IJARASEM

ISSN: 2395-7852

International Journal of Advanced Research in Arts, Science, Engineering & Management

Volume 10, Issue 3, May 2023

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 6.551

International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 6.551 |Bimonthly, Peer Reviewed & Referred Journal|

| Volume 10, Issue 3, May 2023 |

Protecting Sensitive Data and Detecting Suspicious File Migration in Cloud Based Private Network

R. Prabhu Arokiyaraj [1], P. Praneshver [2], Harieesh Rakhavandaran.N.A [3], M. Vengatesh Kumar [4]

[1] Assistant Professor, Department of Computer Science and Engineering, Velammal Institute of Technology, Panchetti, Chennai, India

[2] UG Scholar, Department of Computer Science and Engineering, Velammal Institute of Technology, Panchetti, Chennai, India

[3] UG Scholar, Department of Computer Science and Engineering, Velammal Institute of Technology, Panchetti, Chennai, India

[4] UG Scholar, Department of Computer Science and Engineering, Velammal Institute of Technology, Panchetti, Chennai, India

ABSTRACT: Identifying weaknesses is important to refine an organization's security strategy and ensure the desired level of protection. Vulnerability assessments by top executives can help uncover security flaws that could lead to device or data compromise. User management can also be a source of vulnerabilities, as regular users may have access to sensitive information or systems. Ordinary client organization might give the way to unapproved individuals to access as an approved specialists. Whenever, clients step into online organizations, without realizing them outsider or some other unsafe individual checking their way of behaving. Give the security from malevolent movement, administrator or approved individual likewise check the client organizations, for example, IP address and email. Administrators or authorized personnel should verify user credentials to prevent unauthorized access and ensure security from malicious activity.

I. INTRODUCTION

Cloud computing is a cost-effective way of storing data on third-party servers, allowing for scalability and reliability. However, the default storage of data in the clear presents a risk of unauthorized access or manipulation, making it important to apply software techniques such as encryption keys to ensure data confidentiality. This is especially crucial for sensitive user data, such as personal customer information or medical records. Cloud computing provides a scalable and cost-effective solution for data storage, but data integrity is a major concern due to potential hardware failures, software errors, and human mistakes.[1]

In order have a safe transmission of data over cloud we have a lot security system, one of the previous existing system is built using LAST-HDFS (Location aware storage technique) technique. With the ever-increasing popularity of cloud computing, the demand for cloud storage has also increased exponentially. Computing firms are no longer the only consumers of cloud storage and cloud computing, but rather average businesses, and even end-users, are taking advantage of the immense capabilities that cloud services can provide.

While enjoying the flexibility and convenience brought by cloud storage, cloud users release control over their data, and particularly are often unable to locate the actual their data; this could be in-state, in-country, or even out-of-country.

Lack of location control may cause privacy breaches for cloud users (e.g., hospitals) who store sensitive data (e.g., medical records) that are governed by laws to remain within certain geographic boundaries and borders. Another situation were, this problem arises is with governmental entities that require all data to be stored in the same country that the government operates in; this challenge has seen difficulties with cloud service providers (CSPs) quietly moving data out-of-country or being bought out by foreign companies.

For example, Canadian laws demand that personal identifiable data must be stored in Canada. However, large cloud infrastructure like the Amazon Cloud has more than 40 zones distributed all over the world [1], which makes it very challenging to provide guaranteed adherence to regulatory compliance. Even Hadoop, which historically has been managed as a geographically confined distributed file system, is now deployed in large scale across different regions (see Facebook Prism [2] or recent patent [3]).

Major disadvantages of using LAST-HDFS technique, the implementation of LAST-HDFS may introduce additional computational overhead, potentially slowing down data transfer and access. The LAST-HDFS system may require additional resources to maintain and monitor, which could result in increased costs for the storage provider and, in turn, potentially higher fees for users. The implementation of LAST-HDFS may also require additional expertise and training for the storage provider's staff, which could result in additional costs and potential delays in implementation.

