Secure Data Protected Encryption Key For Deduplicated Cloud Storage System

Assistant Professsor Mrs.V.Hemalatha[1], S.Revathi[2], M.Santhosh[3], N.Suthan[4]
[1]Assistant Professor, Dept of Computer Science Engineering
[2, 3, 4]Dept of Computer Science Engineering
[1, 2, 3, 4] N.S.N College of Engineering and Technology, Karur, India,

Abstract- *Cloud computing has reached a maturity that leads it into a productive phase. This means that most of the main issues with cloud computing have been addressed to a degree that clouds have become interesting for full commercial exploitation. However, concerns over data security still prevent many users from migrating data to remote storage. Client-side data compression in particular ensures that multiple uploads of the same content only consume network bandwidth and storage space of a single upload. The intuition is that outsourced data may require different levels of protection, depending on how popular it is: content shared by many users. Then present a novel idea that differentiates data according to their popularity. Based on this idea, design a deduplicate scheme that guarantees semantic storage for unpopular data and provides weaker security and better storage and bandwidth benefits for popular data. This way, data de-duplication can be effective for popular data storage system. Deduplication technique helps to reduce storage requirements by similarity checking process. This allows organizations save far more data on the same system and extends disk purchase intervals automatically. With the advantage of speed, organizations can store data to disk cost effectively. This way, data de-duplication can be effective for popular data, whilst semantically secure encryption protects unpopular content. The backup recover system can be used at the time of blocking and analyze frequent login access system.*

Keywords- Cloud computing, Data de-duplication , Data Security , Semantic Storage , Remote Storage

I. INTRODUCTION

Cloud computing technology consists of the use of computing resources that are delivered as a service over a network. In cloud computing model users have to give access to their data for storing and performing the desired business operations. Hence cloud service provider must provide the trust and security, as there is valuable and sensitive data in huge amount stored on the clouds. There are concerns about flexible, scalable and fine-grained access control in the cloud computing.

Cloud computing is consistently growing and there are many main cloud computing providers including Amazon, Google, Microsoft, Yahoo and many others who are offering solutions including Software-as-aService (SaaS), Platform-as-a-Service (PaaS), Storage-as-a- Service and Infrastructure-as-a-Service (IaaS). In addition, considering the possibility to substantially minimizing expenses by optimization and also maximizing operating as well as economic effectiveness, cloud computing is an excellent technology. Furthermore, cloud computing can tremendously boost its cooperation, speed, and also range, thus empowering a totally worldwide computing model on the internet infrastructure. On top of that, the cloud computing has advantages in delivering additional scalable, fault tolerant services.

Cloud computing handles resource management in a better way since the user no longer needs to be responsible for identifying resources for storage. If a user wants to store more data they request it from the cloud provider and once they are finished they can either release the storage by simply stopping the use of it, or move the data to a long-term lower-cost storage resource. This further allows the user to effectively use more dynamic resources because they no longer need to concern themselves with storage and cost that accompany new and old resources.

Cloud computing service models are all inside in the cloud sing and laptops, desktops, phones and tablets are acts like clients to get services from the cloud. Cloud computing provides a shared pool of configurable IT resources on demand, in which needs minimal effort of management to get better services. Services are based on various agreement SLA (Service Level Agreement) between service providers and consumers.

II. LITERATURE REVIEW

"Investigating the adoption of hybrid encrypted cloud data deduplication with game theory" is a paper written by Liang et al. in 2020. The paper proposes a protocol called LEVER, which aims to handle encrypted data deduplication between users and cloud storage. The protocol involves three

phases: chunk key derivation, chunk encryption, and a deduplication protocol. LEVER uses symmetric cryptographic twoparty interactions to prevent brute-force attacks on the cloud storage. The paper proves the security of the protocol through ideal/real paradigms and analyzes its performance through simulation. However, client-side data deduplication can breach user privacy by creating a side-channel for attackers to gain information about file existence status.

Shen et al. proposed a cloud storage auditing scheme with deduplication supporting strong privacy protection in 2020. The proposed method generates a file index with an Agency Server, and the key for file encryption is generated using the file and a file label kept secret by the user. This protects the user's files' privacy from the cloud and Agency Server. The proposed scheme also allows users to generate the same ciphertext and authenticators for the same file, improving storage efficiency. The proposed scheme is secure, as demonstrated by the security proof, and requires lightweight computations from users to generate data authenticators, verify cloud data integrity, and retrieve their files.

