Implementation of Data Sharing in Cloud Storage Using Data Deduplication

1. M.JITHENDRA, 2. Dr. G.V.S.N.R.V. PRASAD M.S, Ph.D.

1. M.Tech, Department of CSE, Gudlavalleru Engineering College.

2. Professor & Dean –A .A, Department of CSE, Gudlavalleru Engineering College.

Abstract-Cloud computing innovation is broadly utilized so the information can be outsourced on cloud can be accessed easily. The cloud service provider and users authentication is necessary to make sure, no loss or leak of user's data. On cloud, anyone can share the information as much as they want. Cryptography helps the data owner to share the data in a secured way. The encryption and decryption keys may be different for different set of data.

In Key Aggregate Cryptography (KAC), users encode a message under a public-key, as well as under an identifier of ciphertext called class. Information deduplication is one of essential information compression techniques for eliminating duplicate copies of information, and has been broadly utilized as a part of cloud storage to scale back the number of space for storing and save information measure. We also present several new deduplication constructions supporting authorized duplicate check in a hybrid cloud design.

Index Terms- Cloud Storage, Data Sharing, Key-Aggregate Encryption.

Author names-

1. M.JITHENDRA,
M.Tech Student, Department of CSE,
Gudlavalleru Engineering College, A.P, India.

2. Dr. G.V.S.N.R.V. PRASAD M.S, Ph.D.,
Professor & Dean–A.A, Department of CSE,
Gudlavalleru Engineering College, A.P, India.

1. INTRODUCTION

New computing paradigms are drastically rising. One notable example that can be taken into consideration is replacement of economic computing model created by the advancement in networking technology, where a shopper can leverage a service provider's computing, storage or networking infrastructure. With the new exponential rate of data, there is increasing demand for outsourcing data storage to cloud services like Microsoft's Azure and Amazon's S3,their integrated tools, pre-assembled templates and managed services make it easier to build and manage enterprise, mobile, Web and Internet of Things (IoT) apps faster.

Storing data remotely to the passing versatile on demand manner brings appealing benefits: relief of the weight for capacity administration, universal data access with freelance geographical locations, and shunning valuable on hardware, software, and personnel maintenances etc. Although the infrastructures at a lower price on the cloud unit of activities proved to be much more powerful and reliable than personal computing devices, they are still facing the internal and external threats for data integrity.

ISSN: 2278 – 7798

Secondly, there exist numerous motivations for CSP (cloud service provider) to behave unreliably towards the cloud users concerning the standing of their outsourced data examples. CSP might reclaim storage for financial reasons by discarding the data, it is not been accessed.

Considering data privacy, by the standard implies that it utterly depends upon the server to provide the access management alone. Once authentication, it recommends that any surprising increase can expose all data. Because of its shared atmosphere, things become worst. Data in VM is additionally receiving by instantiating another VM co-resident with the target one. Ordinarily in study schemes, third party auditor (TPA) can check the availability of data on behalf of owner however cloud server does not trust TPA, therefore we have to follow vary theoretic approach permanently. Users need to cipher their own data by victimization their own keys before uploading. Data sharing is crucial utility in cloud storage.

Clearly, user can transfer encrypted data and decipher them, and share with others; but this approach violates value of cloud storage. For example, consider two military camps A, B among the suggested camps. Here camp A can share the secured data only with camp B and does not expose the data to any of the camps. The camp A encrypts all maps with corresponding keys before uploading and send keys firmly to the camp B.

1. Satellite cryptography is a rising innovation in which two parties may simultaneously shared, secret cryptographic key material using the Satellite.

- Camp A encrypts all files with one cryptography key and provides camp B the corresponding secret key directly.
- Camp A encrypts files with distinct keys and sends camp B the corresponding secret keys

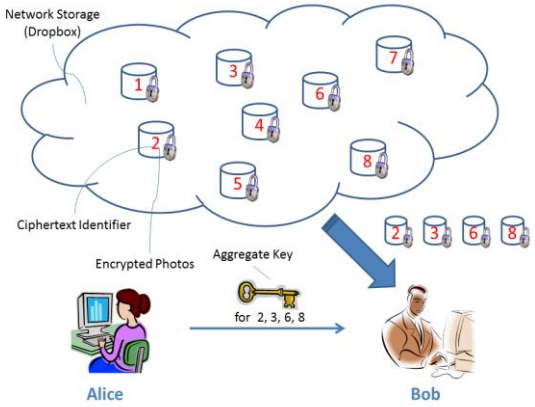

Fig.1 Alice shares files with identifiers 2, 3, 6 and 8 with Bob by sending him multiple keys.

