iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52273

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Data Leakage Detection using Cloud Computing

Unnati Komal[1], Isha Gujarathi[2], Prachi Chandra[3], Kumar Shivam[4], Prof. Nikhil H. Deshpande[5]
Department of Information Technology, Sinhgad College of Engineering, Pune

Abstract: ***"Data Leakage Detection Using Cloud Computing" is a project that aims to address the growing need to protect sensitive data from unauthorized access and leakage in cloud computing environments. With the increasing use of cloud-based services, there is a growing concern about data privacy and security. The project proposes a solution that leverages cloud computing technology to detect and prevent data leakage in real-time .The proposed system employs data classification, access control, and monitoring mechanisms to prevent unauthorized data access and ensure data privacy. The system is designed to monitor user activities and detect any suspicious behavior that may indicate a potential data leak. The system uses machine learning algorithms to detect anomalies in user behavior and classify data based on its sensitivity level. The system then applies access control policies to restrict user access to sensitive data.***

The project also utilizes cloud-based storage and computing resources to provide scalability and flexibility. The system can handle large volumes of data and adapt to changing user requirements. The project aims to provide a comprehensive solution for data leakage detection in cloud computing environments, ensuring data privacy and security.

Keywords: Cloud Computing, Data Leakage

I. INTRODUCTION

In recent years, cloud computing has become an drastically demanding option for storing and processing large amounts of data provided. While cloud-based systems offer numerous advantages such as flexibility, scalability, and cost-effectiveness, they also pose significant security risks. One of the most pressing concerns in cloud security is the risk of data leakage, which occurs when sensitive information is disclosed to unauthorized parties. Data leakage can have severe consequences, including financial losses, reputational damage, and legal consequences.

To address this issue, various techniques have been developed to detect data leakage in cloud-based systems. However, many of these techniques are either too complex or too expensive to implement. This study proposes a novel approach to data leakage detection that leverages the power of cloud computing and machine learning algorithms. The proposed system can monitor and analyze data traffic in the cloud, identifying patterns and anomalies that may indicate a data leak. By doing so, the system can detect data leakage in real-time and provide timely alerts to system administrators.

The main objective of this study is to evaluate the effectiveness of the proposed system for detecting data leakage in cloud-based systems. To achieve this objective, we will conduct experiments using real-world datasets and evaluate the performance of the system in terms of accuracy, precision, and recall. The results of this study will provide insights into the potential of cloud-based solutions for data leakage detection and contribute to the development of more robust and cost-effective security mechanisms in cloud computing.

However, as the popularity of cloud storage increases, security problems and corresponding threats arise, regardless of how many reliable measures are cloud service providers (CSPs). In addition to threats from the external cloud storage system, it was not possible to rely completely on the cloud service provider to hide data loss incidents to maintain a reputation. Although data outsourcing in the cloud offers a broad perspective for large-scale long-term archiving, it cannot provide any reliable guarantee of data integrity and availability. Without solving this problem, cloud computing technology could not move forward.

Although the cloud storage system has mainly been adopted, it cannot meet some important emerging needs, such as the ability to check the integrity of files in the cloud by clients in the cloud and to detect duplicate files from noisy servers. We will identifiy both problems below. The very first and important problem is Integrity control. This cloud server can mitigate customers of weighty storage management and maintenance load. Biggest difference between cloud storage and ancient internal storage is that it to help in data transfering over the Internet and stored in an unsure domain, which is not relatively under the control of known customers, which unavoidable generates great concerns about the Integrity of your data. These concerns stem from the fact that storage cloud is affected by both internal and external cloud security threats and that uncontrolled cloud servers can passively conceal some customer data loss incidents to keep your reputation. The worst thing is that, to save money and space, servers in the cloud could even actively and deliberately discard limited access data files that belong to a common client.

II. LITERATURE SURVEY

The aim of this literature survey is to provide a comprehensive review of the existing literature on data leakage detection in cloud computing. The survey will focus on different techniques and tools used for detecting data leakage, including data mining, machine learning, and statistical analysis. Moreover, the survey will explore the strengths and weaknesses of each approach and provide insights into future research directions in this area.

Moreover, the survey will explore the strengths and weaknesses of each approach and provide insights into future research directions in this area. The survey will start by providing an overview of cloud computing and data leakage, including the different types of data leakage and their consequences. Then, it will delve into the different approaches used for detecting data leakage, including their theoretical foundations and implementation details. The survey will also review some of the popular tools and frameworks used for data leakage detection in cloud computing.

