



A Robust Defence Strategy for Secure Federated Learning Using LOMAR-Based Validation

Sadia Nikhat¹, Adeeba Anjum², Mirza Danish Baig³, Hafsa Mohd Abdul Lateef⁴, Mohammed Abdul Munam⁵

^{1,2}Assistant Professor, Department of CSE (Data Science),
Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.

^{3,4,5}UG Students, Department of CSE (Data Science),
Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.

Abstract— This project presents a secure and efficient federated learning framework designed to improve model accuracy while defending against malicious updates. The system enables a centralized server and multiple clients to collaboratively train machine learning models without sharing raw data. A graphical interface is used to upload and preprocess the MNIST dataset, train models, and communicate with the server. The proposed approach incorporates LOMAR-based techniques to evaluate model integrity and detect poisoned updates. Experimental results show that genuine models achieve high accuracy, while malicious models are identified and rejected by the server. Additionally, model compression techniques are applied to reduce communication overhead, improving efficiency in distributed environments. Comparative graphs demonstrate that the proposed method outperforms traditional approaches in terms of accuracy and model size. Overall, the system ensures secure, scalable, and reliable federated learning suitable for real-world applications.

Keywords— Federated learning, model security, malicious update detection, LOMAR technique, MNIST dataset, model compression, distributed machine learning, data privacy.

I. INTRODUCTION

Federated learning has emerged as a powerful paradigm that enables multiple clients to collaboratively train machine learning models without sharing sensitive data [1][2]. This approach is particularly useful in privacy-sensitive environments such as healthcare, finance, and mobile applications. Instead of transferring raw data to a central server, each client trains a local model and shares only the model updates. The

central server then aggregates these updates to build a global model. This decentralized training process helps preserve user privacy while still benefiting from distributed data sources [2]. However, federated learning systems face several challenges, including communication overhead, model inconsistency, and vulnerability to malicious attacks [5][15]. Ensuring the reliability and accuracy of aggregated models is critical for practical deployment. Therefore, designing secure and efficient federated learning frameworks has become an important research area in modern machine learning systems [1].

One of the major challenges in federated learning is the presence of adversarial or poisoned models [4][9]. Malicious clients may intentionally upload corrupted updates to degrade the performance of the global model. These attacks can significantly impact the accuracy and reliability of the system [10]. Detecting such harmful contributions is difficult because the server does not have access to the original training data. To address this issue, robust validation techniques must be implemented to differentiate between genuine and malicious model updates [11][12]. The proposed system integrates advanced techniques to identify poisoned models based on their performance and behavior. By analyzing accuracy variations and model characteristics, the system can effectively reject suspicious updates. This enhances the overall robustness of the federated learning framework and ensures that only reliable contributions are considered during model aggregation [13].

To improve the reliability of the system, the proposed approach incorporates LOMAR-based techniques for model validation. These techniques allow the server to independently evaluate the received models using its own test dataset. By comparing the performance of incoming models, the server can determine whether they are genuine



or potentially poisoned. Experimental results show that genuine models achieve high accuracy, while malicious models exhibit lower performance and are rejected. This selective aggregation process enhances the quality of the global model. Furthermore, the system provides visual representations of accuracy comparisons through graphs, making it easier to analyze performance differences between techniques. Such evaluation mechanisms are essential for building trustworthy federated learning systems that can operate effectively in real-world scenarios [14][16].

II. RELATED WORK

Konečný et al., [2017] [1] Konečný and colleagues introduced federated learning as a distributed approach focused on improving communication efficiency during model training. Their work emphasizes reducing the amount of data exchanged between clients and the central server by sharing only model updates instead of raw datasets. They proposed optimization strategies such as structured updates and compression techniques to minimize bandwidth usage. The study highlights the importance of scalability in large networks with multiple devices. It also discusses challenges like system heterogeneity and limited client resources. Their approach significantly reduces communication costs while maintaining model performance. This work laid the foundation for modern federated learning systems. It is widely considered a key contribution in enabling privacy-preserving machine learning. Overall, the paper provides important insights into efficient distributed training mechanisms.