So, if we have an inclination to appear at the first methodology, it is not acceptable since all numerous maps might place information along belies conjointly leaked to camp B. For the second methodology, there are good issues on potency. For having distinct key cryptography, sender ought to send multiple keys. Transferring is completed through secure channel and storing of keys wishes secure storage. Succeeding value and quality will increase.

Encryption keys also come with two flavors- symmetric key and asymmetric (public) key. By employing the Satellite Cryptography[1], once camp A wishes the maps to be originated from a third party, sender ought to offer the encrypted her secret key; clearly, typically this can be often not invariably fascinating.

By second approach, public key cryptography offers much flexibility for our

ISSN: 2278 – 7798

applications. In enterprise settings, every employee can transfer encrypted data on the cloud storage server while not the data of the company's master-secret key i.e. public key cipher, much flexibility provides. Therefore, best answers are attending to the camp A, encrypts the map with distinct key however sends alone single cryptography key that is of Constant size. Since in the cryptography, secrete keys are to be sent through a protected channel.

For example, users should not pretend massive storage for cryptography keys among the resource constraint devices like smart cards. Especially, these secret keys generally hold among the tamper-proof memory, which is comparatively valuable. These analysis efforts mainly target minimizing communication radiant like combination sign.

2. Literature Survey

Cloud computing is imagined as structural planning for succeeding generation. It has numerous facilities however having a risk of attacker who can get to the information or leak the user's identity. While setting a cloud users and administration suppliers confirmation is important. The issue arises whether cloud administration supplier or user is not compromises. The information will leak if any one of them is compromised. The cloud should be simple; preserving the privacy is also maintaining user's identity. [1]

The adaptable utilization of cloud storage for user is a need as it getting to information locally. It is critical to investigate the information set on the cloud. Consequently, it is important to allow a public audit for integrity of outsourced data through TPA^2. TPA is likewise helpful for cloud administration supplier. It checks the accuracy of the outsourced information. TPA ought to have the capacity to do public audibility, storage correctness, privacy preserving etc. [2]

Another way for sharing encoded information is Attribute-Based Encryption (ABE). It is likely to encrypt the information with attributes that are proportional to users attribute rather than only encrypting each part of data. In ABE, attributes description considered as set so that only a particular key, which is matches' with attribute, can decrypt the cipher text. The user key and the attribute are matched it can decrypt a particular cipher text. At the point when there is 'k', attributes are overlay among the cipher text and a private key then decryption is granted. [3]

A multi group key management achieves a hierarchical access control by applying an integrated key graph also handling the group keys for different users with multiple access authorities. Centralized key administration plan uses tree structure to minimize the information preparing, communication and storage capacity overhead. It finishes an integrated key graph for every user. [4]

2. In order to maintain the integrity of data, the user takes the assistance of a Third Party Auditor (TPA).

In Identity-based encryption (IBE), the public key of user contains distinct information of user's identity. The key can be textual value or domain name, etc. The

ISSN: 2278 – 7798

identity of the user is used as identity string for public key encryption. A trusted party called private key generator (PKG) in IBE, which has the master secret key and gives secret key to users according to the user identity. The information owner collaborate the public value and the identity of user to encrypt the information. The cipher text is decrypted using secret key. [7]

Numerous cloud users need to transfer the information without giving much personal details to other users. The anonymity of the user is to be preserved so that not to reveal the identity of data owner. Provable Data Possession (PDP) utilizes comparative demonstrating marks to reduce computation on server, and network traffic. Security Mediator (SEM) methodology permits the user to preserve the anonymity. Users are transferring all their information to SEM so that the SEM is not able to understand the information although it is going to produce the verification on data. [8]

Multiattribute-Authority analyzed the attributes with respect to the decryption key and the user must get a particular key related to the attribute while decrypting a message. The decryption keys are distributed freely to users. Multiattribute-Authority encryption allows real time deployment of attribute based privileges as different authorities issue different attributes. The attribute authorities ensure the honesty of the user privilege so central authority maintains the confidentiality. [9]

3. Existing System

In Key-Aggregate Cryptosystem for Scalable Data Sharing in Cloud Storage, users encode a message under a public-key, as well as under an identifier of ciphertext called class. The key proprietor holds a key called master-secret key, which can be utilized to extract secret keys for different classes. The sizes of ciphertext, public-key, and master-secret key and aggregate key in KAC schemes are all of constant size.

