Enhancing Cloud Storage Security in Big Data: A Comprehensive Data Sharing Protocol

Prof. Ashwini Bhosale[1] ,Sujay Bhor[2], Vipul Prayag[3], Sakshi Puri[4]
[1] *Professor, Siddhant College Of Engineering, Pune*
[2]*Student, Siddhant College Of Engineering, Pune*
[3] *Student, Siddhnt College Of Engineering, Pune*
[4] *Student, Siddhant College Of Engineering, Pune*

Abstract : A cloud based system gives advantage of a storage facility provided by a cloud service provider for exchange of data with authorised users. cloud providers store shared data in huge data centres outside the data owner's trust zone, which may raise the issue of data confidentiality. Number of parties, devices, app in cloud increases so number of access point also increases and making perfect access access control is challenging, so cloud data subject to being deleted or can be modified by cloud service provider. This article offers a secret sharing group key management protocol (SSGK) to prevent unwanted access in the communication process and shared data. in SSGK, group key is utilised to encrypt the shared data, and a secret sharing mechanism is employed to distribute the group key. security and performance evaluations show that our approach significantly reduces the security and privacy concerns of data sharing in cloud storage.

Keywords : Big data, Security and Privacy, Cloud Storage, Data Sharing, Encryption, Decryption.

1. INTRODUCTION

1.1 What is cloud?

Cloud computing is a revolutionary approach that transforms how computing resources are accessed, managed, and utilized. In this paradigm, data and programs are stored and accessed on remote servers hosted on the internet, rather than relying solely on local hard drives or local servers. It's like having a vast virtual warehouse of computing power and storage accessible via the internet.

1.2 What is Big Data?

Big data in this context refers to massive and complex datasets that are stored in the cloud. The project is specifically looking at security challenges associated with sharing this kind of data.

1.3 Why big data is relevant:

- **Volume, Variety, Velocity:** Big data is typically huge (volume), comes from many sources and formats (variety), and is constantly changing (velocity). Traditional security methods struggle to handle this complexity.

- **Sharing Challenges:** Securely sharing big data among authorized users is difficult because of the data's siz
e and the dynamic nature of cloud environments.

1.4 Cloud Storage and Big Data

A project is underway to develop a comprehensive security framework for cloud-based big data storage, focusing on secure data sharing among authorized entities. Traditional methods like access control and group key management have limitations, leaving data vulnerable to various threats, including those from cloud service providers. The project aims to overcome these limitations by integrating advanced encryption techniques, access controls, and authentication mechanisms. It seeks to ensure data security while enabling efficient sharing in the dynamic cloud environment. By addressing the unique challenges of big data, the project aims to instill trust in data sharing practices and uphold the highest standards of security and compliance. Through scalable and adaptable security measures, it aims to pave the way for secure data sharing in the era of cloud computing.

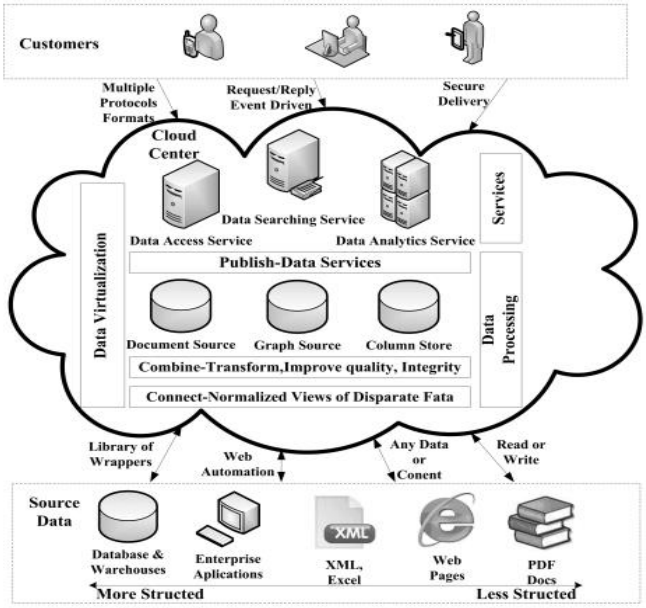

Figure 1. Cloud Storage architecture for Big data.