McMahan et al., [2017] [2] McMahan and co-authors presented a communication-efficient method for training deep neural networks using decentralized data. They introduced the Federated Averaging (FedAvg) algorithm, which allows multiple clients to train models locally and send updates to a central server for aggregation. Their approach significantly reduces the number of communication rounds required for convergence. The study demonstrates how federated learning can be applied to real-world applications such as mobile devices. It also addresses issues related to data privacy by ensuring that user data remains on local devices. Experimental results show that the proposed method achieves high accuracy with reduced communication overhead. The paper highlights the balance between local computation and global model updates. This research is considered a foundational work in federated learning. It has influenced many subsequent studies in distributed machine learning.

Bagdasaryan et al., [2018] [4] Bagdasaryan and team explored the vulnerabilities of federated learning systems to backdoor attacks. Their study demonstrates how malicious participants can manipulate model updates to introduce hidden behaviours into the global model. These attacks are difficult to detect because they maintain high accuracy on normal data while behaving incorrectly on specific inputs. The authors proposed attack strategies that exploit the aggregation process of federated learning. They also discussed the limitations of existing defence mechanisms in detecting such attacks. The research highlights the need for robust validation techniques to ensure model integrity. Their findings reveal serious security concerns in collaborative learning environments. This work has encouraged further research into secure federated learning systems. It remains a significant contribution in understanding adversarial threats.

Fang et al., [2020] [9] Fang and colleagues investigated local model poisoning attacks in federated learning environments. Their research focuses on how attackers can manipulate local updates to degrade the global model's performance. The study evaluates the effectiveness of various Byzantine-robust aggregation methods against such attacks. Experimental results show that even advanced defence techniques can be vulnerable under certain conditions. The authors also proposed improved attack strategies that can bypass traditional defences. Their work highlights the limitations of existing robustness mechanisms in distributed learning. It emphasizes the need for stronger security frameworks to protect against adversarial behaviour. The paper contributes to a deeper understanding of poisoning attacks. It also provides insights into improving model aggregation strategies. Overall, the study is important for enhancing federated learning security.

Blanchard et al., [2017] [11] Blanchard and co-authors introduced Byzantine-tolerant gradient descent methods for secure distributed machine learning. Their work focuses on handling malicious or faulty participants during the training process. They proposed aggregation rules such as Krum to filter out harmful updates before combining them into the global model. The study demonstrates that traditional averaging methods are not reliable in adversarial environments. Their approach ensures that only trustworthy updates contribute to model training. The research provides theoretical guarantees for robustness against Byzantine failures. Experimental evaluations confirm the effectiveness of their proposed method. This work is considered a key contribution to secure distributed optimization. It has influenced many later studies in federated learning defense

mechanisms. Overall, it strengthens the reliability of collaborative learning systems.

III. DATASET DETAILS

The dataset used in this project is the MNIST dataset, which is a widely recognized benchmark in the field of machine learning and computer vision. It consists of grayscale images representing handwritten digits from 0 to 9, making it suitable for classification tasks. Each image is of size 28×28 pixels, where pixel intensity values range from 0 to 255. The dataset contains a total of 70,000 images, divided into 60,000 training samples and 10,000 testing samples. These images are labeled, allowing supervised learning techniques to be applied effectively. The simplicity and structured nature of the dataset make it ideal for evaluating different machine learning models. In this project, the dataset is uploaded through the client interface, enabling users to interactively load and visualize the data. Its balanced distribution across digit classes ensures fair training and testing, which helps in accurately measuring model performance.

Before training the model, the dataset undergoes a preprocessing stage to improve learning efficiency and accuracy. The primary step involves normalization, where pixel values are scaled to a range between 0 and 1. This transformation helps in faster convergence during training and reduces computational complexity. Additionally, the dataset is reshaped and formatted to match the input requirements of the machine learning model. Proper preprocessing ensures consistency across all samples and eliminates variations that may negatively impact performance. In the federated learning setup, each client processes its local dataset independently, preserving data privacy. After preprocessing, the dataset is used to train both genuine and poisoned models for evaluation purposes. The system then analyzes model performance based on accuracy and behavior. This structured dataset handling plays a crucial role in achieving reliable and efficient results in the proposed framework.

