Investigating Information Leakage in Multi-Cloud Environments: A Comprehensive Analysis

Prof. H. P. Bhabad[1], Mr. Pratham Ganesh Kotkar[2], Mr. Yash Pravin Dilwale[3], Miss. Rucha Manav Patil[4], Miss. Janhavi Anil Kolhe[5].

[1,2,3,4,5] Department of Computer Engineering Loknete Gopinathji Munde Institute of Engineering Education and Research, Nashik, India.

Abstract*:*

The increasing adoption of multi-cloud storage solutions has introduced new security challenges, particularly information leakage risks. This study conducts a comprehensive comparative analysis of information leakage risks in popular cloud storage services, including Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), Dropbox, and IBM Cloud. Our research investigates data breaches, side-channel attacks, and insider threats, evaluating the effectiveness of existing security mechanisms. Experimental results and survey findings reveal significant variations in information leakage risks across cloud services, highlighting trade-offs between security, performance, and cost. Our analysis provides valuable insights for cloud storage users, administrators, and providers, informing the development of more secure and efficient multi-cloud storage solutions [1].

Data leakage is a growing insider threat in information security among organizations and individuals. A series of methods have been developed to address the problem of data leakage prevention (DLP). However, large amounts of unstructured data need to be tested in the Big Data era. As the volume of data grows dramatically and the forms of data become much complicated, it is a new challenge for DLP to deal with large amounts of transformed data [4]. We propose an Adaptive Weighted Graph Walk model (AWG) to solve this problem by mapping it to the dimension of weighted graphs [11]. Our approach solves this problem in three steps. First, the adaptive weighted graphs are built to quantify the sensitivity of tested database on its context. Then, the improved label propagation is used to enhance the scalability for fresh data. We proposed a low-complexity score walk algorithm is to determine the ultimate sensitivity. This proposed method can detect leaks of transformed or fresh data fast and efficiently.

Keywords: ***Data Leakage prevention(DLP), Big data, Adaptive Weighed Graph (AWG), Multi-cloud, etc.***

I. Introduction

Data leakage is defined as the accidental or unintentional distribution of private or sensitive data to unauthorized entity. Sensitive data of companies and organizations includes intellectual property (IP), financial information, patient information, personal credit-card data, and other information depending on the business and the industry. Furthermore, in many cases, sensitive data is shared among various stakeholders such as employees working from outside the organizational premises (e.g., on laptops), business partners and customers. This increases the risk of confidential information falling into unauthorized hands [16]. Whether caused by malicious intent, or an inadvertent mistake, by an insider or outsider, exposed sensitive information can seriously hurt an organization.

In our system, we have assumed that distributor's sensitive data is stored in the form of "Relational Database System" and agent (trusted party) is going to request for original database say data object. Before requesting required data object, agent needs to Sign Up by filling registration form. While filling this registration form, unique ID is assigned to each registering agent by system itself. After Sign UP, agent can Sign In to send request for data object.

Unique ID assigned to an agent is used to create fake object (watermark)which is going to be embedded in requested data object and this modified Data Object and Fake Object is is given to an agent, So that even if this modified data is leaked by agent and similar data is found by distributor somewhere, then the distributor must assess the likelihood that the leaked data came from one or more agents.

The widespread adoption of multi-cloud storage services has revolutionized how organizations store, manage, and access data. By leveraging multiple cloud providers, businesses seek greater flexibility, redundancy, and the ability to avoid vendor lock-in. However, this multi-cloud strategy introduces significant security concerns, particularly related to information leakage—the unintentional or unauthorized disclosure of sensitive data.

While multi-cloud environments offer benefits like scalability and resilience, they also create complex security landscapes. The varied configurations, disparate security policies, and inconsistent encryption standards across cloud providers increase the potential for data vulnerabilities. This fragmentation in security measures provides more avenues for exploitation, amplifying the risk of information leakage [5].

The advent of cloud computing has revolutionized the way organizations store, process, and manage data. Multi-cloud environments, where multiple cloud services are used to optimize performance, cost, and scalability, have become increasingly prevalent. However, this trend has introduced new security challenges, particularly information leakage risks. Information leakage, the unauthorized disclosure of sensitive data, can have devastating consequences, including financial losses, reputational damage, and legal liabilities [6].