This cloud-based big data architecture visualizes how data is securely processed, stored, and shared. Raw data from various sources like databases and sensors enters the system. Wrappers translate the data formats for seamless integration. The cloud center, comprised of virtualized resources, stores and processes this data. Data services manage the flow: the access service retrieves data, the processing service cleans and analyzes it, and the analytics service generates insights. Data virtualization presents a unified view of this processed data, regardless of its origin. Finally, authorized users securely access and interact with the data through this architecture. This design facilitates secure data sharing and leverages the cloud's scalability for efficient big data management.

2. LITERATURE SURVEY:

1) **Paper Name**: Secured and Privacy-Preserving Multi-Authority Access Control System for Cloud Based Healthcare Data Sharing(2023)
Author Name: Reetu Gupta, Priyesh Kanungo,Nirmal Dagdee,Golla Madhu
Description: The cloud has become the obvious choice for data sharing. Generally, the data are outsourced to cloud storage servers in encrypted form. Access control methods can be used on encrypted outsourced data to facilitate and regulate access. Multi-authority attribute-based encryption is a propitious technique to control who can access encrypted data in inter-domain applications such as sharing data between organizations, sharing data in healthcare, etc. The data owner may require the flexibility to share the data with known and unknown users.

2) **Paper Name:** Cloud computing security: A survey of service-based models(2021).
Author Name: Fatemeh Khoda Parast, Chandni Sindhav, Seema Nikam
Description: In this study, we surveyed service-based cloud computing security issues to establish the current state of the field. The main contribution of this paper is to analyze the state of cloud security in the last decade and provide a unified taxonomy of security issues over the three-layer model.

3) **Paper Name**: Simple Approach to Realizing Verifiable Secret Sharing for Secure Cloud System (2022).
Author Name: Keiichi Iwamura, Ahmad Akmal Aminuddin Mohd Kamal
Description: In secret sharing system, we can split a secret into several parts (shares), and these parts are used to reconstruct the original secret, Now, there's a challenge when we use this in the cloud. If someone tries to cheat and submits fake shares during the reconstruction process, the reconstructed value may not match the original secret. Our method uses the shares generated during asymmetric secret sharing to reconstruct and verify the secret.

4) **Paper Name:** Security challenges and solutions using healthcare cloud computing(2021)
Author Name: Mohammad Mehrtak, Tayebeh Noori, Amirali Karimi, Omid Dadras
Description: Cloud computing is among the most beneficial solutions to digital problems. Security is one of the focal issues in cloud computing technology, and this study aims at investigating security issues of cloud computing and their probable solutions. Data encryption could be applied to store and retrieve data from the cloud in order to provide secure communication. Besides, several central challenges, which make the cloud security engineering process problematic, have been considered in this study.

5) **Paper Name:** A comprehensive survey of security, privacy, and trust issues in cloud storage for big data (2020).
Author Name: Khan, S., Khan, S. U., Zaheer, R., & Madani
Description: This survey paper presents a comprehensive overview of security, privacy, and trust issues in cloud storage for big data. It discusses data sharing protocols, encryption techniques, access control models, and trust management mechanisms to mitigate risks and ensure secure data sharing in the cloud. These papers provide a comprehensive understanding of the security and privacy challenges in cloud storage and present various data sharing protocols and techniques to mitigate these risks. They offer valuable insights and serve as a foundation for the development of an effective data sharing protocol to minimize security and privacy risks in the big data era.