Overall, this literature survey aims to provide a valuable resource for researchers and practitioners who are interested in data leakage detection in cloud computing. By synthesizing the existing literature, the survey will help to identify gaps in the current research and provide insights into future directions for research in this area.

1) "A Watermarking Technique for Securing Multimedia Data against Digital Data Leakage" by G. Raghavendra Rao, D. D. Doye, and S. Niranjan. This paper proposes a watermarking technique to secure multimedia data against digital data leakage. The proposed technique is based on the discrete wavelet transform and provides robustness against various attacks. The paper also discusses the integration of watermarking with access control and monitoring mechanisms for enhanced security.
2) "A Novel Data Leakage Detection Technique Using Watermarking and Steganography" by S. Sathishkumar and R. Sridhar. This paper proposes a novel technique for detecting data leakage using watermarking and steganography. The proposed technique embeds a watermark in the data and then applies steganography techniques to hide the watermark. The paper also discusses the use of machine learning algorithms for enhanced detection accuracy.
3) "A Hybrid Approach for Detecting Data Leakage in Cloud Environment" by N. R. Jhanjhi, V. Sharma, and M. Singh. This paper proposes a hybrid approach for detecting data leakage in cloud environments. The proposed approach combines watermarking, data classification, and access control mechanisms to detect and prevent data leakage. The paper also discusses the use of a genetic algorithm for optimizing access control policies.
4) "A Survey of Data Leakage Detection Techniques" by M. W. Abu, A. U. M. Emon, and M. Z. H. Bhuiyan. This paper provides a comprehensive survey of data leakage detection techniques, including watermarking techniques. The paper discusses various watermarking techniques and their applicability to different types of data. The paper also highlights the limitations and challenges of watermarking techniques for data leakage detection.
5) "A Comparative Study of Digital Watermarking Techniques for Copyright Protection of Text Documents" by S. M. Bokhari, S. H. Khan, and M. Usman Akram. This paper provides a comparative study of digital watermarking techniques for copyright protection of text documents. The paper discusses various watermarking techniques, including spread spectrum watermarking and discrete cosine transform-based watermarking. The paper also discusses the effectiveness of these techniques in detecting data leakage.

III. PROPOSED SYSTEM

The proposed system for "Data Leakage Detection Using Cloud Computing" consists of several components that work together to provide a comprehensive solution for data leakage detection and prevention in cloud computing environments.

1) *Data Classification:* The first component of the system involves classifying data based on its sensitivity level. The system uses machine learning algorithms to analyze data and classify it into different categories based on its sensitivity level. This component ensures that sensitive data is adequately protected and only accessed by authorized users.
2) *Access Control:* The second component of the system involves access control policies. Access control policies are applied to restrict user access to sensitive data. The system employs role-based access control policies to ensure that only authorized users can access sensitive data.
3) *Monitoring Mechanism:* The third component of the system involves monitoring user activities. The system uses machine learning algorithms to monitor user behavior and detect any suspicious activities that may indicate potential data leakage. The system can detect anomalies in user behavior and take appropriate actions to prevent data leakage.
4) *Cloud-Based Storage and Computing:* The fourth component of the system involves utilizing cloud-based storage and computing resources. The system uses cloud-based storage to store large volumes of data and computing resources to process data in real-time. This component provides scalability and flexibility to the system, ensuring that it can handle large volumes of data and adapt to changing user requirements.

5) *Reporting and Alerting:* The final component of the system involves reporting and alerting. The system generates reports on user activities and data access patterns, enabling administrators to monitor the system's performance. The system also sends alerts to administrators when it detects potential data leakage, enabling them to take immediate action.

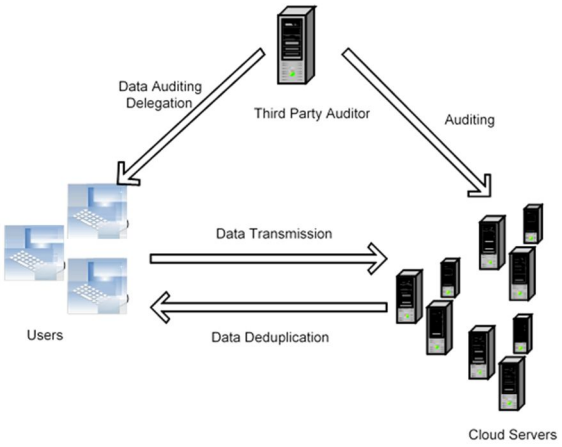


Fig 1. System Architecture

System Architecture is shown in figure 1.

The entire system works in stages.