Despite the growing importance of multi-cloud security, research on information leakage risks in these environments remains limited. Existing studies focus primarily on single-cloud environments or specific leakage scenarios, neglecting the complexities and nuances of multi-cloud settings. This knowledge gap hinders the development of effective security measures, leaving organizations vulnerable to information leakage.

In this paper, we explore how certain aspects of multi-cloud storage services may inadvertently enhance information leakage. We examine the underlying factors, such as improper data partitioning, insufficient encryption, and lack of unified access controls, that contribute to this risk. Understanding these factors is crucial for developing strategies that can mitigate the dangers associated with multi-cloud data management while still benefiting from the advantages of this architecture.

II. Literature Review:

1. **(Hao Zhuang et al., 2022)**, Many schemes have been recently advanced for storing data on multiple clouds. Distributing data over different cloud storage providers (CSPs) automatically provides users with a certain degree of information leakage control, for no single point of attack can leak all the information. However, unplanned distribution of data chunks can lead to high information disclosure even while using multiple clouds. In this paper, we study an important information leakage problem caused by unplanned data distribution in multi cloud storage services.
2. **(Prashant et al., 2022)** We design an approximate algorithm to efficiently generate similarity-preserving signatures for data chunks based on Min-Hash and Bloom filters, and design a function to compute the information leakage based on these signatures. Next, we present an effective storage plan generation algorithm based on clustering for distributing data chunks with minimal information leakage across multiple clouds. Finally, we evaluated our scheme using two real datasets from Wikipedia and GitHub. We show that our scheme can reduce information leakage by up to 60-70 Percent.
3. **(ISHU GUPTA et al., 2022)** A large number of researchers, academia, government sectors, and business enterprises are adopting the cloud environment due to the least upfront capital investment, maximum scalability, and several other features of it. Despite the multiple features supported by the cloud environment, it also suffers several challenges. Data protection is the primary concern in the area of information security and cloud computing. Numerous solutions have been developed to address this challenge. However, there is a lack of comprehensive analysis among the existing solutions and a necessity emerges to explore, classify, and analyze the significant existing work for investigating the applicability of these solutions to meet the requirements. This article presents a comparative and systematic study, and in-depth analysis of leading techniques for secure sharing and protecting the data in the cloud environment.

4. **(G. V. Kapse et al.,2017)** access control is an efficient way to provide the data security in the cloud but due to data outsourcing over untrusted cloud servers, the data access control becomes

a challenging issue in cloud storage systems. Attribute-based Encryption (ABE) technique is regarded as a most trustworthy cryptographic conducting tool to guarantee data owner's direct control on their data in public cloud storage. The previous ABE schemes involve only one authority to maintain the complete attribute set, which can bring a single-point hindrance on both security and performance. Paper proposed the design, an expressive, efficient and revocable decentralized manner data access control scheme for multi-authority cloud storage systems, where there are multiple authorities exist and every authority is able to issue attributes independently.

III. PROBLEM STATEMENT

Information leakage in multi-cloud environments poses a significant threat to organizational data security, compromising confidentiality, integrity, and availability.

Despite the growing adoption of multi-cloud strategies, there is a lack of comprehensive understanding of information leakage risks, effective security mechanisms, and mitigation strategies. Existing research focuses primarily on single-cloud environments or specific leakage scenarios, neglecting the complexities of multi-cloud settings.

The leak of sensitive data on computer systems poses a serious threat to organizational security. Statistics show that the lack of proper encryption on files and communications due to human errors is one of the leading causes of data loss. But most of them worried about security; frequently they used to keep the data in single data server or data chunk. In this case if the data is lost or hacked in a sense entire data will be loose. To circumvent these kinds of vulnerabilities and to achieve better security we are offering of multi-instances data storage technique where the data will be stored in different databases or data chunks means instances.

Objectives

The research objectives are:

1. To identify and categorize information leakage risks in multi-cloud environments.
2. To analyze the effectiveness of existing security mechanisms in mitigating information leakage.
3. To evaluate the impact of multi-cloud configuration and data distribution on information leakage risks.
4. To develop a framework for mitigating information leakage in multi-cloud environments.