So in order to overcome these disadvantages of using LAST-HDFS we have implemented our system using AES algorithm. The major advantage of using our system over LAST-HDFS lies in fact that it ensures the data confidentiality and data integrity. Customary client association could give the best approach to unapproved people to access as an endorsed trained professionals.

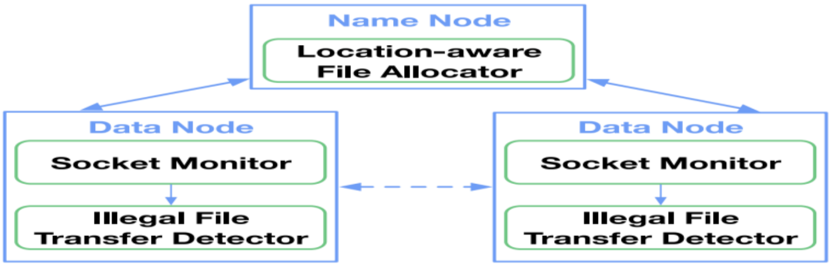
II. RELATED WORKS

Hadoop Distributed File System (HDFS), and design an enhanced HDFS system, called LAST-HDFS. The LAST-HDFS extends HDFS' capabilities to achieve location-aware file allocations and file transfer monitoring.

Specifically, LAST-HDFS provides the following new functions: (i) consistently enforces a location aware data loading and storage by assigning data nodes according to user specified privacy policies; (ii) actively tracks and dynamically corrects possible data migration (due to balancing or data replication needs) within the cluster that might violate data placement policies; (iii) detects potentially illegal data migration, by monitoring socket communication between individual data nodes and correlating it with the constraints imposed by the policy

The idea of our approach is that, once data is allocated per users' location preferences, our framework monitors real time file transfers in the cloud and is capable of detecting potential illegal transfers.

An illegal transfer in our context denotes moving sensitive data outside the legal boundaries specified by the file owner (e.g., storing a file in a physical location other than what the file owner desires).


Name Node
Location-aware File Allocator
Data Node
Socket Monitor
Illegal File Transfer Detector
Data Node
Socket Monitor
Illegal File Transfer Detector

Our approach builds on the observation that users' location preferences are often consistent with privacy laws and regulations. As a result, files can be gathered into groups in which multiple users share the similar, if not the same, location preferences. Accordingly, our system allocates cloud nodes based on the similarity of users' location preferences.

As a result, files can be gathered into groups in which multiple users share the similar, if not the same, location preferences. Accordingly, our system allocates cloud nodes based on the similarity of users' location preferences. More specifically, we model the file transfers among nodes as a weighted graph and then maximize the probability that files with similar privacy preferences will be stored in the same region.

