Data Leakage Detection in Cloud Computing Environment

[1]Mr. V. UDHAYAKUMAR,[2]VIGNESH K

[1]Associate Professor, Department of MCA, Sri Manakula Vinayagar Engineering College, Puducherry-605107, India.

[2]PG Student, Department of MCA, Sri Manakula Vinayagar Engineering College, Puducherry-605107 India.

vigneshvignesh9185@gmail.com[2]

ABSTRACT

The detection of data leaks in cloud computing is achieved by integrating Java, HTML, CSS, and JavaScript. As a result, you get both strong backend support and a handy, user-friendly frontend. Java is the main programming language used to manage server-side responsibilities including logging in, secure files, data access, and finding leakage. Java makes it possible to avoid compatibility problems, grows as needed, and enhances security, which is why it's best for the project's cloud-side components. Thanks to HTML and CSS in the frontend, a simple and user-friendly web interface is built for administrators and users to use. Interactions, data checks, and dynamic changes are made possible using JavaScript, so that the entire experience is continuous. In the project, files are transferred to and from a cloud storage area where everything is being watched for unusual behavior that could result in sensitive data being disclosed. When unusual things like off-limit file access are noticed, the system alerts the administrator and stops anyone else from accessing the file. Because of these technologies, the project is able to show a working method for addressing and solving threats of data leakage in cloud-based services. With this implementation, people can learn about cloud security challenges and the need for preventive data protection measures. The project keeps track of what is happening with sensitive data in a virtual cloud environment and prevents any unauthorized use of this data.

KEYWORDS

Cloud Computing, Data Leakage, Java, Web Security, Anomaly Detection, Cloud Storage, User Authentication, Access Monitoring

1.INTRODUCTION

Cloud computing has caused a major shift in how data is handled by using technology that can expand as needed at a reasonable cost. This helps both companies and individuals keep a lot of data outside their offices, making it easy for everyone to share and access the data worldwide. On the other hand, using cloud computing means encountering newly arising problems, mainly in data security and privacy. For this reason, this project will design and apply a data leakage detection system in a simulated cloud environment. It is essential that the pathfinder allows for secure file storage and access and also to watch for any possible harmful use by a user in real time. When the backend is secure and the frontend is interactive, the system ensures that every access to files is documented, looked at, and tested for unusual patterns.

2.PROBLEM STATEMENT

Since more organizations are using cloud computing, they now count on cloud storage to safely manage and access their sensitive data. Although cloud services are very scalable, convenient, and economical, they still have serious security problems. Unauthorized use, distribution, or exposure of confidential information through data leakage is a major and tough concern in cybersecurity. As more data is saved remotely and accessed with the help of the internet, people are becoming concerned about its protection from attackers, mistakes, and those inside the organization. Simple security measures like firewalls and encryption work well, yet they are not enough to discover all types of hidden and delicate data leaks. Most of the time, data breaches result from unusual user activity or lack of proper control over data access, and this is not detected on time by traditional systems. Also, since cloud computing is not as transparent, it becomes complicated for users and administrators to observe the use of data, identify any unusual actions, or address immediate dangers. With this in mind, there is an immediate need for artificial intelligence that watches over, finds, and stops unauthorized sharing of data on the cloud.

3. LITERATURE SURVEY

Using the internet, people can use cloud computing to store and get to their data even when they are not connected to the same computer. Even so, people have still been worried about the safety of sensitive information kept in the cloud, as more cases of data breach and unauthorized access keep surfacing. Numerous strategies and models aimed at dealing with data leakage have been put forward by experts as it is still one of the biggest dangers in the cloud. The approach most people reference is tracking data provenance, which involves keeping a careful record of how data is handled in the system. An assessment of the history of how data is used allows for detecting suspicious access to a system. The authors Wang et al. (2010) suggested an effective and secure system for finding out the source of data leakage by embedding identifiers in shared documents. Nevertheless, organizing such technologies at a large scale can be difficult and costly because they involve a lot of management. Another major point is the value of detecting abnormal activities and analyzing behaviors. They use either machine learning or rules to keep an eye on user behaviors and point out anything that's unusual. For example, if a user is active with their files outside the usual work time or downloads a lot of information, the system will notify administrators.

Using the internet, people can use cloud computing to store and get to their data even when they are not connected to the same computer. Even so, people have still been worried about the safety of sensitive information kept in the cloud, as more cases of data breach and unauthorized access keep surfacing. Numerous strategies and models aimed at dealing with data leakage have been put forward by experts as it is still one of the biggest dangers in the cloud. The approach most people reference is tracking data provenance, which involves keeping a careful record of how data is handled in the system. An assessment of the history of how data is used allows for detecting suspicious access to a system. The authors Wang et al. (2010) suggested an effective and secure system for finding out the source of data leakage by embedding identifiers in shared documents. Nevertheless, organizing such technologies at a large scale can be difficult and costly because they involve a lot of management. Another major point is the value of detecting abnormal activities and analyzing behaviors. They use either machine learning or rules to keep an eye on user behaviors and point out anything that's unusual. For example, if a user is active with their files outside the usual work time or downloads a lot of information, the system will notify administrators.

4.PROPOSED TECHNIQUES

Step 1: User Authentication and Role Assignment

Everyone who wants to use the website must register and check in with secure information. After logging in, people are assigned the role of Administrator or Regular User. People are permitted certain actions on the system according to the roles assigned to them.