IV. PROPOSED METHODOLOGY

The proposed system follows a federated learning approach where a central server coordinates with multiple client applications to train a global model without sharing raw data. Initially, the server is started, and clients connect to it to participate in the training process. Each client uploads the MNIST dataset and performs preprocessing steps such as normalization to prepare the data. After preprocessing, a genuine model is trained locally using the dataset. This trained model is then sent to

the server, where it is evaluated and aggregated with other models. This process ensures privacy preservation and efficient distributed learning.

To enhance security, the system introduces mechanisms to detect and prevent model poisoning attacks. Along with genuine models, clients may also upload manipulated or poisoned models to test system robustness. The server applies LOMAR-based validation techniques to evaluate incoming models using its own test data. Models with abnormal performance are identified and rejected, ensuring only reliable updates contribute to the global model. Additionally, model compression techniques are used to reduce communication overhead and improve efficiency. Performance comparisons are visualized using graphs, highlighting accuracy and model size differences between various approaches.

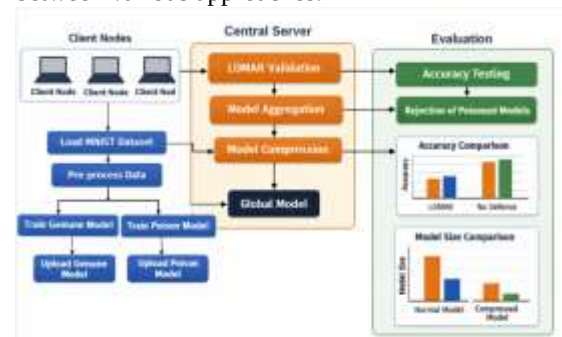


Figure [1] : Federated Learning System Architecture with LOMAR-Based Security

Figure[1] This diagram shows how multiple client nodes train models using local data without sharing it directly. Each client preprocesses the dataset and uploads either genuine or poisoned models to the central server. The server applies LOMAR validation to detect and reject malicious updates before aggregating models. Finally, the system evaluates performance using accuracy and model size comparison graphs.

V.RESULT AND DISCUSSION

The experimental results demonstrate the effectiveness of the proposed federated learning system in terms of accuracy, security, and efficiency. When the genuine model was trained using the MNIST dataset and uploaded to the server, it achieved a high accuracy of around 98% at the client side, while the server independently validated it with approximately 99% accuracy on its test data. This indicates that the model performs consistently across different environments. In contrast, when a poisoned model was uploaded, the system identified irregularities, and the server



rejected the update, preventing it from affecting the global model. The poisoned model showed a significantly lower accuracy of around 88%, which helped in distinguishing it from genuine contributions. Additionally, graphical comparisons showed that the proposed LOMAR-based technique achieved higher accuracy compared to systems without defense mechanisms. Model size comparison results also revealed that compressed models required less storage and transmission space than normal models. These findings confirm that the system not only improves model accuracy but also enhances communication efficiency and security in federated learning.



Figure [2] : Client Application Interface for Federated Learning with LoMar Defense

Figure [2] This screen displays the client-side interface used to perform federated learning tasks. It allows users to upload the MNIST dataset, preprocess data, and train models locally. Users can send both genuine and poisoned models to the server for validation. Additional options provide accuracy comparison and model size analysis through graphs.



Figure [3] : Local Model Training and Dataset Processing Dashboard

Figure [3] The interface shows options to upload and process a dataset for training. It displays details about dataset size, number of classes, and total samples. Training progress and duration are summarized after completion. Options are available to upload models and view accuracy or size graphs.

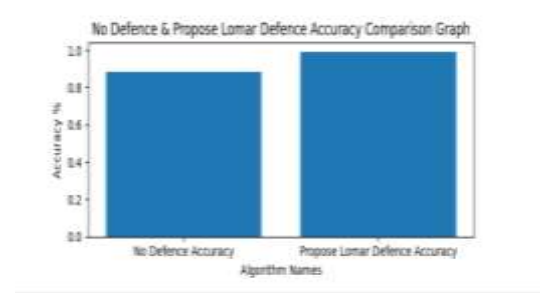


Figure [4] : Accuracy Comparison Between No Defence and Lomar Defence

Figure [4] The graph compares accuracy results of two approaches. The no defence method shows lower accuracy performance. The proposed Lomar defence achieves higher accuracy close to 100%. This indicates improved model performance with the defence mechanism.

