

GENERATION OF KEY DATA SECURE IN CLOUD SERVER (KEY D)

Het Rachhadiya , NAGA JITENDRA ¹ , ARUN M ²

¹ Final Year B. Tech - CSE, DR. MGR Educational and Research Institute, Chennai

² Professor, Department of CSE, DR.MGR Educational and Research Institute, Chennai

hetrachhadiya11@gmail.com , eswarguptha01@gmail.com ,
maddulanagajitendrah05@gmail.com, arasfr@gmail.com .

ABSTRACT :- With the massive increase in data stored in the cloud, deduplication which can reduce storage costs by allowing us to store only one copy of identical data—becomes more important than ever. Before outsourcing, they are often encrypted to protect data confidentiality. Data deduplication is hampered by traditional encryption, which necessarily leads to numerous distinct ciphertexts being generated using the same plaintext by various users' secret keys. Deduplication is made possible by convergent encryption since it automatically converts identical plaintexts into identical ciphertexts. How to manage a sizable number of convergent keys in a reliable and efficient manner is one associated issue. In order to solve the convergent key management problem, a number of deduplication approaches have been proposed. They must, however, either implement key management servers or mandate communication between involved parties. Using Attribute-based encryption, we create the unique client-side deduplication protocol KeyD in this study without the need for a separate key management server.

Users only communicate with the cloud service provider when downloading and uploading data. According to a security analysis, KeyD guarantees data confidentiality, convergent key security, and effectively protects ownership privacy all at once. A rigorous and in-depth comparison results reveals that our plan achieves a superior balance between the storage expense, communication expense, and computation expense. We eliminate duplication and enhance the administration of memory.

Index Terms: Data Deduplication, Attribute based Encryption, Memory Management and Security.

INTRODUCTION

The phrase "cloud computing" refers to a collection of several computing sectors that enables capital cost reduction and offers flexibility in terms of resource provisioning, release, and scaling up/down. With the help of cloud computing, customers can have unlimited access to storage space and data from any location at any time. Rather than being installed on the user's computer, programs are purchased and licenced across the internet, and only the actual usage is charged.

As internet users consume more information in the current day, demand for data storage is rising along with network traffic. Increasingly individuals and businesses are outsourcing their data to cloud services. Due to the data being stored in many locations, there is a significant amount of data duplication, which blocks memory by creating identical copies of the data. It was discovered that the majority of storage capacity is useless when redundant data is floating around taking up the most valuable storage and feigning a lack of space (1). A highly sophisticated method called as deduplication was introduced for reducing redundancy and cleaning the unwanted copies of identical data. In order to store only unique data copies, redundant data is eliminated using the data deduplication process. It results in better storage utilisation and is an effective method for handling similar information. Depending on the type of data, different data deduplication techniques are created. Text, image, and video are the three most frequently utilised data types. Each sort of data has its own distinct storage format and features. For each type of data, a distinct process is used by the deduplication technique to identify and get rid of duplicate copies. If the information is in the form of text, image, or video, it can be challenging to detect and compare. The deduplication procedure employs bit-level

representation. The bare minimum of data is the replication factor. The minimum quantity of identical data copies is known as the replication factor. Every piece of data is replicated by the cloud storage system (2). By creating a storage pool and sharing resources in a large-scale storage system environment, you can prevent different users from setting up their free storage space. Data de-duplication can drastically reduce the amount of backup data stored, as well as the amount of space, energy, and storage required. The top storage companies are releasing similar goods and services, like IBM's IBM System Storage TS7650G. Products from ProtecTIER that use data deduplication can use up to one-quarter of the actual storage capacity. Reduce total expenses and energy use by lowering the requirement for storage hardware. The NetApp V Series offers the removal of redundant data, which can lower the amount of data by at least 35%. The technology for deleting redundant data has evolved into more of a secondary storage system that performs filing and replicates the same material when numerous copies are deleted. The trend of the future will be the real-time deletion of duplicate data from the primary storage system (3). Many methods, including compression algorithms and deduplication, can be employed to increase storage capacity and lessen data replication in order to address this. Data mining and redundant

data detection techniques, when combined, can significantly lower an enterprise's storage needs. The data management businesses have demonstrated that efficient data management is achievable. Today, Dropbox is a provider of online storage that offers free online storage of sensitive data up to 2 GB. A new product named Google Drive, a search engine company, has recently been created. It offers free online storage of up to 5 GB. These businesses are able to do this by developing an effective file system for data deduplication and effectively managing data redundancy. The majority of storage solution providers, including NetApp and EMC, use data compression in their backup. This method is used to cut down on the quantity of data that is daily backed up and sent over the Internet from the client to the backup servers, saving bandwidth and money (4). The primary objective of this paper is to examine the effectiveness and efficacy of current deduplication methods. Deduplication is therefore necessary for the provider of cloud services to lower the high cost, vast volume of storage space demand, and greater network transfer rate. This publication also seeks to connect academics and industry professionals to share knowledge about various deduplication strategies.

