

This paper [13] aspires to make available penetrating a folder over mist environment using manifold keywords demonstrating the file with numerous credible circumstances. The aspire is to make available the protection to its maximum extent by including encryption and decryption methods. Authorization of the users directly by the administrators allows the files involved to transfer more securely. Encryption and decryption of both file name and file which uses asymmetric and symmetric key algorithms respectively. The secret key is generated for each consumer to prevent any other user to misuse the file. The data that protected on cloud complete from any attack that is caused both by external and internal attackers. Most of the inside attacks are utilized by the cloud providers by using comparison significance and analysing the statistical leakage. Based on the usage of the file over ranked manner, it is easy to get all the details of used files through probability prediction. This kind of data leakage should be completely avoided and maximum protection. The solution suggests the same by applying some new concepts to increase the data security.

<table>
<thead>
<tr>
<th>Paper Title</th>
<th>Author / Year</th>
<th>Technique Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improving Security and Efficiency in Attribute-Based Data Sharing.</td>
<td>Junbeom Hur</td>
<td>Here CP-ABE attribute based data sharing technique is used which solves key escrow problem and proxy encryption.</td>
</tr>
<tr>
<td>A Survey on Attribute Based Encryption Scheme in cloud computing</td>
<td>Minu George, Dr. C.Suresh Gnanadhas, Saranya.K</td>
<td>Here in this paper presented a review on different attribute based encryption methods.</td>
</tr>
<tr>
<td>Attribute-Based Encryption for Circuits from Multilinear Maps</td>
<td>Sanjam Garg Craig Gentry Shai Halevi Amit Sahai Brent Waters</td>
<td>Here in this paper proposed a new technique for the attribute based encryption using the concept of Multilinear Maps.</td>
</tr>
</tbody>
</table>
| Fuzzy Identity-Based | Amit Sahai, Brent Waters | A Fuzzy IBE scheme can be applied to enabl
### III. CONCLUSION

Cloud computing enables various users to share or access resources over the internet, but during the data sharing or storage in cloud security plays a vital role and hence various auditing protocols are implemented for the security of these cloud data and also provides privacy preservation between users. Here in this paper survey of all the existing techniques are analyzed and discussed which are based on the criteria of storing Multiple Copies of Dynamic Data in Cloud Computing. By analyzing the various advantages and limitations of the existing techniques a new and efficient is implemented in future.

### REFERENCES


V. M. Arrington, —Gmail Disaster: Reports of Mass Email Deletions,” Dec. 2006;


XIII. Cong Wang, Sherman S.M. Chow, Qian Wang, Kui Ren, and Wenjing Lou, “Privacy-Preserving Public Auditing for Secure Cloud Storage,” IEEE Transactions on Computers (TC), 2011 (A preliminary version of this paper appeared at the 29th IEEE Conference on Computer Communications (INFOCOM10)).

XIV. Cong Wang, Qian Wang, Kui Ren, and Wenjing Lou, “Towards Secure and Dependable Storage Services in Cloud Computing.” To appear, IEEE Transactions on Service Computing (TSC). (A preliminary version of this paper appeared at the 17th IEEE International Workshop on Quality of Service (IWQoS’09)).