IV. PROPOSED SYSTEM ARCHITECTURE

Our goal is to detect and identify the information leakage data by intruders, and if possible to identify the agent that leaked the data information. Perturbation is a very useful technique where the data is modified and made "less sensitive" before being handed to agents. we proposed a model for assessing the "guilt" of agents and develop an algorithms for distributing objects to agents, in a way that improves our chances of identifying a leaked data [10,11]. Finally, we also consider the option of adding “fake” objects to the distributed set. Such objects do not correspond to real entities but appear realistic to the agents. In a sense, the fake objects acts as a type of security for the entire set, without modifying any individual members. If it turns out an agent was given one or more fake objects that were leaked, then the distributor can be more confident that agent was guilty [12].

Proposed Outcome:-

• Protect data leakage the user data during transmission.

• Stop the inside attack where the administrator of the user database reveals the sensitive users data.

• Security and privacy analysis has shown that our proposed new protocols are secure against both outside and inside attacks as long as one data server is not compromised.

V. System Architecture:

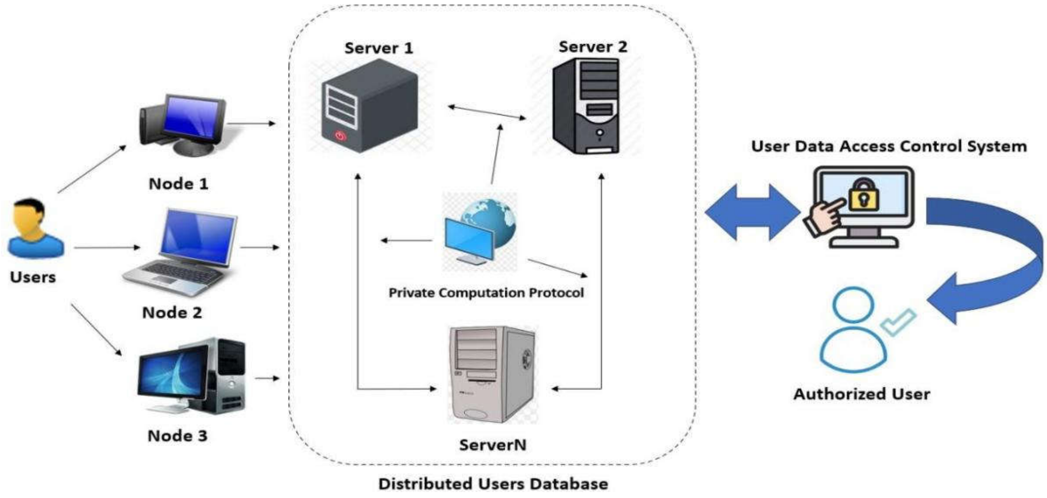
Server 1
Server 2
Node 1
Users
Node 2
Node 3
Private Computation Protocol
ServerN
Distributed Users Database
User Data Access Control System
Authorized User

Figure 1. System Architecture

In recent years, wireless network networks have been widely used in health- care, banking, educational, and industrial applications, such as hospital and home patient monitoring, and also sensitive user information. Wireless networks are more vulnerable to eavesdropping, modification, impersonation and replaying attacks than the wired networks. A lot of work has been done to secure wireless networks. The existing solutions can protect the user data during transmission, but cannot stop the inside attack where the administrator of the user database reveals the sensitive user data. In this proposed system, we propose a practical approach to prevent the inside attack by using multiple data servers to store user data. The main contribution of this system is securely distributing the user's or organization-related data in multiple data servers and employing the SHA crypto-systems to perform statistical analysis on the user/organizational data without compromising its privacy.

Nowadays, data security its miles been playing a important function in phrases of facts storing and lowering the overall price to entrepreneurs. But most of them involved approximately security; typically, they used to hold the information in a single data server example. In this example if the facts are lost or hacked inside the sense whole statistics can be loose. To keep away from these varieties of vulnerabilities and to achieve higher security we're proposing of multi data server or instances in which the information might be stored in one of a kind databases manner multi- instance.