The main goal is to achieve data integrity and deduplication in the cloud. Integrity audit. The first design goal of this work is to provide the possibility to verify the accuracy of data stored remotely. The integrity check also requires two features:

a) Public verification, which allows any person, not only customers who originally filed the file, to perform the verification;

b) Stateless verification, which can eliminate the need to maintain status information on the verifier side between audit actions and data storage.

Data auditing delegation plays a crucial role in the topic of data leakage detection using cloud computing. Auditing is the process of verifying and validating data to ensure its accuracy, completeness, and security. In cloud-based systems, auditing can be challenging due to the distributed nature of the data and the lack of direct control over the infrastructure.

Data auditing delegation allows organizations to delegate the auditing process to trusted third parties, such as auditors or regulators. These third parties can verify that the cloud service provider is complying with the organization's security policies and regulations. They can also perform independent audits of the data stored in the cloud to detect any unauthorized access or data leakage.

Third-party auditors play a crucial role in the topic of data leakage detection using cloud computing. Cloud service providers may offer different levels of security controls and measures to ensure data confidentiality and integrity, but it's difficult for cloud users to independently verify these claims. Third-party auditors provide an objective assessment of the cloud service provider's security posture and help cloud users to evaluate the security of their data in the cloud.

- *Independent Verification:* Third-party auditors independently verify the security controls and measures implemented by the cloud service provider to ensure data confidentiality and integrity. This involves assessing the cloud provider's security policies, procedures, and infrastructure to ensure that they are adequate to protect the data stored in the cloud.
- *Risk Assessment:* Third-party auditors assess the risks associated with storing data in the cloud and provide guidance to cloud users on how to mitigate those risks. This includes identifying potential vulnerabilities and threats that could lead to data leakage and providing recommendations for mitigating those risks.
- *Compliance Verification:* Third-party auditors verify that the cloud service provider is compliant with relevant security and privacy regulations, such as HIPAA, GDPR, and PCI DSS. This provides assurance to cloud users that their data is being handled in accordance with regulatory requirements.
- *Data Auditing:* Third-party auditors perform independent audits of the data stored in the cloud to detect any unauthorized access or data leakage. This helps cloud users to detect and respond to data leaks quickly and prevent further damage.

Cloud servers play a crucial role in the topic of data leakage detection using cloud computing. These servers provide the necessary infrastructure and resources for data processing and analysis, allowing organizations to detect data leaks and prevent further damage.

