2026 Volume: 02 Issue: 05 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-12

FEDERATED LEARNING-BASED INTRUSION DETECTION IN IOT NETWORKS WITH TRUST-AWARE AGGREGATION

*[1]Varsha C. Parihar, [2]Subramanya R., [2]Suraj S. Chinivar

[1]Assistant Professor Dept. of CSE Jyothy Institute of Technology Bangalore, India.

[2]B.E. (CSE) Dept. of CSE Jyothy Institute of Technology Bangalore, India.

Article Received: 08 April 2026

Article Revised: 28 April 2026

Published on: 18 May 2026

***Corresponding Author: Varsha C. Parihar**

Assistant Professor Dept. of CSE Jyothy Institute of Technology Bangalore, India.

DOI: https://doi-doi.org/101555/ijrpa.7499

ABSTRACT

The rapid proliferation of Internet of Things (IoT) devices has significantly increased the volume and complexity of network traffic, making in- trusion detection a critical component of modern cy- bersecurity systems. Traditional Intrusion Detection Systems (IDS) rely on centralized architectures, which introduce challenges such as data privacy risks, high communication overhead, and limited scalability. Fed- erated Learning (FL) has emerged as a promising decentralized machine learning paradigm that enables collaborative model training without sharing raw data, thereby preserving privacy and reducing communica- tion costs.

However, existing FL-based intrusion detection sys- tems assume that all participating clients are trust- worthy, which is not realistic in practical IoT environ- ments where devices may be compromised. Malicious clients can inject poisoned model updates, degrading the performance and reliability of the global model. This paper presents a comprehensive literature sur- vey of Federated Learning-based intrusion detection systems, highlighting their strengths, limitations, and vulnerability to adversarial attacks.

Based on the identified research gaps, a trust-aware federated intrusion detection framework is proposed, incorporating client reliability evaluation and robust aggregation mechanisms. The proposed approach aims to enhance detection accuracy, improve system robust- ness, and ensure resilience against malicious partici- pants while maintaining data privacy. This work pro- vides a foundation for developing secure and scalable intrusion detection systems in distributed IoT environ- ments.

I. INTRODUCTION

The rise of the Internet of Things (IoT) has led to a mas- sive deployment of interconnected

devices such as smart cameras, sensors, wearable systems, and home automation units. These devices continuously generate and exchange network traffic in the form of structured data packets containing attributes such as source, destination, protocol type, packet size, and communication patterns. However, with the growing scale and heterogeneity, the IoT networks become highly vulnerable to the cyber threats. Hence, there is a need of robust and intelligent intrusion detection mechanisms.

Intrusion Detection Systems (IDS) are commonly used to identify malicious activity by detecting anomalies in the usual behavior of a network. Traditional IDS approaches based on signature-based and anomaly-based methods are typically based on centralized architectures. In such architectures, data from multiple devices is collected and processed on a central server. However, these centralized systems suffer from serious challenges, such as data privacy risks, high communication overhead and limited scalability especially in large-scale IoT scenarios.

Federated Learning (FL) is a distributed machine learn- ing paradigm in which a number of clients collaboratively train a global model without sharing their raw data. In FL, each client trains on its local data and sends the model updates to a central server only, which aggregates them by algorithms such as Federated Averaging (FedAvg). This approach can preserve data privacy and reduce the communication costs, which is well suited for IoT based intrusion detection systems.

However, despite these advantages, existing FL-based IDS frameworks generally assume that all the participat- ing clients are trustworthy. In practical IoT environments, devices are often exposed to adversarial attacks and possi- bly compromised. Malicious clients can inject poisoned or manipulated updates during the training process, resulting in model degradation and reduced detection accuracy. Such attacks are often called poisoning or Byzantine at- tacks and are a major concern of the reliability of federated systems.

This paper presents a comprehensive review of the literature on Federated Learning-based intrusion detection systems for IoT networks with an emphasis on finding existing limitations and security vulnerabilities. We pro- pose a federated intrusion detection framework with trust awareness based on the insights, with mechanisms to assess the reliability of the clients and alleviate the impact of malicious participants. The proposed approach aims at im- proving the detection accuracy, increasing the robustness of the system and ensuring secure collaborative learning in distributed IoT environments.