6) **Paper Name:** How to share a secret (1979).
Author Name: A. Shamir
Description: Shamir's Secret sharing is a quantum attack proof algorithm and is used heavily for secret sharing. But it can also be used for authentication protocols as a replacement of hashing. In this paper, we propose an authentication protocol which will use Shamir's secret sharing method to authenticate with server. Hashing may not be able to hide data as effective in post quantum era. So in post quantum era, if any data server get exposed, users credentials can be also compromised as they were hidden by using hashing as an one way encryption. Our protocol will be able to solve this problem in a way that complete data exposure from server will not reveal the actual password provided by the user. So, even if the user uses same password for other online services/systems, these services and systems will not be effected.

7) **Paper Name:** Safeguarding cryptographic keys (1979).
Author Name: G. R. Blakley
Description: Safeguarding cryptographic keys is essential in ensuring the security and confidentiality of sensitive information in today's digital age. Cryptographic keys are the linchpin of encryption algorithms,

enabling the secure transmission and storage of data. This abstract explores the significance of cryptographic key protection and its role in safeguarding digital assets. It delves into key management best practices, including secure storage, access control, and periodic rotation. The abstract also highlights the potential consequences of key compromise, emphasizing the importance of proactive measures to prevent breaches. By prioritizing cryptographic key security, organizations can mitigate risks, protect their data, and maintain trust in an increasingly interconnected and vulnerable digital environment.

3. PROPSED SYSTEM:

3.1 Enhancing Healthcare

Cloud storage offers a cost-effective, scalable, and accessible solution for managing ever-growing medical records. This centralized approach benefits diagnosis and collaboration among healthcare providers. However, security risks abound. Sharing on public channels exposes data, and the numerous participants in the cloud environment increase the chances of unauthorized access by outsiders or even the cloud provider itself. Even authorized parties might conspire to steal records. Robust security measures are essential to safeguard patient privacy in e-healthcare.

3.2 SSGK:

SSGK (Secrete Sharing Group Key) introduces an effective solution for the secure sharing of data on cloud storage, obviating the need for reliance on any trusted third party. Beyond employing symmetric encryption for data encryption, SSGK integrates an asymmetric algorithm and a secret sharing scheme to thwart unauthorized users from accessing the key essential for decoding shared data. In the secret sharing system utilized by SSGK, a dealer partitions a secret into n shares, distributing them among n shareholders.

3.3 Model:

The protocol model consists of three types of entities: the cloud provider, the data owner, and the group members.

- **Cloud Service Provider**

  The cloud service provider offers a public platform for data owners to store and exchange encrypted data. Owners' data access is not controlled by the cloud provider. Any user can freely download the encrypted data.

- **Data Owner**

  Data owner establishes the access policy and encrypts its data using a symmetric encryption method and a group key. A sharing group is made up of group members who met the access policy. The owner then employs a secret sharing mechanism to deliver the encryption key to the sharing group.

- **Group Member**

Each member of the group, including the data owner, is issued a unique and a pair of keys.Members of the group can easily obtain any encrypted material from the public cloud that they are interested in. However, the user can only decrypt the data if and only if the data owner provides the data decryption key.

3.4 Working

This secret can subsequently be reconstructed using any t shares. Within SSGK, certain assumptions are made: the data owner is entirely trusted and impervious to corruption by adversaries, while the cloud provider is deemed semi-trusted. While the cloud provider accurately executes assigned tasks for profit, they may attempt to glean confidential information based on the data uploaded by the data owner. The distribution of the group key is facilitated by running

the secret sharing scheme, enabling group members to aggregate their respective sub-secret shares to reconstruct the group key.

To share data with group members, the owner initiates by generating a secret key, which is then employed to encrypt the data prior to its upload to the cloud. Subsequently, the owner shares the secret key with the group members. However, it's imperative to note that the decryption of data necessitates the collaborative effort of all group members for the confirmation and reconstruction of the key. This collaborative process ensures both the security and accessibility of the key solely to trusted group members.

In essence, SSGK offers a robust framework for secure data sharing in cloud storage environments, circumventing the reliance on potentially fallible third parties. By amalgamating symmetric encryption, asymmetric algorithms, and secret sharing schemes, SSGK provides a multifaceted approach to safeguarding shared data, bolstering trust, and ensuring confidentiality among group members.