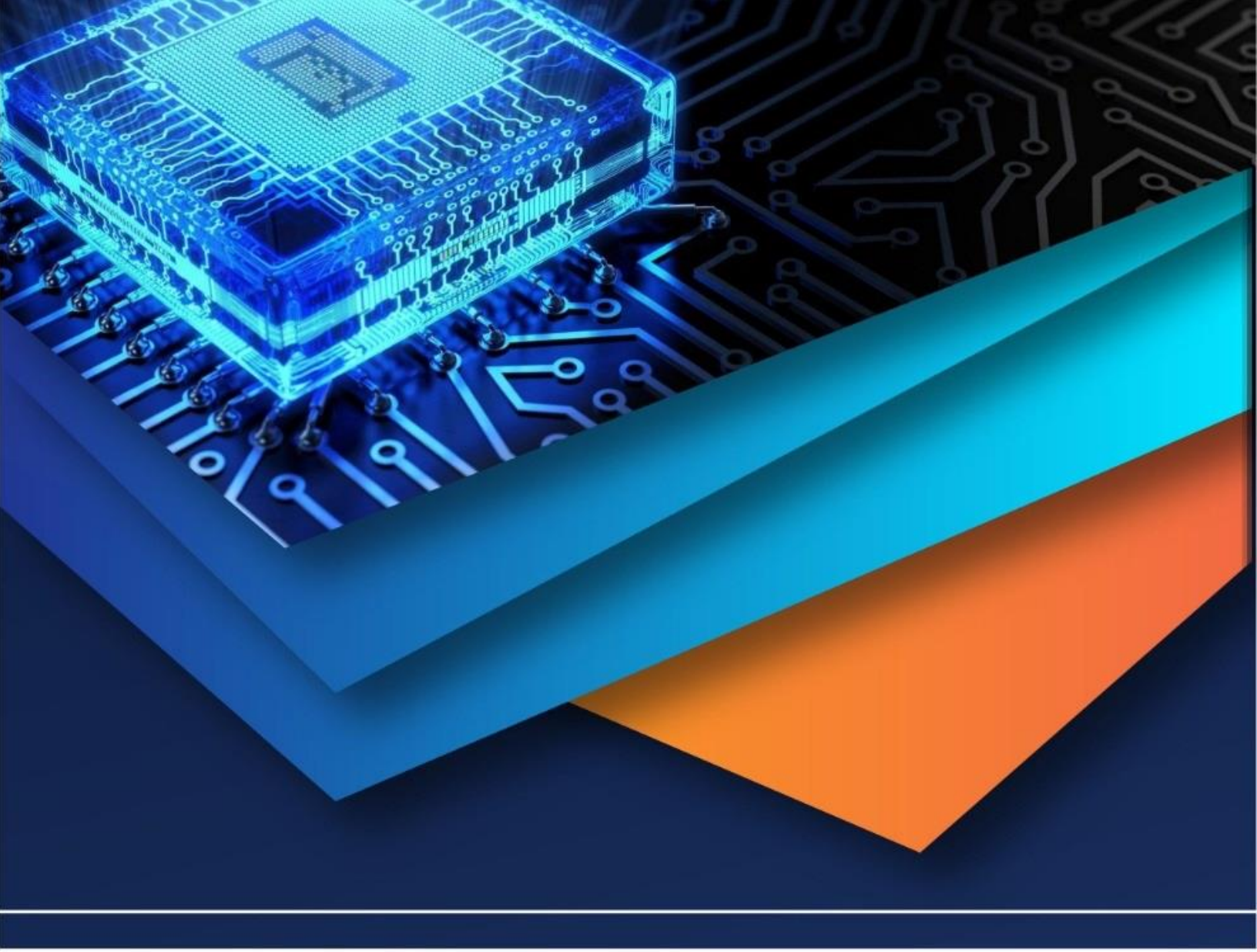
IV. CONCLUSION

The proposed project topic, "Data Leakage Detection using Cloud Computing", aims to develop a system that can detect data leakage in cloud environments. The project involves designing and implementing a data leakage detection system using cloud computing technologies.

The project involves several phases, including problem identification, requirements gathering, design and architecture, implementation, testing, deployment, and maintenance and support. The successful completion of these phases will result in the development of a robust and effective data leakage detection system that can help organizations protect their sensitive data in cloud environments. Overall, the proposed project is a significant contribution to the field of cloud computing security and can help organizations address the growing concerns of data leakage in cloud environments. The project's outcome can have a significant impact on the security of sensitive data and help organizations maintain the confidentiality, integrity, and availability of their critical data assets.

REFERENCES

[1] Q in long Huang, Member, IEEE, Yixian Yang, Wei Yue and Yue He" Secure Data Group Sharing and Conditional Dissemination with Multi-Owner in Cloud Computing", IEEE TRANSACTIONS ON CLOUD COMPUTING , APRIL 2019
[2] Z. Yan, X. Li, M. Wang, and A. V. Vasilakos, "Flexible data access control based on trust and reputation in cloud computing," IEEE Transactions on Cloud Computing, vol. 5, no. 3, pp. 485-498, 2017.
[3] H. Cui, X. Yi, and S. Nepal, "Achieving scalable access control over encrypted data for edge computing networks," IEEE Access, vol. 6, pp. 30049–30059, 2018.
[4] K. Xue, W. Chen, W. Li, J. Hong, and P. Hong, "Combining data owner-side and cloud-side access control for encrypted cloud storage," IEEE Transactions on Information Forensics and Security, vol. 13, no. 8, pp. 2062–2074, 2018.
[5] N. Paladi, C. Gehrmann, and A. Michalas, "Providing user security guarantees in public infrastructure clouds," IEEE Transactions on Cloud Computing, vol. 5, no. 3, pp. 405-419, 2017.
[6] Q. Huang, Y. Yang, and J. Fu, "Secure data group sharing and dissemination with attribute and time conditions in Public Clouds," IEEE Transactions on Services Computing, 2018.
[7] L. Jiang, and D. Guo "Dynamic encrypted data sharing scheme based on conditional proxy broadcast re-encryption for cloud storage," IEEE Access, vol. 5, pp. 13336 – 13345, 2017.
[8] K. Xu, Y. Guo, L. Guo, Y. Fang, and X. Li, "My privacy my decision: control of photo sharing on online social networks," IEEE Trans. On Dependable and Secure Computing, vol. 14, no. 2, pp. 199-210, 2017.
[9] L. Fang, L. Yin, Y. Guo, Z. Wang, and Fenzhua Li, "Resolving access conflicts: an auction-based incentive approach," Proc. IEEE Military Communications Conference (MILCOM), pp. 1-6, 2018.
[10] Q. Huang, W. Yue, Y. He, and Y. Yang, "Secure identity-based data sharing and profile matching for mobile healthcare social networks in cloud computing," IEEE Access, vol. 6, pp. 36584–36594, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY