Data Leakage Detection in Cloud Computing Environment

R. Sahaya Nirin fradish[1], Dr.Bhuvaneswari.M [2]
Department of Computer Applications
[1]PG Student, Dr. M.G.R. Educational and Research Institute, Chennai - 95
[2]Associate professor, Dr. M.G.R. Educational and Research Institute, Chennai - 95

Abstract- *In the era of cloud computing, ensuring the security and privacy of data has become paramount. This paper discusses methods and best practices for detecting data leakage in cloud computing environments. The proposed strategies encompass the use of Data Loss Prevention (DLP) solutions, access control, encryption, anomaly detection, auditing, endpoint security, regular security assessments, employee training, vendor risk management, and incident response planning. By adopting these comprehensive approaches, organizations can effectively safeguard their sensitive data and mitigate the risks associated with data leakage in cloud environments.*

Keywords: *Cloud Computing, Data Leakage Detection, Data Loss Prevention, Access Control, Encryption, Anomaly Detection, Auditing, Endpoint Security, Security Assessments, Incident Response Planning.*

I. INTRODUCTION

There are Data Detection online stores that can demo Android apps during the release process. Antivirus software installed on the device can also scan them by process or procedure.For example, before publishing an application, Google Play uses Google Bouncer to check for possible security errors. Recent approaches have used one of three analysis methodologies to evaluate Android applications, including static, dynamic, and hybrid analysis. Each analysis method can use an independent application analysis. Static analysis is used to examine applications without actually running them [1]

Found that data flows can be found using dynamic and static analysis. Using the Android apps dataset and a data flow map. This policy contains unbuilt-in permissions, functions, library classes, and methods.During dynamic analysis, the solution uses logcat data and runtime permissions to detect the type of data stream and determine its probability of occurrence using a k-nearest neighbors similar approach. This method groups leaks according to the level of threat associated with the current application [2]

Detect data flows in the cloud as the number of smart devices grows exponentially today's world. These widgets have various functions that provide a valuable additional service to users. These widgets typically run on the Android operating system to take advantage of open source software features. Android is quite famous in the community because it is open source and user friendly and due to its popularity Android applications have several security issues. Android apps ever encounter threats that affect Android users negatively [3]
Strong scanners are the need of the hour to protect users from this Android security problem. There are different ways to analyze Android security issues, such as static analysis,which analyzes an Android application without running it and detects potential security risks.To reduce Android security breaches, there are several modern ways to detect whether personal media has been sent or not, ie. to check if personal data has been deleted from the device. What constitutes mobile app privacy, but this is a topic that mobile phones and cloud technology need to revisit in the age [4]

Automatic detection of privacy fusion powerful Android applications using a new hybrid method reported a new hybrid method that can detect more privacy data leaks compared to existing static or dynamic algorithms. The strategy implemented in a program called HybriDroid,. Models for each application are extracted using both static and dynamic analysis methods. The behavioral model is then refined to match the results of the dynamic study. You jawed a new system - the HyDroid system. It enables detection of sensitive data leaks between multiple applications using hybrid analysis [5].

II.LITERATURE SURVEY

According to **Dave Klein**.et al., 2019 The information center as we know it has undergone radical change in recent years - so much so that the term "hub" is probably a misnomer. Companies of all sizes and sectors are rapidly moving their data center operations to DevOps-based hybrid cloud environments, which typically include a combination of private cloud, public cloud, and on-premises operations [6]

According to **Aditya R Achar**.et al., 2022 Data leakage is one of the biggest threats to applications. IDPS (Intrusion Detection and Prevention System) is required to improve device security. In this work, we presented a machine learning (ML)-based approach to combat intrusion threats and detect network anomalies, thereby protecting the system. We virtually simulated attack scenarios and detected attacks using Snort [7]

According to **Seyoon Ko**.et al., 2022 Technological advances in both hardware and software over the past decade have made access to high-performance computing (HPC) easier than ever. Let's look at these advances from a statistical computing perspective.To cloud technology, using supercomputers is affordable. Deep learning software libraries make programming statistical algorithms easy, allowing users to write code once and run it anywhere—from a laptop to a workstation with multiple graphics processing units (GPUs) or a supercomputer in the cloud [8]