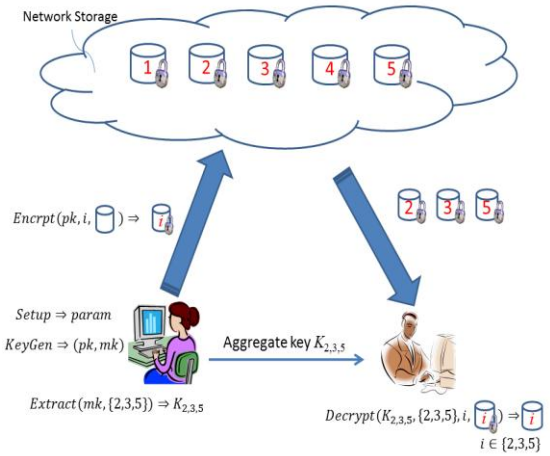

Fig.2 KAC for data sharing in cloud storage

B. Algorithm

1. Setup (1^{λ},n): Setup ensures that the owner of the info will construct the general public system parameter andhe\sheproduces account on cloud. When coming into the input, the overall of cipher text categories ‘n’ and a security level parameter 1^{λ}.

2. KeyGen: It is for generation of public/master-secret key pair (pk, msk).

3. Encrypt (pk, i, m): Run anyone who wish to change over plaintext into cipher text, on the input public key and index.

4. Extract (msk, S): Provide input as master secret key and ‘S’ indices of various

ISSN: 2278 – 7798

cipher text categories. They are extracting by the info owner himself.

5. Decode (Ks, S, i, C): The decrypted original message 'm' is displays on coming into S, C, and I if and as long as I belongs to the set S.

4. PROPOSED SYSTEM

A Hybrid Cloud could be a joined type of private clouds and public clouds in which some some basic information resides within the enterprise's private cloud whereas alternative information is keep and accessible from a public cloud. Hybrid clouds get chance to convey the benefits of adaptability, dependability,, rapid deployment and potential expense savings of public clouds with the protection and administration and administration of individual clouds.

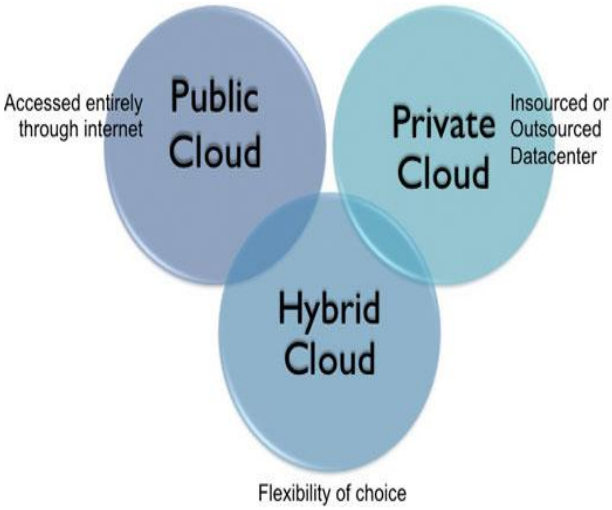

Fig.3 Architecture of Hybrid cloud.

The crucial challenge of cloud storage or cloud computing is the administration of the endlessly increasing volume of information. Data deduplication or Single Instancing basically refers to the elimination of redundant data. Within the deduplication technique, copy data is erased, just one copy of the info to be hold on. On the other hand, indexing of all information is stays kept up should that data ever be required. In general the data deduplication disposes the duplicate copies of repeating data. The information is encoded before outsourcing it on the cloud or network. This encryption requires needs more to cipher the information. In case of huge information storage the encryption becomes even more complex and critical. By exploitation the information deduplication within a hybrid cloud, the encryption will become less complicated.

Many vast scale network utilizes the information cloud to store and share their data on the network. The node or client, which is present in the network have full rights to transfer or download information over the network. But many times different user transfers the same information on the network. Which will make a duplication inside the cloud. If the user wants to retrieve the information or download the information from cloud, every time he\she has needs to utilize the two encrypted files of same information. The cloud will do same operation on the two duplicates of information files. Because of this the information privacy and the security of the cloud get violated. It creates the burden on the operation of cloud.