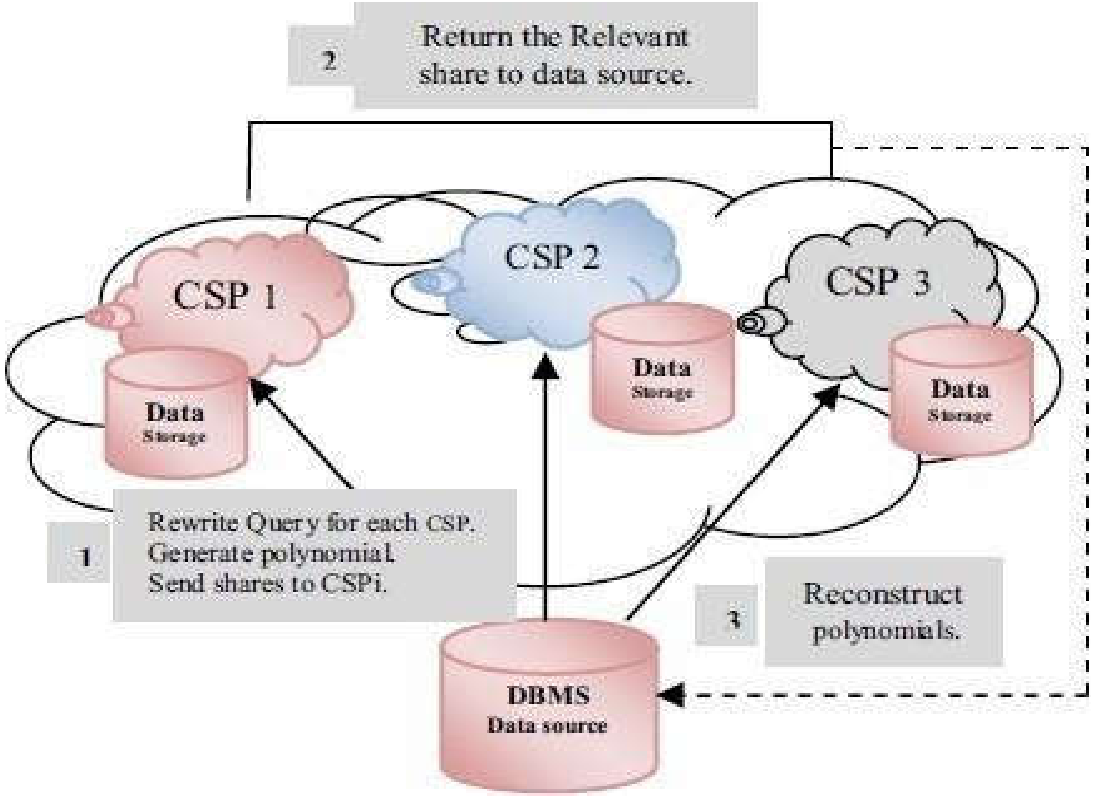

Figure 2. Database Architecture

On this proposed system we propose a two non-colluding server times to behavior a relaxed database carrier, in which the facts are stored in a single data server, at the same time as the expertise of the question sample is well partitioned into two parts, and understanding simplest one can't reveal any personal records. In this proposed work we're implementing the concept of a couple of data server storage together with improved security the usage of encryption techniques wherein instead storing whole document on single server system. The machine will encrypt and split the record in specific chunks and keep on special cloud. The data required for decrypting and rearranging that file may be stored in metadata management server for efficient retrieval and decryption of unique document. In this section, we describe the data flow from DBMS to the multi-instance providers in our proposed model. DBMS divides the data into n shares and stores each share in a different CSP.

VI. METHODOLOGY

Data computing and security is one of the inclining computing standards in which benefits of the computing framework which is given as a service over the Internet. This standard additionally delivers many people new tasks for data security and access control when clients outsource sensitive data for data servers, which are not inside the same trusted dominion as data possessor.

Many services like Email, Net Banking and so on are given on the Internet so that customer can make use of them from any place at any time they want indeed cloud storage is one of the adaptable services, how the security and protection are accessible for the export data that turns into vital concern.

A data User, such as an IT company, wants to outsource the database to server, which contains valuable and sensitive information (such as transaction logs, account information, medical information) and then access to the database. Due to the hypothesis that the service provider is honest, but curious, the server might try the best to get private information for its own benefits. Worse still, the server could convey such sensitive information to for-profit business competitors, which is an unacceptable operational risk.

Algorithm

1. AES:

Data Input and Key Initialization:

The diagram starts with the input data and the AES encryption key (which could be 128, 256 bits).

Data Segmentation:

The data is split into 128-bit blocks for AES encryption. Each block will undergo encryption independently, adding a layer of security.