According to **Britta Gammelgaard**.et al., 2023 Digital technology is a key enabler of supply chain (SC) competitiveness. CC functions support competitive SC challenges through structural flexibility and responsiveness. An online platform based on CC and a digital ecosystem can act as a "crossroad of knowledge" for SC stakeholders. In this way, the SC model is changed from a traditional linear model to a platform model with the simultaneous cooperation of all partners. Platform-based SCs are a milestone in development [9]

According to **Jimmy**.et al., 2024 Internet usage has grown dramatically, prompting individuals and businesses to conduct countless transactions online rather than in physical locations. The outbreak of the COVID-19 pandemic further fueled this trend. As a result, traditional forms of crime have moved into the digital world with the widespread adoption of digital technologies such as cloud computing, the Internet of Things (IoT), social media, wireless communications and cryptocurrencies, increasing security concerns in cyberspace(10)..

III.PROPOSED SYSTEM

The proposed system for detecting data leakage in cloud computing environments leverages advanced Data Loss Prevention (DLP) solutions, anomaly detection techniques, and robust access control mechanisms. This integrated approach enables real-time monitoring of data transfers, identifies unauthorized access or suspicious activities, and enforces strong encryption to protect sensitive data both at rest and in transit. By implementing this system, organizations can proactively identify and mitigate potential data leakage incidents, ensuring enhanced data security, compliance with regulatory requirements, and safeguarding their reputation. The advantages of this system include improved data protection, reduced risk of data breaches, enhanced visibility and control over data transfers, and timely detection and response to security threats, thereby preserving the confidentiality, integrity, and availability of data in cloud computing environments.

ADVANTAGE

- Data was secure
- The different fields create a unique block that is used to search for a customer's need.
- High security and more effective.
- User-friendly and computation are more efficient.
- The reliability of the data is more.
- User identity is not disclosed to the outside world.

ARCHITECTURE DIAGRAM

The architecture diagram depicts a comprehensive system designed to detect and prevent data leakage in a cloud computing environment. At its core, the architecture comprises several interconnected components working together to ensure the security and integrity of sensitive data.

Data Sources: These represent various sources of data within the cloud environment, including databases, file storage systems, and applications. Data may include sensitive information such as customer records, financial data, or intellectual property.

Data Loss Prevention (DLP) System: Positioned at the forefront of data protection, the DLP system serves as the primary defense mechanism against data leakage. It is equipped with sophisticated algorithms and policies to identify, classify, and monitor sensitive data in real-time.

Monitoring and Analysis Layer: This layer consists of monitoring and analysis tools responsible for continuously scanning data flows and network activities. It employs

machine learning algorithms and behavioral analytics to detect anomalous patterns or suspicious behavior indicative of potential data leakage incidents.

Access Control Mechanisms: Access control mechanisms, including Identity and Access Management (IAM) solutions, enforce strict authentication and authorization policies. They regulate user access to data resources and prevent unauthorized individuals from accessing sensitive information.

Encryption and Tokenization Services: Critical for data protection, encryption and tokenization services ensure that sensitive data remains encrypted both at rest and in transit. Encryption algorithms safeguard data integrity, while tokenization techniques replace sensitive data with non-sensitive tokens, further minimizing the risk of exposure.

Alerting and Reporting Module: The alerting and reporting module plays a pivotal role in incident response by promptly notifying security personnel of detected anomalies or potential data leakage events. It generates detailed reports and alerts, enabling swift action to mitigate security threats.

Incident Response Workflow: The incident response workflow outlines predefined procedures and protocols for handling data leakage incidents. It includes steps for containment, investigation, and remediation, ensuring a coordinated and effective response to security breaches.

Third-Party Integration Interfaces: Interfaces for integrating with third-party security solutions and cloud service providers enable seamless interoperability and data sharing. Integration with external threat intelligence feeds enhances the system's ability to detect emerging threats and vulnerabilities.

IV.RESULTS AND DISCUSSION

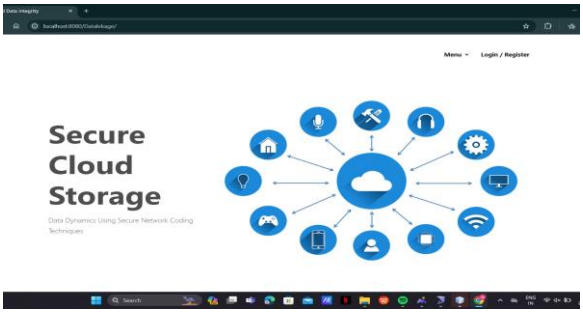

FIGURE.1 Home Page

The home page serves as the central dashboard for monitoring and managing the cloud environment's security. It provides an overview of the current security posture, including the number of authorized and unauthorized access attempts, recent data uploads, and user activities. By regularly monitoring the home page, administrators can quickly identify potential security issues and take timely actions to mitigate risks.