III. PROPOSED SYSTEM

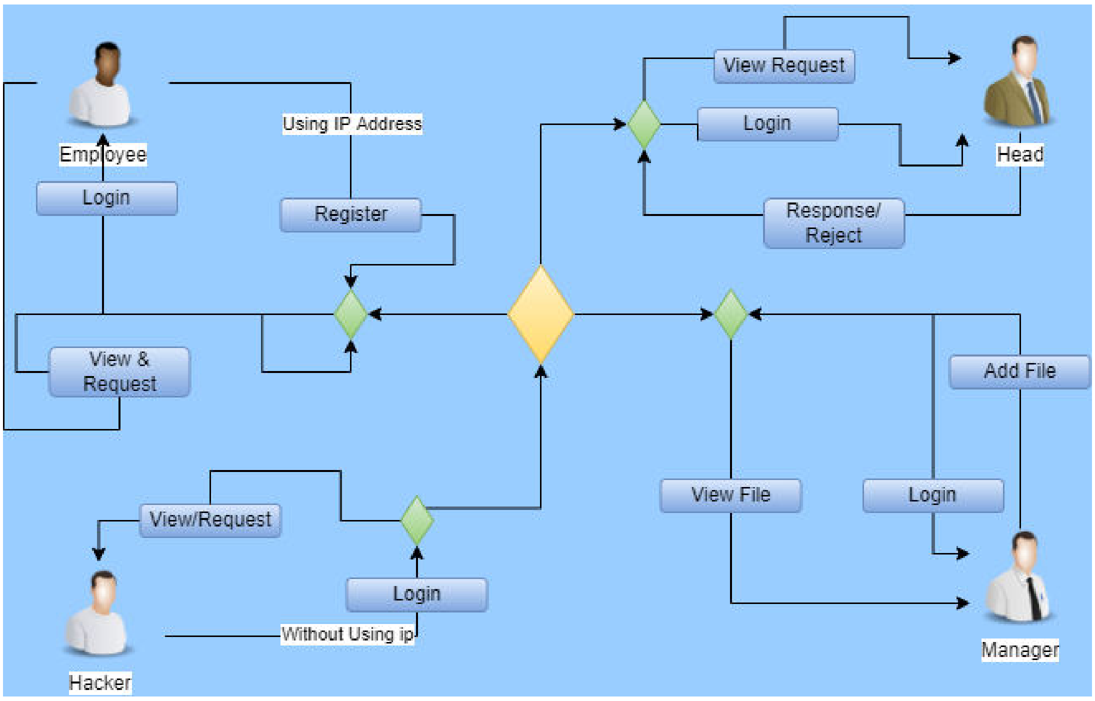
Our proposed system majors focus on the confidentiality and integrity of data. It ensures both extreme confidentiality of data which is being transmitted between parties. It also ensures safe transmission of data within the connected parties inside the organization. It also acts as a preventive measure that it does not allow unauthorized users to access the information without an administrator permission.

Our proposed system also ensures safe delivery of data to the requested party without a data being stolen or hacked.Give the security from malicious development, manager or supported individual moreover really look at the client associations, for instance, IP address and email.

We have designed in such a way that only authorized members whose IP address is connected with our system can only make a transmission. This is done by Request-Response method where request is made at one end and response to the request is made at the other end.At each end that is both request end and response end key has been generated.

To handle in key we have used AES algorithm and in order to ensure the overall security we have used SHA algorithm.

3.1 Architecture diagram



IV. DATA DESCRIPTION

For this application there are in total six data modules were implemented each with their own priority and their functionality respectively. Those data modules were discussed below

4.1.LOGIN :-

In everyday PC use, logon is the methodology used to gain admittance to a working framework or application, as a rule in a far off PC. Quite often a logon expects that the client have (1) a client ID and (2) a secret key. Frequently, the client ID should adjust to a restricted length, for example, eight characters and the secret key should contain something like one digit and not match a characteristic language word.

4.2.ADD FILE

The Java record transfer Servlet will contain a do Post technique to deal with the structure accommodation. In this do Post strategy, the transferred document will handle in parts. After the record transfers, the Java Servlet saves each part to a like-named document in the server's record system. This explanation likewise permits the designer to set different record transfer properties, like the max File Size, max Request Size and the file Size Threshold.

4.3.VIEW

This module to help us the staff view the file. But the staff cannot be open the file until the Head office accept the file. After response the head office the file can view by the staff with the help of secret key.

4.4.REQUEST

In this module, during the transmission the user who wants a data can make a request to the admin. After the request the user should wait for the response.

4.5.Response

In this module, response will be sent from the admin side to user who made a request. Along with response key is also sent to the user who requested the data.

4.6.Download

This example just downloads the record, be that as it may, you can stretch out the usefulness to introduce or execute documents utilizing the execute shell order in specialist systems. This module is help to staff download the record after head office support the solicitation. The information recover in the data set, they view at the download page. The assignment finished once endorse.

5. Strategic approach

The software engineering process is often depicted as a spiral, with each turn representing a phase in the software development lifecycle. The spiral begins with system engineering, where the role of software is defined, and leads to software requirement analysis, where the functions, behavior, performance, constraints, and validation criteria for software are established. The spiral then moves inward, with design and coding phases, each turn decreasing the level of abstraction.

