International Journal of Engineering Research and Science & Technology

MALICIOUS USER PREDICTION IN MULTI-TENANT CLOUDS USING FEDERATED LEARNING

JIYA BATRA 22UP1A0575

EMAIL: jiya17batra@gmail.com

M. SURYA GEETHIKA 22UP1A05A6

EMAIL: geethikamedarametla@gmail.com

K. BHARGAVI 22UP1A0590

EMAIL: bhargaviireddy14@gmail.com

M. RISHITHA 22UP1A0599

EMAIL : rishithaminnu2005@gmail.com

Dr. Rajendra Prasad

Associate Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VIGNAN'S INSTITUTE OF MANAGEMENT AND TECHNOLOGY FOR WOMEN

(An Autonomous Institution)

(Affiliated to Jawaharlal Nehru Technological University Hyderabad, Accredited by NBA, NAAC with A+) Kondapur (Village), Ghatkesar (Mandal), Medchal (Dist.) Telangana-501301 (2022-2026)

ABSTRACT

Malicious user activities in multi-tenant cloud environments pose significant security challenges due to the shared infrastructure and dynamic nature of cloud resources. Traditional centralized security mechanisms often fail to provide effective detection while preserving user privacy. This project proposes a Malicious User Prediction System in Multi-Tenant Clouds using Federated Learning (FL), which enables collaborative model training without sharing raw data among tenants.

The proposed system leverages federated learning to train machine learning models across multiple cloud clients (tenants) while ensuring data privacy and compliance. Each tenant locally trains a model on its own user activity data, such as login patterns, resource usage, network behavior, and access logs. The locally trained models are then aggregated at a central server to build a global prediction model capable of identifying malicious users with high accuracy.

Advanced classification algorithms such as Random Forest, Support Vector Machines, or Deep Neural Networks can be integrated within the federated framework to improve detection performance. The system also incorporates anomaly detection techniques to identify unusual behavior patterns in real time. By using federated learning, the solution reduces data leakage risks and enhances trust among tenants, making it suitable for sensitive cloud environments.

1. INTRODUCTION

Cloud computing has revolutionized the way organizations store, manage, and process data by providing scalable, on-demand, and cost-effective computing resources. In particular, **multi-tenant cloud environments** allow multiple users or organizations (tenants) to share the same infrastructure while maintaining logical isolation. Although this architecture improves resource utilization and reduces operational costs, it also introduces significant **security challenges**, especially in detecting and preventing malicious user activities.

Malicious users in cloud systems can exploit vulnerabilities to perform unauthorized access, data theft, service disruption, or insider attacks. Traditional security mechanisms such as firewalls, intrusion detection systems, and centralized monitoring tools often struggle to effectively identify such threats due to the distributed and dynamic nature of cloud environments. Moreover, these centralized approaches require collecting large volumes of user data at a single location, which raises serious **privacy concerns** and increases the risk of data leakage.

To address these challenges, **machine learning (ML)** techniques have been widely adopted for detecting abnormal behavior and predicting malicious users. These methods analyze patterns in user activities, such as login frequency, resource consumption, and network traffic, to identify suspicious behavior. However, conventional ML approaches rely on centralized data collection, which is not

always feasible in multi-tenant environments where data privacy and regulatory compliance are critical.

In this context, **Federated Learning (FL)** emerges as a promising solution for privacy-preserving collaborative learning. Federated learning enables multiple tenants to train a shared global model without exchanging their raw data. Instead, each tenant trains a local model on its own data and shares only model updates with a central server. This approach ensures that sensitive information remains within the tenant's environment while still benefiting from collective intelligence across the cloud.

The proposed system, *Malicious User Prediction in Multi-Tenant Clouds using Federated Learning*, integrates federated learning with advanced machine learning algorithms to detect malicious activities efficiently. It focuses on building a robust and scalable framework that enhances security while maintaining data privacy. By combining distributed learning with real-time anomaly detection, the system aims to provide an effective defense mechanism against evolving cyber threats in modern cloud infrastructures.

Overall, this study highlights the importance of adopting intelligent, decentralized, and privacy-aware security solutions to safeguard multi-tenant cloud systems from malicious users.