LITERATURE REVIEW

Xuwei et al Ma present a novel server-

side deduplication strategy for encrypted data. Additionally, we employ an initial upload check method to ensure that only the first uploader needs to execute encryption and we use an access control technique that validates the legitimacy of the data users before they download data in order to reduce the communication overhead. Our server-side deduplication proposal provides superior performance compared to existing schemes in terms of security, efficacy, and practicability, as shown by our security study and performance assessment. However, our approach successfully withstands collusion and repeated fake attacks(5). Kanika et al Data are saved in an encrypted format using the Advanced Encryption Standard (AES) algorithm to ensure data confidentiality. It is feasible to execute data deduplication to save storage space by allowing data owners to share copies of the same data. The aforementioned problems have prompted research on encrypted data deduplication. In this publication, we provide an encrypted data de-duplication mechanism that enhances privacy protection by enabling the cloud storage server to get rid of duplicate cypher texts(6). Leesakul et al presents dynamic nature of cloud storage, data usage evolves over time. Some data chunks may be commonly read at one point in time but not at another. Although some datasets may need a high amount of redundancy due to dependability requirements, others

may be routinely accessed or modified by several users concurrently. Support for this dynamic feature in cloud storage is therefore essential. However, the majority of current techniques are static in nature, which restricts their full applicability to the dynamic nature of data in cloud storage. We provide a dynamic deduplication strategy for cloud storage in this paper with the goal of increasing storage effectiveness while preserving redundancy for fault tolerance (7). Lou et al source model that takes into account probabilistic replacements. With the information of repetition length, the performance of two modified forms of fixed-length deduplication is investigated and shown to fall within a constant factor of optimal. We also examine the variable-length approach, demonstrating that when entropy decreases, the compressed string's size diminishes in comparison to the length of the uncompressed string(8).Conde et al a brand-new different cluster duplication technique that outperforms existing methods while focusing on pivots to simplify calculation. To assess the performance of the algorithm, we offer a theoretical comparative study of the computational costs and experimental findings. The suggested deduplication method achieves the same rates as brute force or naive techniques while improving previous methods in terms of cost(9).

METHODOLOGY:

We suggest a brand-new deduplication strategy. To accomplish secure and effective convergent key management, we specifically combine convergent encryption and attribute-based encryption without the use of additional independent key management servers or reliable outside parties. Our system guarantees the security of convergent keys and the confidentiality of data files, according to security analyses. A thorough performance comparison between KeyD and various recent efforts is provided, demonstrating how well our method balances the costs of storage, communication, and computing.

EXISTING METHOD:

In existing system amount of data being saved in the cloud has dramatically increased, making deduplication—which allows us to store identical material in only one copy—unprecedentedly important. They are typically encrypted before outsourcing in order to guarantee data confidentiality. Data deduplication is hampered by traditional encryption since numerous distinct ciphertexts would inevitably be generated from the same plaintext using different users' secret keys. Since convergent encryption naturally converts identical plaintexts into identical ciphertexts, deduplication is made possible. How to effectively and reliably manage a sizable number of convergent keys is one of the accompanying issues. The convergent

key management problem has been addressed by a number of deduplication approaches. They must, however, either implement important management servers or mandate communication between data owners. Our work falls under the category of client-side compression, and it can operate without the help of key production server or other reliable outside parties both at the file-level and block-level. Therefore, the data in the cloud is not secure.

PROPOSED SYSTEM

Without a separate key management server, we create the unique client-side deduplication protocol KeyD for the proposed system by using attribute-based encryption. Users only communicate with the cloud service provider when downloading and uploading data. According to a security analysis, KeyD guarantees data confidentiality, convergent key security, and effectively protects ownership privacy all at once. A rigorous and in-depth performance comparison reveals that our plan achieves a superior balance between the storage expense, communication expense, and computation expense. The security study demonstrates that our KeyD effectively preserves user ownership privacy while also guaranteeing the

confidentiality of

Figure:1 System Architecture

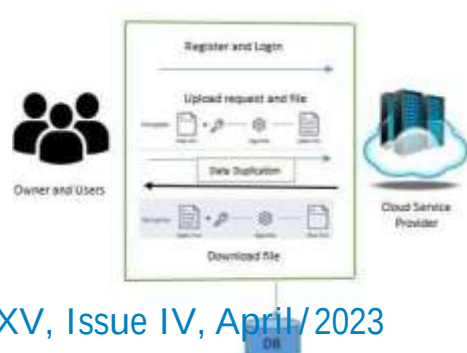
data and the security of convergent keys. A brand-new deduplication approach is what we suggest. To accomplish secure and effective convergent key management, we specifically combine convergent encryption and attribute-based encryption without the use of additional independent key management servers or reliable outside parties.