Testing is an integral part of the software development process and can also be viewed in the context of the spiral. The testing strategy begins at the vertex of the spiral with unit testing, which focuses on each unit of software implemented in source code. As testing progresses, the focus moves outward along the spiral to integration testing, where the software architecture is tested for design and construction. Another turn on the spiral leads to validation testing, where software requirements are validated against the constructed software. Finally, the spiral reaches system testing, where the software and other system elements are tested as a whole. A strategic approach to software testing involves ensuring that each phase of the spiral is thoroughly tested before moving on to the next phase. This approach ensures that issues are identified and addressed early in the development process, reducing the risk of costly errors and delays further down the line. It is also important to consider the specific testing needs of each phase, using appropriate testing techniques and tools to ensure that software meets the necessary quality standards. By following a strategic approach to software testing, developerscan ensure that their software is reliable, robust, and meets the needs of their end-users which is demonstrated in below figure 2.

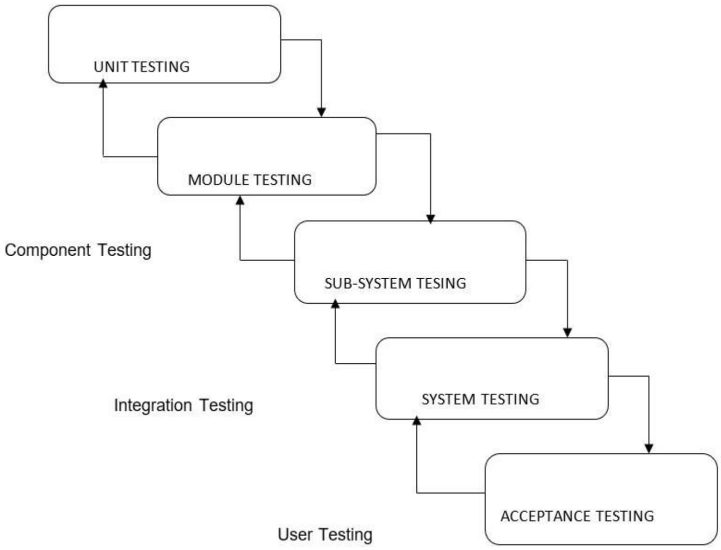


Figure – 2 Testing

V. CONCLUSION

Our system concentrates on safe transmission of data among users who are connected together. And it also focuses on two main factors which is dataconfidentiality and integrity. So we can say that by using our system the users can transfer sensitive data among themselves. It especially designed for the organization who are working on a project.

To effectively detect suspicious file migration or replication, it is important to have a comprehensive security strategy in place that includes regular security audits, user education, and the implementation of security best practices. Organizations should also have a clear incident response plan that outlines the steps to be taken in the event of a security breach.

By implementing these measures, organizations can reduce the risk of unauthorized access to their sensitive data and protect their reputation and financial well-being.

And we also conclude that by combination of AES and SHA algorithm in our system enhances the overall system security. For future works, or in future enhancements we can able to replace these two algorithms with the latest algorithm which will much more enhanced security.

REFERENCES

1] Amazon, "Aws global infrastructure," 2017. [Online]. Available: https://aws.amazon.com/about-aws/global-infrastructure/

[2] C. Metz, "Facebook tackles (really) big data with project prism," 2012. [Online]. Available: https://www.wired.com/2012/08/ facebook-prism/

[3] K. V. Shvachko, Y. Aahlad, J. Sundar, and P. Jeliazkov, "Geographically-distributed file system using coordinated namespace replication," 2014. [Online]. Available: https://www.
google.com/patents/WO2015153045A1?cl=zh

[4] C. Liao, A. Squicciarini, and L. Dan, "Last-hdfs: Location-aware storage technique for hadoop distributed file system," in Proc. IEEE Int. Conf. Cloud Comput., 2016, pp. 662–669.