III. EXISTING SYSTEM

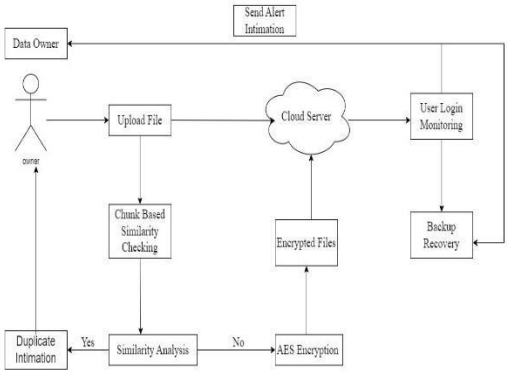
Currently, the most popular CDC approaches use Rabin fingerprints to determine chunk boundaries. However, this approach is timeconsuming as it computes and judges fingerprints byte by byte. To address this, other hash algorithms have been proposed. Fast CDC enhances and simplifies the judgment to reduce CPU operations during CDC. It pads zero bits into the mask value in its hash-judging statement to enlarge the sliding window size to the size of 48 Bytes used by Rabin-based CDC. Fast CDC also employs a normalized Content-Defined Chunking scheme to address cut-point skipping. However, it has some limitations such as only checking compression with file names, not content, and not achieving secure access control in a dynamic ownershipchanging environment. It may also lead to security degradation of the cloud service.

IV. PROPOSED SYSTEM

Introduction

This paragraph discusses the importance of compression for storage providers, and how it can be used to improve storage efficiency. There are four different compression strategies, and client-side compression is most beneficial. Security is also important, and the system must ensure data confidentiality for all files, including predictable ones. The server is trusted for user authentication and access control, but not for data confidentiality. The system must be designed to withstand attacks from potential attackers, including employees and the cloud storage provider. The system also implements a backup and recovery scheme, with alert systems and mobile notifications to ensure proper data recovery. Overall, this system provides improved compression and storage efficiency while maintaining data security.

System Architecture



Architecture Description: Data deduplication is a process by which a storage provider only stores a single copy of a file (or of its part) owned by multiple users. There are four different deduplication strategies, depending on whether deduplication happens at the client side (i.e. before the upload) or at the server side, and whether it happens at a block level or at a file level. Client-side data deduplication is more beneficial than server-side since it ensures that multiple uploads of the same content only consume network bandwidth and storage space of a single upload. Objective of this project is to reduce the amount of physical space consumed by removing identical blocks of data and using metadata to associate the logical copies of data to the physical ones. In the public cloud, the deduplication capabilities of the storage platform aren't exposed to the user. The architecture diagram shows the secure deduplication scheme to multiple types of data files at the time data storage and retrieval and using acknowledge system to know the status of login time. Data owner can be uploading the files with various file formats. And check the similarity of data using parity based similarity checking approach. File name and file content should be analysed. Here files are chunking and checked with database for redundancy prediction. If both the contents are same means, server rejects the files. Otherwise file encrypted using AES (Advanced Encryption Standard) algorithm. Server can implement self-destruction system to recover the data from blocked account and provide alert system at the time of recovering. User can

get all back up files in user manager mail with real time mobile intimation. Also provide acknowledgement about mail delivery.

Module List:

- Cloud Storage Framework
- De-duplication Checking
- File Encryption
- Storing in AWS by SQL Server
- Block Chain Security

Cloud Storage Framework:

Cloud computing and storage solutions enable users and businesses to store and process their data in data centers owned by either themselves or third parties, located at varying distances. Resources are shared to achieve coherence, with two types of users: data owners and providers. The cloud service owner, who can be the cloud consumer or provider, legally owns the service within the cloud. Cloud providers offer storage space to users, with multiple data owners sharing the space by uploading files for future use.

De-Duplication Checking:

In computing, data compression is a specialized data compression technique for eliminating duplicate copies of repeating data. Related and somewhat synonymous terms are intelligent (data) compression and single-instance (data) storage. In this module, we can check the files using file name with file contents. Encrypted files are spited into chunks. Service provider checks the chunks at the time of uploading files. Data owner only upload. Original file so save storage space in cloud system. We can compression in text file, document file and image files.