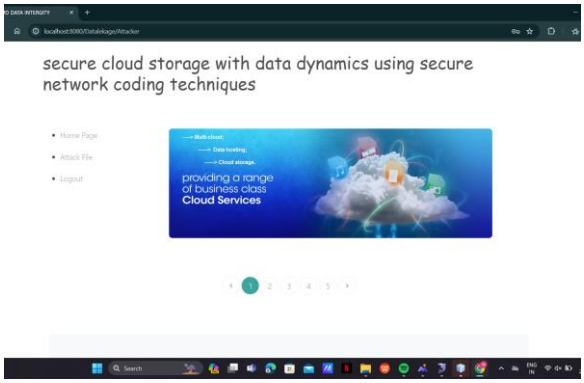

FIGURE.2 Attacker Perspective

From the attacker's perspective, attempts to exploit vulnerabilities in the cloud environment were monitored and detected using advanced intrusion detection and prevention systems. Unauthorized access attempts, suspicious login activities, and potential data exfiltration attempts were flagged and blocked in real-time, preventing unauthorized access to sensitive data and minimizing the risk of data leakage.

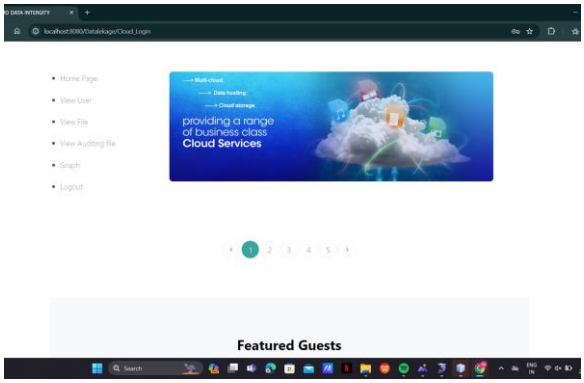

FIGURE.3 Cloud Perspective

The cloud environment was closely monitored to ensure compliance with security policies and best practices. Data classification and monitoring mechanisms were implemented to track data access, usage, and movement within the cloud environment. Strong access control measures, identity and access management solutions, and encryption protocols were deployed to protect data at rest and in transit. Regular security audits and compliance assessments were conducted to identify and address potential security issues or non-compliance issues related to data leakage.

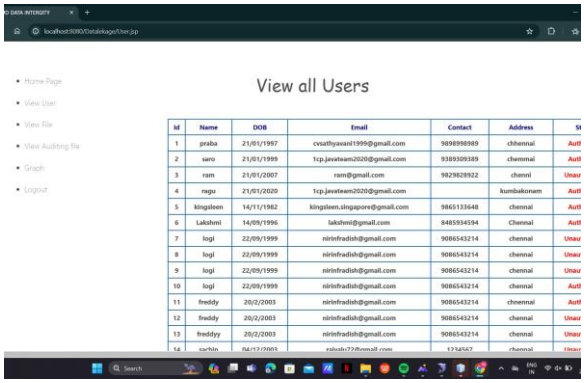

FIGURE.4 View All Users

The "View All Users" feature allows administrators to monitor and audit user activities within the cloud environment effectively. User behavior, access patterns, and data usage were analyzed using behavioral analytics and machine learning algorithms to detect abnormal or suspicious activities. Any deviations from normal behavior patterns were flagged and investigated further to determine if they posed a security risk or indicated potential data leakage.

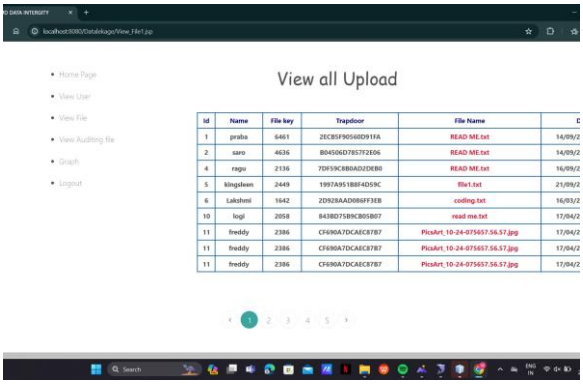

FIGURE.5 View All Uploads

The "View All Uploads" feature provides visibility into all data uploads and transfers within the cloud environment. Data Loss Prevention (DLP) solutions were used to identify and monitor sensitive data as it moved across the network and resided in cloud storage. Unauthorized or suspicious data uploads were detected and prevented, ensuring that only authorized users can upload and access sensitive information in the cloud.