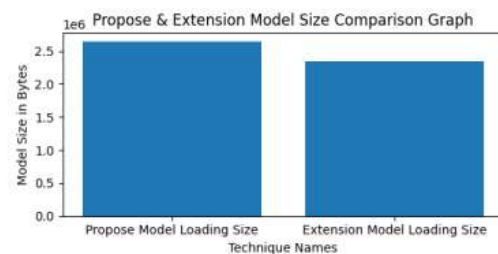


Figure [5] : Model Size Comparison Between Proposed and Extension Techniques

Figure[5] The graph presents the model loading size for two techniques. The proposed model has a slightly larger size compared to the extension model. The extension model reduces the overall model size. This shows improved efficiency in storage with the extension approach.

DISCUSSION

The results highlight the importance of incorporating security mechanisms in federated learning systems to handle adversarial conditions effectively. The ability of the server to differentiate between genuine and poisoned models ensures that only reliable updates contribute to the global model, improving overall performance. The use of LOMAR validation techniques proved to be effective in detecting malicious behavior based on model accuracy and performance patterns. This approach reduces the risk of model corruption and enhances trust in distributed learning environments. Furthermore, the implementation of model compression plays a crucial role in optimizing communication between clients and the server, especially in bandwidth-limited scenarios. By reducing model size without significantly affecting accuracy, the system becomes more scalable and efficient. The visual graphs provided in the system



help in better understanding performance variations and comparing different approaches. Overall, the combination of security validation and efficiency improvements makes the proposed system suitable for real-world applications where data privacy and reliability are critical concerns.

VI. CONCLUSION

The proposed system demonstrates an effective approach to secure and efficient federated learning by integrating validation and optimization techniques. By allowing clients to train models locally, the system ensures data privacy while still achieving high accuracy through collaborative learning. The implementation of LOMAR-based validation helps in identifying and rejecting poisoned models, thereby protecting the integrity of the global model. Experimental results confirm that genuine models maintain strong performance, while malicious updates are successfully filtered out. In addition, the use of model compression reduces communication overhead, making the system more efficient and scalable. The graphical analysis further supports the improvement in both accuracy and model size. Overall, the system provides a balanced solution that addresses key challenges in federated learning, including security, efficiency, and reliability, making it suitable for real-world distributed applications.

REFERENCES

1. J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, Federated learning: Approaches to enhance communication efficiency, 2017.
2. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Arcas, Efficient training of deep neural networks using decentralized data, presented at Artificial Intelligence and Statistics, 2017, pp. 1273-1282.
3. F. Dabek, R. Cox, F. Kaashoek, and R. Morris, Vivaldi: A system for decentralized network coordinate estimation, ACM SIGCOMM Computer Communication Review, vol. 34, pp. 15-26, 2004.
4. E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, Backdoor attack strategies in federated learning systems, arXiv preprint arXiv:1807.00459, 2018.
5. A. N. Bhagoji, S. Chakraborty, P. Mittal, and S. Calo, Federated learning under adversarial conditions, in Proceedings of the International Conference on Machine Learning, 2019, pp. 634-643.
6. L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, Information leakage risks in collaborative learning environments, IEEE Symposium on Security and Privacy, 2019, pp. 691-706.
7. H. Xiao, H. Xiao, and C. Eckert, Label manipulation attacks on support vector machines, presented at ECAI, 2012, pp. 870-875.
8. C. Fung, C. J. Yoon, and I. Beschastnikh, Addressing sybil attacks in federated learning poisoning scenarios, arXiv preprint arXiv:1808, 2018.
9. M. Fang, X. Cao, J. Jia, and N. Gong, Poisoning attacks targeting local models in robust federated learning, USENIX Security Symposium, 2020, pp. 1605-1622.
10. A. N. Bhagoji, S. Chakraborty, S. Calo, and P. Mittal, Exploring model poisoning threats in federated learning, presented at the SecML Workshop, NeurIPS, 2018.
11. P. Blanchard, R. Guerraoui, J. Stainer, et al., Gradient descent techniques resilient to Byzantine failures, Advances in Neural Information Processing Systems, 2017, pp. 119-129.
12. D. Yin, Y. Chen, R. Kannan, and P. Bartlett, Towards optimal robust distributed learning under Byzantine settings, International Conference on Machine Learning, 2018, pp. 5650-5659.
13. E. M. El Mhamdi, R. Guerraoui, and S. Rouault, Hidden weaknesses in Byzantine distributed learning systems,



International Conference on Machine Learning, 2018.