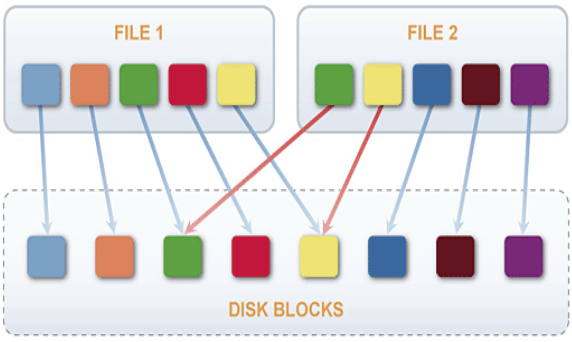

Fig.4 Example of De-duplication

ISSN: 2278 – 7798

To make data management adaptable in cloud computing, deduplication has been a well-known strategy and has attracted more and more attention recently. Information deduplication is a specific information compression technique for eliminating duplicate copies of repeating information in storage. The strategy is utilized to enhance storage utilization. Rather than keeping numerous information duplicates with the same substance, deduplication eliminates redundant information by keeping one and only physical duplicate. Deduplication can occur at either the file level or the block level. For filelevel deduplication, it disposes duplicate copies of the same file. At the block level, which eliminates duplicate blocks of data that occur in non-identical files.

In cloud storage, the quantity of cipher texts usually grows rapidly. Therefore, we have to hold enough ciphertext classes for the future extension. Information deduplication is one of important data compression techniques for eliminating duplicate copies of repeating data, and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To secure the privacy of sensitive information while supporting deduplication, the convergent encryption method has been proposed to encrypt the information before outsourcing. To better protect data security, this paper makes the first endeavour to formally address the problem of authorized data deduplication.

Algorithm:

- **SHA-1:-**

  This SHA-1 algorithm consists following categories.

  - **File Tag (File) -** It processes SHA-1 hash of the File as File Tag.
  - **Token Req (Tag, User ID) -** It requests the Private Server for File Token generation with the File Tag and User ID.
  - **Dup Check Req (Token) -** It requests the Storage Server for Duplicate Check of the File by sending the file token received from private server.
  - **Share Token Req (Tag, {Priv.}) -** It requests the Private Server to generate the Share File Token with the File Tag and Target Sharing Privilege Set.
  - **File Encrypt (File) -** It encrypts the File with Convergent Encryption using 256-bit AES algorithm in cipher block chaining (CBC) mode.
  - **File Upload Req (File ID, File, Token) -** It transfers the File Data to the Storage Server if the file is Unique and updates the File Token stored.

ISSN: 2278 – 7798

5. RESULTS AND DISCUSSION

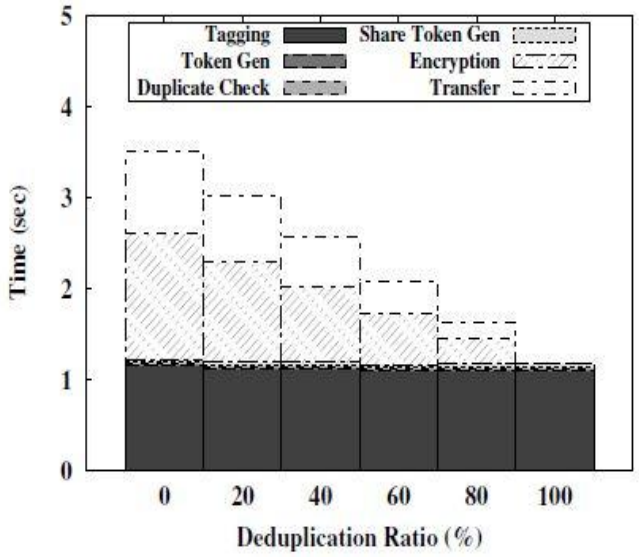

Fig.5 Time Breakdown for Different Deduplication Ratio

Deduplication Ratio:-

To assess the effect of the deduplication proportion, we plan two kind information sets, each of which consists of 50 100MB records. We first transfer the first set as an initial transfer. For the second transfer, we pick a portion of 50 records, as per the given deduplication ratio, from the beginning set as copy documents and remaining records from the second set as remarkable documents.

The average time of transferring the second set is displayed in Figure 5. As transferring and encryption would be skipped in the event of copy record, the time spent on them two declines with expanding deduplication proportion. The time spent on copy check likewise diminishes as the seeking would be finished when copy is found. Aggregate time spent on transferring the document with deduplication proportion at 100% is just 33.5% with exceptional records.

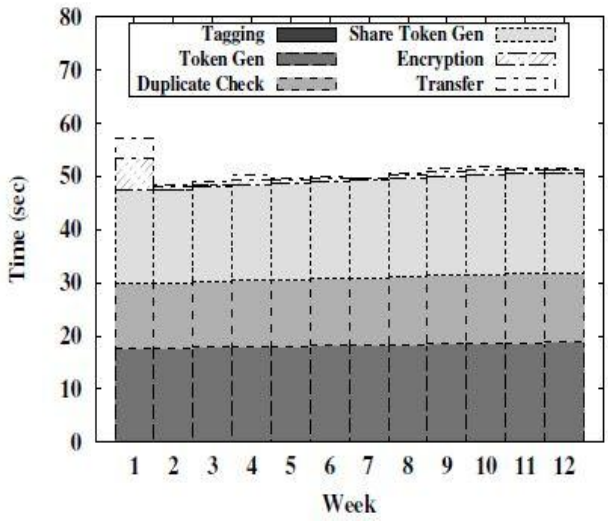

Fig.6 Time Breakdown for the VM dataset.