Advantages:

- Our system guarantees the security of convergent keys and the confidentiality of data files, according to security analyses.
- High performance and security are provided.
- Make a better balance the cost of storage, communication overhead, and computational overhead.

SYSTEM DESIGN:

In this Figure :1 shows our proposed system work. The user has access to a lot of data but typically has limited storage and computational power. He therefore desires to outsource the data to the cloud for subsequent access. Only the data that hasn't been stored in the cloud is uploaded by the user in order to conserve network capacity during the data upload phase. If duplication occurs, the CSP can be used to demonstrate



ownership, and it can also be used to update materials pertaining to convergent key attribute-based encryption.

MODULE DESCRIPTION:

In this project has two modules:

- Cloud Service Provider
- Users

Cloud Service provider:

Enter the correct username and password to get into the account. View every user and give them permission. View complete details for each user. View the status of every upload request. View each file that has been uploaded. To download the file, view the download request and send the key. view all file that was downloaded. Every result displayed is based on the graph.

Users:

Create an account with the necessary details to access the website. After receiving authorization from the cloud, the user can log in using the proper username and password. Views are the owners of their profiles, which they created while registering. Send the cloud service the upload file request. Uploading a file using the data owner's secret key requires the data owner. Delete the file if you no longer want it. Send the request for a download. As soon as you have the key, download it.

We don't keep files that are based on deduplication if they already exist.

RESULT AND DISCUSSION:

In order to lower the cost of disc energy consumption, data compression can reduce the number of discs utilized in the process. Remove duplicate data from large data center information technology system backup system a comprehensive, mature, safe and reliable, More environmentally friendly save the backup data storage technology solutions, has a very high value and great academic value., with very high application value and significant academic research value.

To compare the effectiveness of our new SW-PD to earlier efforts, we take the data deduplication ratio into account. The amount of data after deduplication is divided by this ratio to arrive at this ratio. based on the entire amount of data that could be backed up (i.e., the data that was checked for duplicates). Higher deduplication ratios are to be expected when there is more redundant data or when edit probability are lower.

In a more general setting than ours, an environment with mainly Windows servers and comparable files will result in far higher deduplication ratios than one with numerous platforms and

different operating systems.

SCREENSHOT

Home Page



Here we have all the modules to choose from like : Cloud Service Provider, Safe User

Admin Module



Admin Login page, where admin can enter the credentials to login and access the admin level controls.

Authorize



Authorize is to access user to use the account in our project.

View Users



Here admin can view the registered users and admin can view their details of the users.

Uploaded Files



View all uploaded files page to view the files which are uploaded by the users.

Download Key



Download key is to send key through mail to the respective user for to download their file after they request for to download.

Downloaded Files



In view download files page Admin can view who all downloaded the files.

User Data



User login page where user can enter

the credentials to login , and access the user level controls.

Profile



Profile page uses for to view the details of the user.

Send Upload Request



Send upload request page is to Upload the file to cloud, to upload the file have to send Request admin have to accept the request then the file uploaded successfully.

Manage File



User can Manage their file accordingly.

Send Download Request



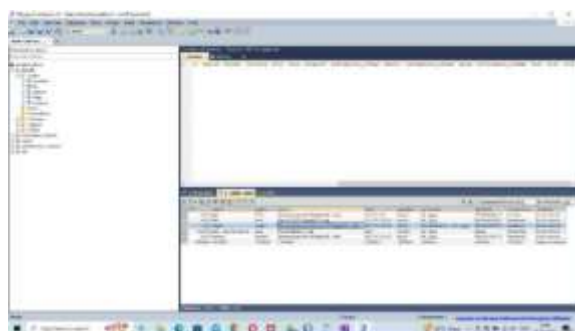
User has to send download Request to the admin from this page to get the generated key in their mail.

Key Mailer



The Key has send to User in mail, which has a unique keys. It can be used only once after that user has to request for another key from the admin.