File Encryption:

Encryption is the most effective way to achieve data security. To read an encrypted file, you must have access to a secret key or password that enables you to decrypt it. Unencrypted data is called plain text ; encrypted data is referred to as cipher text. There are two main types of encryption: asymmetric encryption (also called public-key encryption) and symmetric encryption. We can implement symmetric encryption for encrypt the data files using single key approach. Symmetric key algorithms are algorithms for cryptography that use the same cryptographic keys for both encryption of plaintext and decryption of cipher text. The keys may be identical or there may be a simple transformation to go between the two keys. The keys, in practice, represent a shared secret between two or more parties that can be used to maintain a private information link. Encrypted data can be stored in cloud server.

Amazon Web Services:

Amazon Web Services offers a broad set of global cloud-based products including compute, storage, databases, analytics, networking, mobile, developer tools, management tools, IOT, security, and enterprise applications: on-demand, available in seconds, with pay-as-you-go pricing. From data warehousing to deployment tools, directories to content delivery, over 200 AWS services are available. New services can be provisioned quickly, without the upfront fixed expense. This allows enterprises, start-ups, small and medium-sized businesses, and customers in the public sector to access the building blocks they need to respond quickly to changing business requirements. This whitepaper provides you with an overview of the benefits of the AWS Cloud and introduces you to the services that make up the platform.

Block Chain Security:

Block chain network is initialized by the Block chain as a Service (BaaS) mode. This mode provides on-chain service through the Internet which means users do not need to consider the stability of the block chain itself or participate in the generation of blocks. This makes the security of block chain is depend on the whole network computational power of all participating nodes and we will discuss it in security analysis. It also allows every authorized participant joins the network and gains the block chain service at any time.

V. CONCLUSION

This Application includes distributed compression systems to improve the reliability of data while achieving the confidentiality of the users and also shared authority outsourced data with an encryption mechanism. Four constructions were proposed to support file-level and blocklevel data compression. The security of tag consistency and integrity were achieved. It Implemented our compression systems using the secret sharing scheme and demonstrated that it incurs small encoding/decoding overhead compared to the network transmission overhead in regular upload/download operations. Identified a new privacy challenge during data accessing in the cloud computing to achieve privacypreserving access authority sharing for similarity files.

REFERENCES

[1] Usharani, A. V., and Girija Attigeri. "Secure EMR Classification and Deduplication Using MapReduce." IEEE Access 10 (2022): 34404-34414.

[2] Zhang, Yucheng, Ye Yuan, Dan Feng, Chunzhi Wang, Xinyun Wu, Lingyu Yan, Deng Pan, and Shuanghong Wang. "Improving restore performance for in-line backup system combining deduplication and delta compression." IEEE Transactions on Parallel and Distributed Systems 31, no. 10 (2020): 2302-2314.

[3] Liang, Xueqin, Zheng Yan, and Robert H. Deng. "Game theoretical study on client-controlled cloud data deduplication." Computers & Security 91 (2020): 101730.

[4] Shen, Wenting, Ye Su, and Rong Hao. "Lightweight cloud storage auditing with deduplication supporting strong privacy protection." IEEE Access 8 (2020): 44359-44372.

[5] Liang, Xueqin, Zheng Yan, Robert H. Deng, and Qinghua Zheng. "Investigating the adoption of hybrid encrypted cloud data deduplication with game theory." IEEE Transactions on Parallel and Distributed Systems 32, no. 3 (2020): 587-600.

[6] Pooranian, Zahra, Mohammad Shojafar, Sahil Garg, Rahim Taheri, and Rahim Tafazolli. "LEVER: Secure Deduplicated Cloud Storage With Encrypted Two-Party Interactions in Cyber-Physical Systems." IEEE transactions on industrial informatics 17, no. 8 (2020): 5759-5768.

[7] Ni, Fan, and Song Jiang. "RapidCDC: Leveraging duplicate locality to accelerate chunking in CDC-based deduplication systems." In Proceedings of the ACM Symposium on Cloud Computing, pp. 220-232. 2019.

[8] Yu, Chia-Mu, Sarada Prasad Gochhayat, Mauro Conti, and Chun-Shien Lu. "Privacy aware data deduplication for side channel in cloud storage." IEEE Transactions on Cloud Computing 8, no. 2 (2018): 597-609.

[9] Yan, Zheng, tLifang Zhang, D. I. N. G. Wenxiu, and Qinghua Zheng. "Heterogeneous data storage management with deduplication in cloud computing." IEEE Transactions on Big Data 5, no. 3 (2017): 393-407.