14. N. Baracaldo, B. Chen, H. Ludwig, and J. Safavi, Preventing poisoning attacks using data provenance techniques, ACM Workshop on Artificial Intelligence and Security, 2017, pp. 103–110.
15. M. Barreno, B. Nelson, A. Joseph, and J. Tygar, Security challenges in machine learning systems, Machine Learning Journal, vol. 81, no. 2, pp. 121–148, 2010.
16. C. Xie, O. Koyejo, and I. Gupta, Generalized stochastic gradient descent methods tolerant to Byzantine faults, arXiv preprint arXiv:1802.10116, 2018.
17. Gummadi, V. P. K. (Ed.). (2025). MuleSoft intelligent document processing: Transforming enterprise document workflows through AI-driven automation. Journal of Computational Analysis and Applications, 34(12). <https://doi.org/10.48047/jocaaa.2025.34.12.16>
18. Akinapalli, S. (2026). An AI-powered data trust and quality scoring framework for enterprise decision intelligence systems. International Journal of Data Science and IoT Management System, 5(1), 946–950.
19. Babburi, S. Lightweight Distributed Provenance Framework for Edge and IoT Data Systems.
20. Gaddam, S. From Fixed Specifications to Self-Adapting Systems: A Machine Learning Perspective on Software Engineering.
21. Immadi, S. K. (2025). Optimizing ERP for Human Capital Management. Applied Research for Growth, Innovation and Sustainable Impact, 377–384. <https://doi.org/10.1201/9781003684657-63>
22. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
23. Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
24. Mahimalur, R. K., Vasgam, M., & Manoharan, D. Devops Lifecycle Management And Cloud Migration Assessments: A Security-Driven CICD Perspective.
25. Viswanathan, V. Generative AI for Smarter Workforce Planning and Enterprise Resource Decisions.
26. Gajula, S. (2026, March). Two Pillars of Banking Intelligence: A Comparative Analysis of AI Techniques for Fraud Prevention and Churn Mitigation. In 2026 14th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE.
27. Maturi, S. Y. (2023). Crowdsourced frontier: Unveiling autonomous adversarial cybercapabilities via open AI competition. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 275–284.
28. Venkata Ramana, P. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. International Journal of Research in Information Technology and Computing, 8(4).
29. Srikanth Kavuri. (2025). AI-DRIVEN TEST AUTOMATION FRAMEWORKS: ENHANCING EFFICIENCY AND ACCURACY IN SOFTWARE QUALITY ASSURANCE. International Journal of Applied Mathematics, 38(10s), 699–710. <https://doi.org/10.12732/ijam.v38i10s.990>
30. Venkata Pavan Kumar Gummadi. (2026). Infrastructure Optimization Techniques for Enterprise Integration Platforms: A Comprehensive Analysis. Computer Fraud and Security, 37–44. <https://doi.org/10.52710/cfs.875>
31. Shashank, A. (2025). AI-Enhanced ETL Processes: Leveraging Artificial Intelligence for Optimized Data Integration Systems. Journal Of Multidisciplinary, 5(8), 219–225.
32. Kandula, S. T. R., Susarla, R. S., & Boyapati, P. K. (2025, July). Enhanced Cyber Security Using Global Local Artificial Neural Network Based Intrusion Detection in Big Data Environment. In 2025 IEEE 4th World Conference on



Applied Intelligence and Computing (AIC)
(pp. 426-431). IEEE.

33. Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. IJSAT-Int. J. Sci. Technol. 16 (2).

<https://doi.org/10.71097/IJSAT.v16.i2.3219>

9