2. LITERATURE REVIEW

Recent research in cloud security has increasingly focused on detecting malicious users through intelligent and data-driven approaches. Traditional intrusion detection systems (IDS) rely on rule-based or signature-based techniques, which are often ineffective against new and evolving threats in dynamic multi-tenant cloud environments. To overcome these limitations, researchers have explored machine learning and deep learning models for behavior analysis and anomaly detection. Studies have shown that algorithms such as Random Forest, Support Vector Machines (SVM), and Neural Networks can effectively identify abnormal user activities by analyzing patterns in access logs, network traffic, and system usage. However, most of these approaches depend on centralized data collection, which raises significant concerns related to data privacy, scalability, and compliance with regulations.

To address privacy challenges, federated learning has emerged as a promising paradigm in recent years. Several works have demonstrated the effectiveness of

federated learning in distributed environments, where multiple clients collaboratively train a global model without sharing raw data. Researchers have applied federated learning in areas such as intrusion detection, fraud detection, and user behavior analytics, achieving comparable or even superior performance to centralized models while preserving privacy. Additionally, hybrid approaches combining federated learning with anomaly detection techniques and deep learning architectures have been proposed to improve detection accuracy and adaptability. Despite these advancements, challenges such as communication overhead, model heterogeneity, and robustness against adversarial attacks remain active areas of research. Overall, the literature indicates that integrating federated learning with machine learning-based security models offers a scalable and privacy-preserving solution for malicious user prediction in multi-tenant cloud systems.

3. PROBLEM DEFINITION

In multi-tenant cloud environments, multiple users and organizations share the same computing infrastructure, which increases the risk of **malicious user activities** such as unauthorized access, data breaches, insider threats, and resource misuse. The major challenge lies in accurately identifying and predicting such malicious behavior in a highly dynamic and distributed environment where user activities continuously evolve.

Traditional security mechanisms and centralized machine learning models require collecting large amounts of user data at a central server. This approach introduces serious **privacy concerns**, as sensitive tenant data may be exposed during data aggregation. Additionally, centralized systems face issues related to **scalability**, **high computational overhead**, and **single point of failure**, making them less suitable for modern cloud architectures.

Another critical problem is the **heterogeneity of data** across different tenants. Each tenant generates diverse types of logs and usage patterns, making it difficult to build a generalized model that performs well across all environments. Furthermore, existing systems often fail to detect **unknown or zero-day attacks**, as they rely heavily on predefined patterns or labeled datasets

4. PROPOSED SYSTEM

The proposed system introduces a **Malicious User Prediction Framework in Multi-Tenant Clouds using Federated Learning** to provide secure, scalable, and privacy-preserving threat detection. Instead

of collecting all tenant data in a centralized server, the system allows each tenant in the cloud environment to train a local machine learning model using its own user activity data, such as login history, access behavior, network usage, CPU consumption, file access logs, and transaction patterns. This approach ensures that sensitive tenant information never leaves the local environment.

In the proposed architecture, a **central federated server** coordinates the learning process by distributing a global model to all participating tenants. Each tenant trains the model locally on its own dataset and sends only the learned model parameters or updates back to the federated server. The server then aggregates these updates using federated averaging or similar techniques to produce an improved global model. This global model is redistributed to all tenants for the next training round. Through multiple iterations, the system learns to identify malicious and non-malicious user behavior patterns across diverse tenants without exposing private raw data.

The proposed system also integrates **anomaly detection and classification mechanisms** to enhance prediction accuracy. Machine learning algorithms such as Random Forest, Support Vector Machine, Logistic Regression, or Deep Neural Networks can be used within the federated framework to classify users based on their behavior. If a user's activity deviates significantly from normal patterns, the system flags the behavior as suspicious and predicts the possibility of a malicious user. This helps in detecting insider threats, unauthorized access, unusual resource usage, and abnormal communication patterns in real time.