[5] N. Paladi and A. Michalas, ""One of our hosts in another country": Challenges of data geolocation in cloud storage," in Proc. Int. Conf. Wireless Commun. Veh. Technol. Inf. Theory Aerosp. Electron. Syst., 2014, pp. 1–6.

[6] Z. N. Peterson, M. Gondree, and R. Beverly, "A position paper on data sovereignty: The importance of geolocating data in the cloud," in Proc. 3rd USENIX Conf. Hot Topics Cloud Comput., 2011, pp. 9–9. [7] A. Squicciarini, D. Lin, S. Sundareswaran, and J. Li, "Policy driven node selection in MapReduce," in Proc. 10th Int. Conf. Security Privacy Commun. Netw., 2015, pp. 55–72.

[8] J. Li, A. Squicciarini, D. Lin, S. Liang, and C. Jia, "Secloc: Securing location-sensitive storage in the cloud," in Proc. ACM Symp. Access Control Models Technol., 2015, pp. 51–61.

[9] E. Order, "Presidential executive order on strengthening the cybersecurity of federal networks and critical infrastructure," 2017. [Online]. Available: https://www.whitehouse.gov/the-pressoffice/2017/05/11/presidential-executive-order-strengtheningcybersecurity-federal

[10] "Hdfs architecture," (2018). [Online]. Available: http://hadoop. apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/ HdfsDesign.html

[11] R. Miller, "Inside amazon cloud computing infrastructure," 2015. [Online]. Available: http://datacenterfrontier.com/inside-amazoncloud-computing-infrastructure/

[12] T. Bujlow, K. Balachandran, S. L. Hald, M. T. Riaz, and J. M. Pedersen, "Volunteer-based system for research on the internet traffic," Telfor J., vol. 4, no. 1, pp. 2–7, 2012.

[13] M. Geist, "Location matters up in the cloud," (2010). [Online]. Available: http://www.thestar.com/business/2010/12/04/ geist_location_matters_up_in_the_cloud.html

[14] Z. N. Peterson, M. Gondree, and R. Beverly, "A position paper on data sovereignty: The importance of geolocating data in the cloud," in Proc. 8th USENIX Conf. Netw. Syst. Design Implementation, 2011, pp. 9–9.

[15] K. Benson, R. Dowsley, and H. Shacham, "Do you know where your cloud files are?" in Proc. 3rd ACM Workshop Cloud Comput. Security Workshop, 2011, pp. 73–82.

[16] M. Gondree and Z. N. Peterson, "Geolocation of data in the cloud," in Proc. 3rd ACM Conf. Data Appl. Security Privacy, 2013, pp. 25–36.

[17] G. J. Watson, R. Safavi-Naini, M. Alimomeni, M. E. Locasto, and S. Narayan, "Lost: Location based storage," in Proc. ACM Workshop Cloud Comput. Security Workshop, 2012, pp. 59–70.

[18] A. Albeshri, C. Boyd, and J. G. Nieto, "Geoproof: Proofs of geographic location for cloud computing environment," in Proc. 32nd Int. Conf. Distrib. Comput. Syst. Workshops, 2012, pp. 506–514.

[19] A. Albeshri, C. Boyd, and J. G. Nieto, "Enhanced geoproof: Improved geographic assurance for data in the cloud," Int. J. Inf. Security, vol. 13, no. 2, pp. 191–198, 2014.

[20] A. Michalas and K. Y. Yigzaw, "Locless: Do you really care where your cloud files are?" in Proc. IEEE Int. Conf. Cloud Comput. Technol. Sci., 2016, pp. 515–520.

[21] D. Lin, P. Rao, R. Ferrini, E. Bertino, and J. Lobo, "A similarity measure for comparing XACML policies," IEEE Trans. Knowl. Data Eng., vol. 25, no. 9, pp. 1946–1959, Sep. 2013.

[22] P. Rao, D. Lin, E. Bertino, N. Li, and J. Lobo, "Fine-grained integration of access control policies," Computer. Security, vol. 30, no. 2–3, pp. 91–107, 201