Real-World VM Images:-

We consider a dataset of week after week VM picture previews gathered over a 12-week compass in a college programming course; while the same dataset is additionally utilized as a part of the earlier work. We perform piece level deduplication with a settled square size of 4KB. The introductory information size of a picture is 3.2GB. Following 12 weeks, the normal information size of a picture increments to 4GB and the normal deduplication proportion is 97.9%. For security, we just gathered cryptographic hashes on 4KB settled size squares. Figure 6. demonstrates that the time taken in token era and copy checking increments directly as the VM picture develops in information size. The time taken in encryption is low in view of the high deduplication proportion. Time taken for the first week is the most noteworthy as the starting transfer contains more one of a kind information.

ISSN: 2278 – 7798

6. CONCLUSION

This public key cryptosystem systematically manufacture cipher texts of constant size such economical delegation of secret writing rights for any set of cipher texts is possible. This not exclusively enhances user privacy and confidentiality of information in cloud storage, but it will supporting the distribution or appointing of secret keys varied for diverse cipher text classes and generating keys by varied derivation of cipher text class properties of the data and its associated keys.

As there are prescribed limits for the storage data in the quantity of cipher text classes. With the advancement and exponential growth, there is a demand for reservation of cipher text classes for future use. For this purpose, this paper ensures "Data Deduplication" concept for eliminating the same files. This sums up the scope of our paper.

REFERENCES

[1] Key –Aggregate Crypto system for Scalable Data Sharing in Cloud Storage Cheng-Kang Chu, Sherman S. M. Chow, Wen-GueyTzeng, Jianying Zhou, and Robert H. Deng.

[2] C Wang, S. S. M. Chow, Q. Wang, K. Ren, and W. Lou, "PrivacyPreserving Public Auditing for Secure Cloud Storage," IEEE Trans.Computers, vol. 62, no. 2, pp. 362–375, 2013.

[3] V. Goyal, O. Pandey, A. Sahai,and B. Waters, "Attribute-Based Encryption for Fine-Grained Access Control of Encrypted data,"in Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS '06). ACM, 2006, pp. 89–98.

[4] F. Guo, Y. Mu, Z. Chen, and L. Xu, "Multi-Identity Single-Key Decryption without Random Oracles," in Proceedings of Information Security and Cryptology (Inscrypt '07), ser.

[5] Jin Li; Yan Kit Li; Xiaofeng Chen; Lee, P.P.C.; Wenjing Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication," in *Parallel and Distributed Systems, IEEE Transactions on* , vol.26, no.5, pp.1206-1216, May 1 2015

[6] S. S. M. Chow, C.-K. Chu, X. Huang, J. Zhou, and R. H. Deng,"Dynamic Secure Cloud Storage with Provenance," in Cryptography and Security: From Theory to Applications - Essays Dedicated to Jean-Jacques Quisquater on the Occasion of His 65th Birthday, ser. LNCS, vol. 6805. Springer, 2012, pp. 442–464.

[7] S. S. M. Chow, Y. J. He, L. C. K. Hui, and S.-M. Yiu, "SPICE - Simple Privacy-Preserving Identity-Management for Cloud Environment," in Applied Cryptography and Network Security – ACNS 2012, ser. LNCS, vol. 7341. Springer, 2012, pp. 526–543.

[8] B. Wang, S. S. M. Chow, M. Li, and H. Li, "Storing Shared Data on the Cloud via Security-Mediator," in International Conference on Distributed Computing Systems - ICDCS 2013. IEEE, 2013.

[9] F. Guo, Y. Mu, Z. Chen, and L. Xu, "Multi-Identity Single-Key Decryption without Random Oracles," in Proceedings of Information Security and Cryptology (Inscrypt '07), ser. LNCS, vol. 4990. Springer, 2007, pp. 384–398.

1. M.JITHENDRA
M.Tech Student, Department of CSE, Gudlavalleru Engineering College. A.P., India.

2. Dr. G.V.S.N.R.V. PRASAD
M.S, Ph.D.
Professor & Dean –A.A,
Department of CSE, Gudlavalleru Engineering College, A.P., India.

ISSN: 2278 – 7798