A. Problem Statement

Such traditional intrusion detection systems rely on cen- tralized data collection, which presents serious challenges in terms of data privacy, communication overhead, and scalability in large scale Internet of Things environments. Federated Learning addresses the above concerns by allow- ing decentralized model training without the need to share raw data. However, current Federated Learning based intrusion detection systems assume that all the clients participated in the training process are honest.

In practical IoT deployments, devices are often vul- nerable to compromise and can act as adversaries. Such malicious clients can inject manipulated or poisoned model updates during the aggregation phase, which leads to degradation of the global model performance and reliabil- ity. These types of attacks, also called poisoning or Byzan- tine attacks, seriously threaten the integrity of federated learning systems.

However, existing federated learning-based intrusion de- tection frameworks lack effective trust management and robust aggregation mechanisms, representing a critical research gap. Addressing this limitation is essential for achieving secure, reliable, and resilient intrusion detection in distributed IoT environments.

B. Objectives

- Investigate existing intrusion detection techniques in IoT environments and their limitations
- To analyze the role of Federated Learning in enabling privacy-preserving intrusion detection
- Examine vulnerabilities in current FL-based systems, particularly in the presence of adversarial or compro- mised clients
- To design a trust-aware Federated Learning-based in- trusion detection framework with robust aggregation mechanisms
- To enhance detection accuracy, reliability, and re- silience of the system against malicious model updates

II. LITERATURE SURVEY

A. Federated Learning Fundamentals

Federated Learning (FL) has emerged as a decentralized machine learning paradigm that enables multiple clients to collaboratively train a global model without sharing raw data. McMahan et al. [1] introduced the concept of Federated Learning and proposed the Federated

Averag- ing (FedAvg) algorithm, which aggregates locally trained model updates to form a global model. This approach significantly reduces communication overhead while pre- serving user data privacy.

Kairouz et al. [2] provided a comprehensive survey of Federated Learning, covering its theoretical foundations, system architectures, and practical applications. Their work highlights key advantages such as privacy preser- vation and scalability, while also identifying challenges including communication efficiency, system heterogeneity, and robustness.

B. Federated Learning in IoT Environments

The application of Federated Learning in IoT networks has gained significant attention due to the distributed and resource-constrained nature of IoT devices. Nguyen et al.

[3] explored the integration of FL in IoT systems and identified major challenges such as device heterogeneity, intermittent connectivity, and limited computational re- sources. These factors can impact model convergence and system performance.

Li et al. [4] further discussed about the challenges in Federated Learning, including statistical heterogeneity of data, communication bottlenecks, and scalability issues. Their work emphasizes the need for efficient communica- tion protocols and adaptive learning mechanisms in large- scale distributed environments.

C. Intrusion Detection Systems using Machine Learning

Intrusion Detection Systems (IDS) are essential for identifying malicious activities in network environments. Traditional IDS approaches can be broadly classified into signature-based and anomaly-based methods. Signature- based systems rely on known attack patterns, while anomaly-based systems detect deviations from normal behavior.

Recent advancements in machine learning have signif- icantly improved IDS performance. Shone et al. [5] pro- posed a deep learning-based IDS using stacked autoen- coders for feature learning, achieving improved detection accuracy. Similarly, Alrawashdeh et al. [6] developed an online intrusion detection system using deep learning tech- niques, demonstrating the effectiveness of neural networks in detecting complex attack patterns.

However, these approaches rely on centralized data collection, which raises concerns related to data privacy, scalability, and communication overhead in distributed IoT environments.

D. Federated Learning-Based Intrusion Detection

To address privacy concerns, several studies have ex- plored the use of Federated Learning for intrusion detec- tion. Albanbay et al. [7] evaluated a Federated Learning- based IDS for IoT networks and demonstrated its ability

to detect anomalies while preserving data privacy. Their results indicate that FL can achieve comparable perfor- mance to centralized approaches.

Despite these advantages, their approach assumes that all participating clients are trustworthy, which is not real- istic in practical IoT deployments. The absence of mecha- nisms to evaluate client reliability limits the robustness of such systems.

E. Security Threats in Federated Learning

While Federated Learning enhances privacy, it intro- duces new security challenges. Zhang et al. [8] studied poisoning attacks in Federated Learning and demonstrated how malicious clients can manipulate local model updates to degrade the performance of the global model. These attacks can significantly impact model convergence and accuracy.