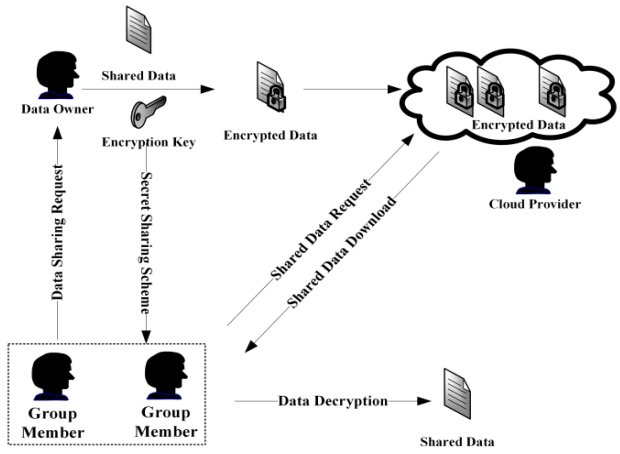

Figure 2. Sharing Data Protocol with SSGK

4. ALGORITHM

4.1 Encryption and Decryption

Encryption and decryption are the two fundamental processes involved in securing communication and data storage.

- **Encryption** is the process of transforming readable data (plaintext) into an unreadable form (ciphertext) using a specific algorithm and a secret key.
  - The data owner acts like a gatekeeper, deciding who can access the data.
  - They define an access policy that specifies authorized users.
  - The data owner encrypts the data using a symmetric encryption algorithm, which requires a secret key for both encryption and decryption. This secret key is called the group key.

- **Decryption** is the reverse process of encryption. It uses the same algorithm and the secret key to convert the ciphertext back into its original plaintext format.
  - A group member can only decrypt the data if they possess the entire group key.
  - Since the group key is split into sub-shares, a member needs to collect all the sub-shares assigned to them based on the access policy.
  - Once they have all the necessary sub-shares, they can combine them using the secret sharing scheme to reconstruct the original group key.

- Finally, the group member can use the reconstructed group key to decrypt the data and access its contents.

- **Secret Sharing:**

  - To distribute the group key securely, the data owner utilizes a secret sharing scheme.
  - This scheme splits the group key into multiple fragments (sub-shares).
  - Each sub-share is then distributed to a specific group member following the defined access policy.

4.2 AES:

- AES stands for Advanced Encryption Standard.
- It is a symmetric block cipher, which means the same key is used for both encryption and decryption.

In this data sharing model, AES is used for encrypting the actual data itself. The data owner (O) first creates a secret key and uses that key to encrypt the data with the AES encryption algorithm. This encrypted data, denoted as Cipher(D), is then uploaded to the cloud for storage.

It's important to note that AES operates on the actual data, not the secret key. The secret key is used to scramble the data into an unreadable format. AES itself is not involved in distributing the secret key.

- **AES Encryption Algorithm.**

  Cipher(D) = AESK (D).

4.3 RSA

- RSA stands for Rivest-Shamir-Adleman, which is a public-key cryptography system.
- It's one of the most widely used methods for secure communication and involves using a pair of mathematically linked keys: a public key and a private key.

In the context of the data sharing model you described, RSA plays a crucial role in securely distributing the secret key fragments (sub-shares) among participants.

In this data sharing model, RSA is used to securely distribute the secret key fragments (sub-shares) among the participants. Here's how it works:

1. **Sub-share Encryption:** The data owner (O) encrypts each sub-share (s1, s2, ..., sn) using the public key (PKi) of a specific participant (Pi). This public key is obtained from the cloud provider.
2. **Public Channel Distribution:** Encrypted sub-shares (Cipher(si)) are then sent over the public communication channel to their designated participants.

RSA ensures that only the intended recipient (Pi) can decrypt the sub-share using their private key. This protects the sub-shares from unauthorized users who might intercept them on the public channel. Even if an attacker manages to get hold of an encrypted sub-share, they wouldn't be able to decrypt it without the corresponding private key.

