Data Leakage Detection Using Cloud

Akshata Puthran[1], Vidhya Paradkar[2], Nikita Mandwal[3]
[1,2,3]*Department of Information Technology, Patkar Varde College*

***Abstract*— The transmission of data from an organization to unauthorized parties to external source is known as data leakage. This is a major problem faced by many people nowadays. The needs to be safeguarded and secured and provide the expected service to the user. We propose 'Data Leakage Detection using Cloud' that offers various advantages, such as the ability to leverage advanced technologies like Machine Learning and Big Data Analytics, and algorithms such as Intrusion Detection System (IDS), Anomaly Detection, Digital Rights Management and Cloud Access Security Brokers (CASB). Cloud-based data leakage detection systems can provide real-time alerts to notify users of any suspicious activities. This allows the user for immediate action to mitigate potential data breaches. These solutions can be easily integrated with existing IT infrastructure and applications, making it easier to implement comprehensive data leakage detection and prevention strategies. One can utilize machine learning, and AI algorithms to analyze vast amounts of data and identify patterns that may indicate data leakage.**

***Index Terms*—Anomaly Detection, AI algorithms, Cloud, Data Leakage, Encryption, Intrusion Detection System (IDS), Machine Learning.**

I. INTRODUCTION

Cloud Computing is a software where data and programs are stored and accessed on distant servers hosted online rather than on a local server or the computer's hard drive. Another name for cloud computing is Internet-based computing, which is a technology that allows users to access resources as a service via the Internet. It offers flexibility, scalability, and cost- efficiency by providing on-demand services. Nowadays, it is adopted by every company, whether it is an MNC or a startup. Cloud computing leverages powerful data centers that host various resources, making them accessible to users globally. Due to increasing use of cloud- based services data privacy and security becomes mandatory. The primary concern regarding cloud services is the risk of data leakage.

Data leakage, or data leaking, refers to the exposure of sensitive information to cybercriminals. Such information can either be personal or corporate or organizational in nature. The leakage is usually through the internet or by email, but it can also be done physically through gadgets like laptops and other hardware or through storage like USBs and external hard drives. For an individual this can be very devastating personally and for an organization it can lead to large scale effects like damaging public reputation and massive financial losses.

To solve this problem, a number of methods for identifying data leaks in cloud-based systems have been developed. But putting many of these strategies into practice is either too expensive or too complicated. In order to detect data leaks, this study suggests a new method that makes use of machine learning algorithms and cloud computing. In order to spot trends and irregularities that can point to a data breach, the suggested system can track and examine cloud data traffic. By doing this, the system is able to identify data leaks rapidly and notify system administrators in a timely manner.

II. MODULES USED AND PREVENTION

"Data Leakage Detection Using Cloud Computing" is made up of a number of components that collaborate to offer an end-to-end solution for data leakage detection and prevention in cloud computing environments.

1) Data Classification: The first part of the system is to classify data according to its sensitivity level. The system applies machine learning algorithms to learn the data and classify it into various categories according to its sensitivity level. The first part of the system ensures that sensitive data is properly secured and only accessed by authorized users.

2) Access Control: The second element of the system includes access control policies. Access control policies are used to limit user access to sensitive information. The system uses role-based access

control policies to allow only approved users to access sensitive information.

3) Monitoring Mechanism: Monitoring of user behavior is the third aspect of the system. User activities are tracked by the system with the aid of machine learning algorithms. These algorithms track the user behavior to detect any odd patterns that signal probable data leakages. Anomalies in the user activity are detected and responded to effectively to avoid any data leakage.

4) Cloud-Based Storage and Computing: The fourth element of the system is employing cloud-based computing and storage capabilities. The system employs cloud-based storage to place large amounts of data and cloud-based computing to process data in real-time. This element facilitates scalability and adaptability to the system, and it ensures the system can accept large amounts of data and cater to evolving users' needs.

5) Reporting and Alerting: The final component of the system is reporting and alerting. The system provides reports of user behavior and patterns of data access, allowing administrators to track the performance of the system. The system also alerts administrators upon detection of possible data leakage, allowing them to respond immediately.

The whole system operates in phases. The primary objective is to attain data integrity and duplication in the cloud. Integrity audit. The first design objective of this work is to offer the ability to verify the correctness of data stored remotely. The integrity check also needs two features:

a. Public verification, which enables any individual, not just customers who initially submitted the file, to conduct the verification;

b. Stateless verification, which is capable of reducing the necessity for the preservation of status information on the verifier side between data storage and audit actions.

Delegation of data auditing is also central to the discussion of cloud computing and data leakage detection. Auditing is verification and validation of data in order to maintain accuracy, completeness, and security of data.

Third-party auditors are an important aspect of the subject of data leakage detection through cloud computing. Cloud service providers can provide varying levels of security controls and measures to protect data confidentiality and integrity, but it is hard for cloud users to independently validate these assertions. Third-party auditors offer an independent evaluation of the security posture of the cloud service provider and assist cloud users in assessing the security of their data in the cloud.

III. POSSIBLE CLOUD ATTACKS

Sr.	Cloud Attacks	
	Attack Name	Attack Process
1	CrossVM attack - Attacks on Confidential Cloud data	Uses Side channels and attacker places the malicious Virtual Machine on the Clients physical server location.
2	Malicious SysAdmin Attack on Confidential Cloud data	Authorized system admin on the cloud provider may access customer memory of users VM.
3	Data loss or Manipulation Attack-Attack on Cloud Data Integrity	Due to large amounts of cloud storage in distrusted servers. Attacks can be initiated using data migrations, loss of data by taking advantage of cloud owners' loss of control.
4	Dishonest Computation – Attack On Cloud Data Integrity	Due to mis- configurations, outdated and vulnerable code at cloud servers.
5	Flooding attack - Attack on Cloud Data Availability causing to Denial of service attacks	Due to huge number of requests makes the service unavailable.
6	SLA violation Attack on	Due to improper allocation of resources to cloud customer by cloud provider.