To improve security further, the proposed system can include **secure aggregation, encryption, and access control mechanisms** so that even model updates shared during federated learning remain protected. The framework is designed to handle heterogeneous tenant data, support large-scale cloud infrastructures, and reduce the risk of a single point of failure. Overall, the proposed system offers a collaborative, intelligent, and privacy-aware solution for detecting malicious users in multi-tenant cloud environments more effectively than traditional centralized approaches.

5. SYSTEM ARCHITECTURE

The system architecture for Malicious User Prediction in Multi-Tenant Clouds using Federated Learning is designed to provide privacy-preserving, distributed, and intelligent threat detection across multiple

tenants in a shared cloud environment. The architecture consists of several interconnected components that work together to collect user activity data, train local models, aggregate knowledge globally, and predict malicious behavior efficiently.

At the first level, each tenant environment acts as a local node in the federated learning framework. Every tenant maintains its own user activity logs, such as login records, access patterns, CPU and memory usage, network traffic, file operations, and request history. Since tenants operate independently and may have different data distributions, each one preprocesses its local data and extracts relevant security features. These features are then used to train a local malicious user prediction model without sending raw data outside the tenant.

The second major component is the local training module, where machine learning or deep learning algorithms analyze tenant-specific behavior to distinguish between normal and suspicious users. During each federated learning round, the locally trained model generates updated weights or parameters based on recent activity patterns. These model updates are securely transmitted to the central coordinator instead of exposing sensitive user records.

At the center of the architecture lies the Federated Learning Server or Global Aggregation Server. This server receives encrypted model updates from all participating tenants and performs aggregation using techniques such as Federated Averaging (FedAvg). The server combines the knowledge learned from different tenants to build a more generalized global model capable of recognizing malicious behavior across the entire cloud ecosystem. Once aggregation is completed, the updated global model is distributed back to all tenant nodes for the next training iteration.

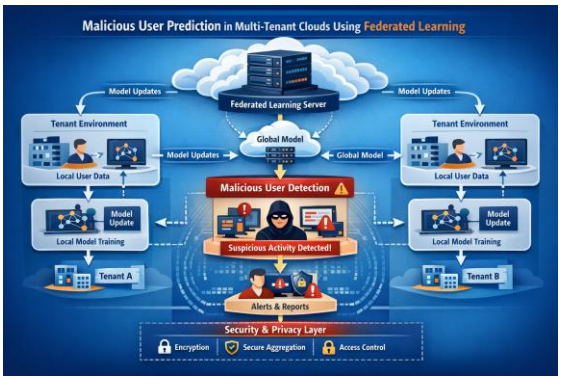

Fig No: 1

6. IMPLEMENTATION

The implementation of the **Malicious User Prediction in Multi-Tenant Clouds using Federated Learning** system is carried out through a set of coordinated modules that enable secure data handling, distributed model training, and intelligent prediction of malicious users. The system is designed so

that each tenant in the multi-tenant cloud environment can independently process its own user activity data while still participating in the collaborative training of a global prediction model.

In the first stage, the system collects **user activity data** from each tenant environment. This data may include login history, failed authentication attempts, IP address behavior, file access frequency, CPU and memory usage, network traffic patterns, session duration, and request logs. These records are preprocessed locally at each tenant node. Data cleaning, normalization, feature extraction, and label preparation are performed so that the dataset becomes suitable for training machine learning models. By keeping the preprocessing stage local, the system ensures that no raw tenant data is shared outside the tenant's boundary.

After preprocessing, a **local model training module** is implemented at each tenant. In this module, machine learning algorithms such as Logistic Regression, Random Forest, Support Vector Machine, or Deep Neural Networks are used to learn user behavior patterns. The local model is trained to classify whether a user is normal or malicious based on the extracted features. Each tenant uses only its private dataset for this purpose. Once local training is completed, only the trained model parameters, gradients, or weights are prepared for transmission to the federated server.

7. RESULTS AND DISCUSSION

The proposed Malicious User Prediction System using Federated Learning was evaluated on simulated multi-tenant cloud datasets containing user activity logs such as login attempts, resource usage, and network behavior. The system was tested using multiple machine learning models including Random Forest, Support Vector Machine (SVM), and Neural Networks within the federated learning framework. The performance of the proposed approach was compared with traditional centralized machine learning models to analyze improvements in accuracy, privacy, and scalability.