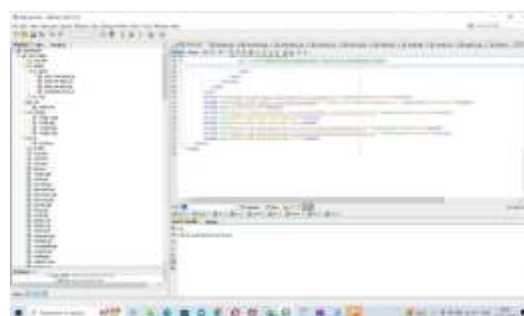
MySQL Database



For the Database, we have use My SQL yog software to store the data in a table

form.

Net Beans



For the implementation, we are using the NetBeans IDE 8.0.1 software for the Integrated development environment (IDE).

CONCLUSION:

In order to maintain convergent keys efficiently, we propose in this work the secure client-side deduplication technique KeyD. In our concept, data deduplication is accomplished by communication between data owners and the cloud service provider, without the involvement of other reliable parties or cloud service providers for key management.

The security study demonstrates that our KeyD effectively preserves user ownership privacy while also guaranteeing the confidentiality of data and the security of convergent keys. Experimental findings show that our scheme performs well in terms of security and memory management. In order to secure the identity privacy of data owners, which is not taken into account in our scheme, we will

endeavour to find solutions in our future work.

REFERENCE:

1. Chhabra, N., & Bala, M. (2018). A Comparative Study of Data Deduplication Strategies. 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC).
2. Vijayalakshmi, K., & Jayalakshmi, V. (2021). Analysis on data deduplication techniques of storage of big data in cloud. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC).
3. Qinlu He, Zhanhuai Li, & Xiao Zhang. (2010). Data deduplication techniques. 2010 International Conference on Future Information Technology and Management Engineering.
4. Manogar, E., & Abirami, S. (2014). A study on data deduplication techniques for optimized storage. 2014 Sixth International Conference on Advanced Computing (ICoAC).
5. Xuewei Ma, Wenyuan Yang, Yuesheng Zhu. (2022). A Secure and Efficient Data Deduplication Scheme with Dynamic Ownership Management in Cloud Computing. 2022 IEEE International Performance, Computing, and Communications Conference (IPCCC).
6. Kanika S. Gandhi, Devashree S, Garima M. (2020). Secure Encrypted Data Deduplication using Hashing Technique in Cloud . (2020) International Research Journal of Engineering and Technology (IRJET).
7. Leesakul, W., Townend, P., & Xu, J. (2014). Dynamic Data Deduplication in Cloud Storage. 2014 IEEE 8th International Symposium on Service Oriented System Engineering.
8. Lou, H., & Farnoud, F. (2020). Data Deduplication with Random Substitutions. 2020 IEEE International Symposium on Information Theory (ISIT).
9. Conde-Canencia, L., & Hamoum, B. (2020). Deduplication algorithms and models for efficient data storage. 2020 24th International Conference on Circuits, Systems, Communications and Computers (CSCC).
10. J. Liu, N. Asokan, and B. Pinkas, Secure Deduplication of Encrypted Data without Additional Independent Servers. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. ACM, 2015: 874-885.
11. M. Bellare, S. Keelveedhi, and T.

- Ristenpart, DupLESS: ServerAided Encryption for Deduplicated Storage[J]. IACR Cryptology ePrint Archive, 2013, 2013: 429.
12. M. Miao, J. Wang, H. Li, and X. Chen, Secure Multi-Server-Aided Data Deduplication in Cloud Computing[J]. Pervasive and Mobile Computing, 2015, 24: 129-137.
 13. C. Delerabl e, Identity-Based Broadcast Encryption with Constant Size Ciphertexts and Private Keys[C]. International Conference on the Theory and Application of Cryptology and Information Security. Springer, Berlin, Heidelberg, 2007: 200-215
 14. V. Javaraiah, "Backup for cloud and disaster recovery for consumers and SMBs," in Advanced Networks and Telecommunication Systems (ANTS), 2011 IEEE 5th International Conference on, 2011, pp. 1-3.
 15. L. L. You, K. T. Pollack, and D. D. E. Long, "Deep Store: An Archival Storage System Architecture," presented at the Proceedings of the 21st International Conference on Data Engineering, 2005.
 16. T. Yujuan, J. Hong, F. Dan, T. Lei, Y. Zhichao, and Z. Guohui, "SAM: A Semantic-Aware Multi-tiered Source Deduplication Framework for Cloud Backup," in Parallel Processing (ICPP), 2010 39th International Conference on, 2010, pp. 614-62.