IV. ALGORITHMS USED

1) *K-anonymity Algorithm:* Wakhare Yashwant R (2016) presented an approach for detecting data leakage in clouds. “K-anonymity Algorithm” is employed to eliminate the disadvantage of the watermarking method and the generation of fake objects, data allocation, employing guilt probability model to identify the guilty agent and generating sensitive data employing k-anonymity algorithm.
2) *SHA Algorithm:* Visnu Dharsini, Mrinal Pramanik (2019) presented an approach of SHA. SHA stands for Secure Hashing Algorithm. Hashing algorithms are mathematical functions, used to compress and encrypt the data passed as input to it and produce some outputs of seemingly random values known as hash or hash values. These

random values are actually encrypted or coded form of the input data. Computers prefer to use the hash values of data instead of the actual data because hashes allow it to be easy for the computer to do multiple operations or calculations on files and strings of data.

3) *Advanced Encryption Standard (AES):* Riya Naik.et.al (2019), in the paper titled "Data Leakage Detection in cloud using Watermarking Technique" presented an approach for detecting data leakage in clouds. For detection purposes, the algorithm is referred to as "Advanced Encryption Standard (AES)". A frequency domain approach is employed in the watermarking algorithm, which makes it more efficient and robust. It is mostly information retrieval and encryption that make up the data transfer phase. To identify the data uniquely, information is collected about it. Once the information is retrieved, the message is generated based on that information and the client ID of the receiver. Various steps are involved in the generation of QR codes.
4) Data allocation: Sushil Kumar N. Holambe 2015 Data allocation with guilt probability calculation The research presents a method based on sending imaginary objects to identify guilty agents and then calculating the guilt probability to detect the leakage.

V. ADVANTAGES

1) *Automated controls*: Sensitive data can be encrypted, blocked, or prompted automatically via cloud computing data leakage protection.
2) *Frequent audits:* At predetermined times, cloud computing data leakage protection can audit and scan data stored on the cloud.
3) *Resources and infrastructure:* Cloud servers offer the infrastructure and resources required for data processing and analysis, which can aid in identifying and stopping data leaks.
4) *Security standard compliance:* Cloud computing data leak prevention can assist businesses in keeping data visible and controls in place to meet security standards.

VI.DISADVANTAGES

1) *Dependency on Internet Connectivity:* Continuous internet access is required for cloud-based detection systems to function effectively. Any disruption in connectivity can impact the monitoring and detection process.
2) *Data Privacy Concerns:* Storing sensitive data in the cloud can raise privacy issues. Users must trust that their cloud service provider is adequately securing their data.
3) *Regulatory Compliance:* Different countries have varying data protection laws. Ensuring compliance with these regulations can be complex when data is stored and processed in the cloud.

VII. SURVEY ANALYSIS

1) Cloud Adoption: High adoption rate, with 64.3% of users utilizing cloud services.
2) Security Concerns: Mixed opinions about the overall security of cloud platforms.
3) Common Security Measure: Multi-factor authentication (MFA) is widely used for enhanced protection.
4) Security Audits: Regular security audits need improvement to address potential vulnerabilities.
5) Preferred Provider: Google Cloud is the preferred choice for detecting data leakage.

In Summary, Ensuring cloud security requires vigilant practices, including adopting MFA and conducting frequent security audits to reduce risks effectively.

VIII. CONCLUSION

In conclusion, securing sensitive data requires the use of cloud computing for data leak detection. Scalability is provided by cloud platforms, however security issues like data mobility and multi-tenancy are also introduced. Organizations can successfully identify and stop data leaks by utilizing cutting- edge strategies including encryption, real-time monitoring, and machine learning. To guarantee the confidentiality and integrity of data in the cloud, users and providers of clouds need to collaborate and enforce robust security. Causal Productions has attempted to the best of its abilities to make sure that the templates appear in the same way.

ACKNOWLEDGMENT

I would like to sincerely acknowledge and thank Patkar Varde College for providing me with the opportunity to undertake this research. We are

particularly grateful to the faculty and staff for their guidance, encouragement, and expertise, which greatly contributed to the depth and quality of this research. This opportunity has not only enriched my academic experience but also significantly enhanced my research skills, for which I am truly appreciative.

REFERENCE

[1] Volume 11 Issue VMay2023doclink: https://doi.org/10.22214/ijraset.203.52273
[2] Bassim M. Saliha, Omer K. Jasim Mohammad * a Computer Center, University of Fallujah, 60 street, Fallujah 31002, Iraq, Iraq. b Quality Assurance and Accreditation, University of Fallujah, 60 street, Fallujah 31002, Iraq, Iraq, "Cloud Data Leakage, Security, Privacy Issues and Challenges.
[3] Wakhare Yashwant R et al,/(IJCSIT) International Journal of Computer Science and Information Technologies, Vol.7 (4), 2016, 1911-1915
[4] S.Visnu Dharsini, Mrinal Pramanik, Sanidhya Gupta, Saurav Pahadiya. International Journal of Scientific & Technology Research Volume 8, Issue 11, November 2019.