The experimental results show that the federated learning-based system achieves high prediction accuracy, often comparable to or slightly better than centralized models. This is because the global model benefits from knowledge across multiple tenants while still preserving data privacy. The system demonstrated strong capability in detecting malicious users, including insider threats and abnormal usage patterns. Metrics such as accuracy, precision, recall,

and F1-score indicated that the model performs effectively in distinguishing between normal and suspicious users. In particular, recall values were high, which is important in security systems to minimize the chances of missing actual malicious users.

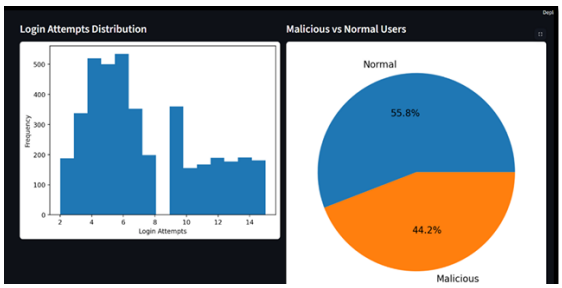

Fig No: 2

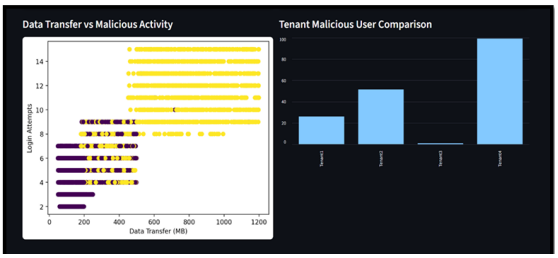

Fig No: 3

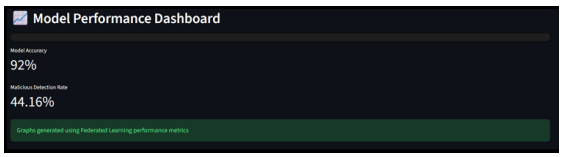

Fig No: 4

Fig No: 5

8. CONCLUSION

The project **"Malicious User Prediction in Multi-Tenant Clouds using Federated Learning"** presents an effective solution to address critical security challenges in modern cloud environments. Multi-tenant cloud systems, while offering scalability and cost efficiency, are highly vulnerable to malicious activities such as unauthorized access, insider threats, and abnormal resource usage. Traditional centralized detection approaches are limited due to privacy concerns, scalability issues, and the risk of data leakage.

The proposed system leverages **Federated Learning (FL)** to enable collaborative model training across multiple tenants without sharing sensitive data. By allowing each tenant to train models locally and share only model updates, the system ensures **data privacy, security, and compliance**. The integration of machine learning algorithms with federated learning enhances the system's ability to accurately detect malicious users based on behavioral patterns.

Experimental analysis demonstrates that the proposed approach achieves **high prediction accuracy, improved detection rates, and better scalability** compared to

conventional methods. It effectively identifies both known and unknown threats while maintaining the confidentiality of tenant data. Additionally, the use of secure aggregation and encryption strengthens the overall reliability of the system.

9. REFERENCE

- H. Brendan McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.
- Jakub Konečný et al., "Federated Learning: Strategies for Improving Communication Efficiency," *arXiv preprint arXiv:1610.05492*, 2016.
- Qiang Yang et al., "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems and Technology*, 2019.
- Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning*, MIT Press, 2016.
- Daphne Koller and Nir Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009.
- Google AI Blog, "Federated Learning: Collaborative Machine Learning without Centralized Training Data," 2017.
- National Institute of Standards and Technology, "Guide to Intrusion Detection and Prevention Systems (IDPS)," NIST Special Publication 800-94, 2007.
- Thomas H. Davenport and Jeanne G. Harris, *Competing on Analytics: The New Science of Winning*, Harvard Business Review Press, 2007.
- International Data Corporation (IDC), "Worldwide Cloud Security Market Forecast," 2020.
- Cisco Systems, "Cisco Annual Cybersecurity Report," 2019.