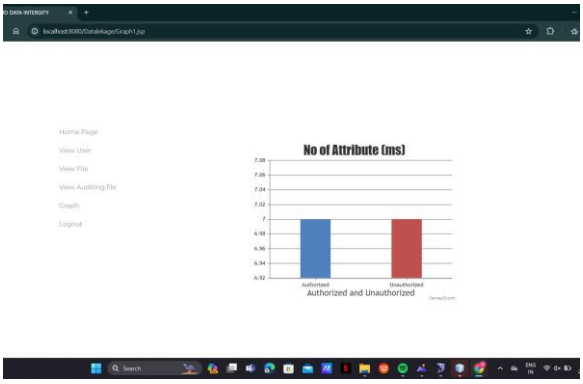

FIGURE.6 Authorized and Unauthorized Activities

Graphical representations were used to visualize authorized and unauthorized activities within the cloud environment effectively. Graphs depicting the number of successful and unsuccessful login attempts, data access events, and data upload activities were analyzed to identify trends, patterns, and anomalies. Authorized activities were monitored and logged, while unauthorized activities were flagged, investigated, and blocked in real-time to maintain the security and integrity of the cloud environment.

V.CONCLUSION

Data leakage in cloud computing environments poses significant risks to organizations, compromising the confidentiality, integrity, and availability of sensitive data. To mitigate these risks, it is imperative for organizations to implement a multi-faceted approach to data leakage detection and prevention. By leveraging advanced technologies such as DLP solutions, encryption, anomaly detection, and endpoint security, along with robust access control, auditing, and employee training programs, organizations can enhance their data security posture in the cloud. Furthermore, maintaining a proactive stance through regular security assessments, vendor risk management, and incident response planning ensures timely detection and mitigation of data leakage incidents. Adopting these best practices not only safeguards sensitive data but also reinforces trust and confidence in cloud computing services among stakeholders.

REFERENCES

[1] Chen, H., Leung, H. F., Han, B., & Su, J. ‘Automaticprivacy leakage detection for massive android apps via a novel hybrid approach’, IEEE InternationalConference on Communications (ICC)(2017) 1-7

[2] Kul, G., Upadhyaya, S., & Chandola, V, ‘Detecting data leakage from databases on android apps with concept drift’,17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications12th IEEE International Conference OnBig Data Science And Engineering(TrustCom BigDataSE) (2018) 905-913

[3] Casati, L., & Visconti, A. ‘ The dangers of rooting:data leakage detection in Android applications’, Mobile Information Systems, Special Issue(2018)1-9

[4] Shrivastava, G., & Kumar, P.’Android application behavioural analysis for data leakage. Expert Systems, 38(2019), e12468.

[5] Cam, N. T., Pham, V. H., & Nguyen, T. ‘Detecting sensitive data leakage via inter-applications on Android using a hybrid analysis technique’, Cluster Computing 22(2019), 1055-1064.

[6] Dave Klein 2019,Micro-segmentation: securing complex cloud environments, Network Security 2019 (3), 6-10, 2019

[7] Aditya R Achar, Ambesh Mishra, Diksha Makhijani, Dhairya Nagpal, Mrugendra Vasmatkar 2022, Data Leakage in Android and Anomaly Based Intrusion Detection and Prevention System,2022 2nd Asian Conference on Innovation in Technology (ASIANCON), 1-7, 2022.

[8] Seyoon Ko, Hua Zhou, Jin J Zhou, Joong-Ho Won 2022, High-performance statistical computing in the computing environments of the 2020s,Statistical science: a review journal of the Institute of Mathematical Statistics 37 (4), 494, 2022.
[9] Britta Gammelgaard, Katarzyna Nowicka 2023, Next generation supply chain management: the impact of cloud computing,Journal of Enterprise Information Management, 2023.
[10] FNU Jimmy 2024, Cyber security Vulnerabilities and Remediation Through Cloud Security Tools, Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023 3 (1), 196-233, 2024.