Example:

Admins manage all aspect of files, but regular users can only work with certain types of files.

Step 2: Access Control and File Operations

People are able to upload, download, or view files if the action is permitted. Validations for actions are made according to a person's role and permissions. Operations that should not be done are blocked straight away.

Example:

Trying to access a restricted file will lead to access denial and the attempt will be noted in the log.

Step 3: Activity Logging

All actions by users such as logging in, moving files, and logging out are tracked and recorded in the system.

- Username
- Timestamp

This kind of name belongs to an IP address.

- Things happen

Example:

If a user makes many quick downloads, the system notes all the actions for examination.

Step 4: Behavior Monitoring and Anomaly Detection

The system constantly tracks how people access different websites. Limiters are put in place to decide what is considered normal behavior. Anyone who behaves in a way that is not usual is marked as suspicious.

Example:

When a user downloads 10 files within 2 minutes, it surpasses the allowed level and messages are sent out.

Step 5: Real-Time Alerting and Action

Alert and take action whenever needed. Once something unusual is noticed, the system does the following: The administrator is notified about the incident as soon as it happens. The user's access is stopped while the account is being looked into further.

Example:

A warning is displayed on the admin screen and the account is locked to be checked for wrongdoing.

Step 6: Integrity Verification (Optional Feature)

To ensure security, the system can use a hashing function called SHA-256 on every file it gets. After finishing, a check is made on the hash to ensure no changes happened to the data.

Example:

A file cannot be downloaded if its hash does not match what was recorded in blockchain.

5. SYSTEM ARCHITECTURE

Fig 1 Architecture Diagram

Massive amounts of data are kept by the cloud server on behalf of its users. However, a 37 a malicious cloud server may choose to delete some of the client's rarely accessed data files not everything needs to be brought with you. Two-party confidential procedures are used to make sure the client's data is protected. the server ensures that the client's data does not get changed when it is being transmitted. The definition of the protocols depends on the kind of outsourced data, and these categories are known as secure cloud. storage protocols deal with static data (SSCS) and dynamic data (DSCS). For the cases where data is static, the client . After the data has been handed over to the service provider for outsourcing, it is not allowed to edit. Dynamic data allow more freedom since the client can change her information as frequently as required. Things are done securely on the cloud. because of storage protocols, the client is able to review the data without seeing everything in detail . Despite using file, you would still find out if a malicious server has messed with the data in any way. During an With audit, the client asks for a proof that certain data is leveraged from the server.

Based on the data that has been saved, the system computes an answer to that challenge. Safe methods for using the cloud can be checked by anyone as Auditors (TPAs) can perform an audit on them . If only the public can verify the data, it is said to be public parameters, otherwise, it needs some private information for an auditor. client. What parties are tied to a secure cloud storage protocol and their ways of interacting.

6.CONCLUSION AND FUTURE WORKS

The project clearly shows that risks to data in the cloud can be identified and managed in an effective way. When Java is used for the backend and HTML, CSS, and JavaScript for the frontend, the system offers a full experience just like real cloud storage. Having user authentication, role-based access control, activity logging, anomaly detection, and quick alerting feature together, the system becomes better at finding suspicious access and suspect behaviors. Its ability to watch and study human activity means that data is safeguarded even if there are possible attacks from employees or outside parties. Besides strong security, it shows how to use proper strategies to store and protect data in the cloud.

Fig 2. Data Leakage Detection

Eventually, the current data leakage detection system might become much better if advanced features for increased security, scalability, and usability are introduced. It would be helpful to use machine learning to examine login patterns and catch anything suspicious more efficiently, which would limit the number of false alarms and protect against fresh

types of security threats. Also, deploying modern encryption features makes sure that even if secret information is accessed without permission, it will be secure. Adding MFA further improves the login process since attackers will find it much more difficult to use stolen usernames and passwords to gain access. Another step would be to make the project compatible with leading cloud services such as AWS and Microsoft Azure.

REFERENCES

1. S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," *IEEE INFOCOM 2010*, pp. 534–542,Mar.2010.
doi: 10.1109/INFCOM.2010.5462174

2. Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 5, pp. 847–859, May 2011.
doi: 10.1109/TPDS.2010.183

3. M. Armbrust et al., "A view of cloud computing," *Communications of the ACM*, vol.53,no.4,pp.50–58,Apr.2010.
doi: 10.1145/1721654.1721672

4. A. C. Squicciarini, S. Sundareswaran, and D. Lin, "Preventing information leakage from indexing in the cloud," *IEEE Transactions on Services Computing*,vol.5,no.3,pp.302–319,Jul.-Sep.2012.
doi: 10.1109/TSC.2011.14

5. R. H. Zhan and J. J. Liu, "Anomaly detection for cloud computing using rule-based and statistical approaches," *Proceedings of the 2014 IEEE International ConferenceonCloudComputing*,pp.113–120,Jun.2014.
doi: 10.1109/CLOUD.2014.24

6. L. Wang, G. Laszewski, M. Kunze, J. Tao, and A. Castellanos, "Cloud computing: A perspective study," *New Generation Computing*, vol. 28, no. 2, pp.137–146,Apr.2010.
doi: 10.1007/s00354-008-0081-5