Initial Round (AddRoundKey):

The initial round applies the XOR operation between the data block and the encryption key (shown as a lock symbol in the diagram).

Distribution Across Clouds:

The encrypted data blocks are then split and distributed across different cloud storage services, reducing the risk of leakage by storing only parts of the data in each cloud.

Decryption Process:

To decrypt the data, the encrypted blocks are retrieved from the clouds, and the AES decryption process (reverse of encryption) is performed to reconstruct the original data.

2. SHA-256:

Data Input:

The input is any data (e.g., files, messages) that you want to hash. This data will undergo hashing using the SHA-256 algorithm for integrity verification and security purposes.

Preprocessing (Padding and Parsing):

The input data is padded to ensure its length is a multiple of 512 bits. It is then split into 512-bit blocks, each processed individually by the algorithm.

Initial Hash Values:

SHA-256 starts with a set of initial hash values, a series of 8 predefined 32-bit constants.

Final Hash Generation:

After processing all blocks, the final hash is produced. This 256-bit hash is the unique "fingerprint" of the input data, which can be stored or distributed.

Multi-Cloud Storage:

The hash can be distributed across multiple cloud storage services to ensure data integrity. Whenever data is retrieved, the hash can be re-calculated and compared to ensure that no unauthorized modifications have occurred.

3. SHAMIR SECRETE ALGORITHM:

Input (Secret) -

The secret (e.g., an encryption key) is represented by the lock in the center. This is the data you want to protect and share across multiple cloud providers.

Polynomial Creation -

A polynomial is created based on the secret. The constant term of the polynomial (the a_0 coefficient) is the secret itself. The degree of the polynomial is determined by the threshold k (e.g., if you need 3 shares to reconstruct, it's a degree-2 polynomial).

Generating Shares -

Each share is generated by evaluating the polynomial at different points (e.g., x_1, x_2, ..., x_n). Each cloud provider gets a unique share.

These shares are shown as distributed across multiple clouds (in the diagram), ensuring that no single cloud provider holds enough information to recover the secret.

Share Distribution -

The shares are securely distributed across multiple clouds or entities. No single entity has enough information to recover the entire secret without the required number of shares (k out of n).

Secret Reconstruction -

When the secret is needed, k shares are gathered from different cloud providers, and Lagrange Interpolation is used to reconstruct the polynomial and recover the secret (represented by the lock).

VII. Future Scope

Develop and deploy a comprehensive data leakage prevention (DLP) solution for cloud environments. This project includes designing policies, implementing content inspection, user behavior analysis, real-time prevention, and encryption to proactively protect sensitive data and ensure compliance with regulations. It aims to reduce the risk of data breaches in cloud computing environments.

In conclusion, this research provides a crucial step towards securing multi-cloud environments against information leakage. By adopting our framework and recommendations, organizations can significantly reduce the risk of sensitive information exposure.

VIII. Conclusion

We have investigated the security and privacy as well as data leakage issues in the wireless network data collection storage and queries and presented a complete solution for the privacy-preserving wireless network. To secure the communication between user and data servers. To keep the privacy of the user's data, we proposed a new data collection protocol that splits the user's data into three numbers and stores them in three data servers, respectively. As long as one data server is not compromised, the privacy of the user's data can be preserved. For the legitimate user e.g. receiver to access the user's data, we proposed an access control protocol, where three data servers cooperate to provide the user with provide security to the user's data.

IX. References

[1] Hao Zhuang, Member, IEEE, Rameez Rahman,Pan Hui, Member, IEEE, and Karl Aberer, Member, IEEE, "Optimizing Information Leakage in Multicloud Storage Services", VOL.14, NO. 8,JANUARY 2022

[2] Prashant Shankar Kapse1, Onkar Dhananjaya Swami2, Dadaso Keskar3, Ruturaj Avinash Kadam4, Dr. S. N. Gujar5, "StoreSim: Optimizing Information Leakage in Multi-Cloud Storage Services", Volume 10 Issue V May 2022

[3] ishu gupta 1, (member, ieee), ashutosh kumar singh 2, (senior member, ieee),chung-nan lee1, (member, ieee), and rajkumar buyya 3, (fellow, ieee)," secure data storage and sharing techniques for data protection in cloud Environments: A Systematic Review, Analysis, and Future Directions", VOLUME 10, 2022

[4]Yong Zhang, "Data Leakage Detection System" – Aims to detect and respond to internal and external data leakage in multi-cloud environments,2019

[5] IRJET, "Minimizing Information Leakage in Multi-Cloud Systems using StoreSim" – Focuses on the MinHash algorithm to reduce data leakage,2021

[6] IJARIIE, "Enhancing Security in Multi-Cloud Environments with Privacy-Preserving Techniques" – Discusses the integration of differential privacy and secure data sharing

[7] L. Chen, "Secure Data Deduplication in Multi-Cloud Storage Systems" – Proposes deduplication techniques to reduce redundancy without compromising data security,2019

[8]Xiang Liu, "Multi-Cloud Data Integrity Verification" – Presents cryptographic methods for verifying data integrity across multiple cloud services,2020

[9] P. Belsis and G. Pantziou.A k-anonymity privacy-preserving approach in wireless medical monitoring environments. Journal Personal and Ubiquitous Com-puting, 18(1): 61-74, 2014

[10] D. He, N. Kumar, H. Wang, L. Wang, and K.-K.-R. Choo,''Privacy-preserving certificateless provable data possession scheme for big data storage on cloud,'' Appl. Math. Comput., vol. 314, pp. 31–43, Dec. 2017.

[11] W. Shen, J. Yu, H. Xia, H. Zhang, X. Lu, and R. Hao, ''Light-weight and privacy-preserving secure cloud auditing scheme for group users via the third party medium,'' J. Netw. Comput. Appl., vol. 82, pp. 56–64, 2017

[12] W. Shen, G. Yang, J. Yu, H. Zhang, F. Kong, and R. Hao, ''Remote data possession checking with privacy-preserving authenticators for cloud storage,'' Future Gener. Comput. Syst., vol. 76, pp. 136–145, Nov. 2017.

[13] R. Ding, Y. Xu, J. Cui, and H. Zhong, ''A public auditing protocol for cloud storage system with intrusion-resilience,'' IEEE Syst. J., vol. 14, no. 1,pp. 633–644, Mar. 2020, doi: 10.1109/JSYST.2019.2923238.

[14] D. He, S. Zeadally, and L. Wu, ''Certificateless public auditing scheme for cloud-assisted wireless body area networks,'' IEEE Syst. J., vol. 12, no. 1,pp. 64–73, Mar. 2018.

[15] J. Yu, K. Ren, and C. Wang, ''Enabling cloud storage auditing with verifiable outsourcing of key updates,'' IEEE Trans. Inf. Forensics Security,vol. 11, no. 6, pp. 1362–1375, Jun. 2016

[16] . K. Singh and I. Gupta, ''Online information leaker identification scheme for secure data sharing,'' Multimedia Tools Appl., vol. 79, no. 41, pp. 31165–31182, Nov. 2020.

[17] E. Zaghloul, K. Zhou, and J. Ren, ''P-MOD: Secure privilege-based multilevel organizational data-

sharing in cloud computing,'' IEEE Trans. Big Data, vol. 6, no. 4, pp. 804–815, Dec. 2020
[18] L. Zhang, Y. Cui, and Y. Mu, ''Improving security and privacy attribute based data sharing in cloud computing,'' IEEE Syst. J., vol. 14, no. 1, pp. 387–397, Mar. 2020
[19] W. Shen, J. Qin, J. Yu, R. Hao, and J. Hu, ''Enabling identity-based integrity auditing and data sharing with sensitive information hiding for secure cloud storage,'' IEEE Trans. Inf. Forensics Security, vol. 14, no. 2, pp. 331–346, Feb. 2019
[20] Gupta and A. K. Singh, ''A confidentiality preserving data leaker detection model for secure sharing of cloud data using integrated techniques, in Proc. 7th Int. Conf. Smart Comput. Commun. (ICSCC). Sarawak, Malaysia: Curtin Univ., Jun. 2019.
[21] G. V. Kapse1, Dr. V. M. Thakare2, Prof. S. S. Sherekar3, A. V. Kapse4," Multi-Authority Data Access Control For Cloud Storage System With Attribute-Based Encryption", IOSR- JCE[e-ISSN: 2278-0661],2017