- **RSA Encryption Algorithm.**

  RSA - Cipher(si) = RSAPKi (si)
  Cipherj(si) = RSA−1 PKj (si)

Notation Table

NOTATION	DESCRIPTION OF NOTATION
Cipher	Cipher Text
Si	Sub Share of Key
Pki	Public Key of Participants
Ski	Secret Key of Pi.

5. CONCLUSION:

In this work, we present a unique group key management mechanism for cloud storage data sharing A data sharing protocol is essential in addressing security and privacy challenges in the era of big data and cloud storage. The Secret Sharing Group Key Management Protocol (SSGK) enhances data confidentiality, minimizes unauthorized access risks, and optimizes storage, marking a significant advancement in secure big data sharing in the cloud.

This research proposes a new group key management protocol (SSGK) for secure data sharing in cloud storage, specifically for big data. Traditional data security and privacy concerns are addressed with this protocol.

SSGK offers several benefits. First, it encrypts data and uses a secure key distribution mechanism to minimize unauthorized access. Second, it's designed to be storage-efficient, making it suitable for large datasets. Third, data owners retain control over their outsourced data without relying on a third party. Finally, a verified security scheme ensures that only authorized parties can access the data. The authors analyze potential security risks and defenses, demonstrating the protocol's strength. They also claim SSGK is less storage and computation intensive than existing solutions.

6. REFERENCES

[1] Reetu Gupta, Priyesh Kanungo,Nirmal Dagdee,Golla Madhu, “Secured and Privacy-Preserving Multi-Authority Access Control System for Cloud-Based Healthcare Data Sharing” in Sensors 2023, 23, DOI:10.3390/s23052617.

[2] Fatemeh Khoda Parast, Chandni Sindhav, Seema Nikam, “Cloud Computing Security: A Survey on Service-based Models” in December 2021, DOI:10.1016/j.cose.2021.102580

[3] Keiichi Iwamura, Ahmad Akmal Aminuddin Mohd Kamal, “Simple Approach to Realizing Verifiable Secret Sharing for Secure Cloud System” in IEEE (Volume: 10), 20 July 2022, DOI: 10.1109/ACCESS.2022.3192645

[4] Yabanku Nikhitha., S. Choudhaiah., “A Data Sharing Protocol To Minimize Security And Privacy Risks Of Cloud Storage In Big Data Era” in JES 08, August/2021(Voume:l 12)

[5] Mohammad Mehrtak, Tayebeh Noori, Amirali Karimi, Omid Dadras, “Security challenges and solutions using healthcare cloud computing” in J Med Life. 2021 Jul-Aug; 14(4): 448–461, doi: 10.25122/jml-2021-0100

[6] V. Casola, A. Castiglione, K. K. Choo, and C. Esposito, ‘‘Healthcare related data in the cloud: Challenges and opportunities,’’ IEEE Cloud Comput., vol. 3, no. 6, pp. 10–14, Apr. 2016.

[7] K. Xue et al., ‘‘RAAC: Robust and auditable access control with multiple attribute authorities for public cloud storage,’’ IEEE Trans. Inf. Forensics Secur., vol. 12, no. 4, pp. 953–967, Apr. 2017.

[8] Si Han ; Ke Han ; Shouyi Zhang, “A Data Sharing Protocol to Minimize Security and Privacy Risks of Cloud Storage in Big Data Era”, in IEEE (Volume: 7), 03 May 2019, **DOI:** 10.1109/ACCESS.2019.2914862

[9] H. He, R. Li, X. Dong, and Z. Zhang, ‘‘Secure, efficient and fine-grained data access control mechanism for P2P storage cloud,’’ IEEE Trans. Cloud Comput., vol. 2, no. 4, pp. 471–484, Oct./Dec. 2014

[10] H. liu, Y. huang, and J. K. Liu, ‘‘Secure sharing of Personal Health Records in cloud computing: CiphertextPolicy Attribute-Based Signcryption,’’ Future Gener. Comput. Syst., vol. 52, pp. 6.