Adversarial participants may perform data poisoning or model poisoning attacks, commonly referred to as Byzantine attacks, where malicious updates are intention- ally crafted to disrupt the learning process. Existing FL frameworks lack robust mechanisms to defend against such threats, making them vulnerable in adversarial environ- ments.

F. Datasets for Intrusion Detection

Datasets play a crucial role in evaluating intrusion detection systems. The UNSW-NB15 dataset [9] is widely used for benchmarking IDS performance, as it contains a diverse set of modern attack scenarios and realistic network traffic features. It provides a balanced representa- tion of normal and malicious activities, making it suitable for training and evaluating machine learning-based IDS models.

G. Research Gap

From the literature, it is evident that while Feder- ated Learning provides a promising solution for privacy- preserving intrusion detection, existing approaches pri- marily assume that all participating clients are reliable. This assumption is unrealistic in real-world IoT environ- ments, where devices may be compromised.

Furthermore, current systems lack effective trust man- agement and robust aggregation mechanisms to mitigate the impact of malicious clients. Addressing these limi- tations is essential for developing secure, reliable, and scalable Federated Learning-based intrusion detection sys- tems.

III. Comparative Analysis

A comparative analysis of existing Federated Learning- based intrusion detection systems is essential to under-stand their strengths, limitations, and applicability in real- world IoT environments. Table I summarizes key contri- butions, techniques, and limitations of

prominent works discussed in the literature.

Recent studies have also explored Federated Learning for IoT security applications [10], [11]. Datasets such as Edge-IIoTset [12] further enhance evaluation of intrusion detection systems. Systems such as DïoT [13] demonstrate practical implementations of federated anomaly detection. Adversarial perspectives on Federated Learning have also been explored in [14].

From the comparative analysis, it is evident that while Federated Learning significantly improves privacy and scalability, most existing approaches lack mechanisms to handle adversarial or compromised clients. Additionally, traditional intrusion detection systems, although effective in detecting complex attack patterns, rely on centralized architectures and fail to preserve data privacy.

Furthermore, existing FL-based intrusion detection frameworks primarily focus on improving detection ac- curacy and communication efficiency, but do not ade- quately address security threats such as model poisoning and Byzantine attacks. This highlights the necessity of integrating trust-aware mechanisms and robust aggrega- tion strategies to enhance the reliability and resilience of Federated Learning systems in IoT environments.

IV. Proposed Framework

A. System Architecture

The proposed system is a Federated Learning-based Intrusion Detection System (FL-IDS) designed for dis- tributed IoT environments. In this framework, multiple IoT devices or edge nodes act as federated clients that locally collect and process network traffic data. Each client trains a local intrusion detection model using its own data without sharing raw information.

A central server coordinates the learning process by ag- gregating local model updates received from clients. Unlike traditional Federated Learning approaches, the proposed system incorporates a trust-aware aggregation mechanism to evaluate the reliability of each participating client. This helps mitigate the impact of malicious or compromised devices.

The overall workflow consists of iterative communica- tion rounds, where clients perform local training and send model updates to the server, which then computes an updated global model and redistributes it to the clients.

B. Trust-Aware Aggregation Mechanism

To address the limitations of traditional Federated Learning, a trust-based mechanism is introduced during the aggregation phase. Each client is assigned a trust score

TABLE I: COMPARATIVE ANALYSIS OF EXISTING FEDERATED LEARNING-BASED INTRUSION DETECTION SYSTEMS.

Reference	Technique Used	Dataset	Strengths	Limitations
McMahan et al. [1]	Federated Averag- ing (FedAvg)	N/A	Introduced Federated Learning; reduced communication cost	Does not address security or adversarial clients
Kairouz et al. [2]	FL Survey	Multiple	Comprehensive overview of FL concepts and challenges	Lacks implementation-specific solutions
Nguyen et al. [3]	FL in IoT Systems	IoT datasets	Addresses scalability and IoT integration	High communication over-head and heterogeneity is-sues
Li et al. [4]	FL Optimization	Multiple	Discusses robustness, scalability, and challenges	Does not provide concrete IDS solutions
Albanbay et al. [7]	FL-based IDS	UNSW-NB15	High detection accuracy with privacy preservation	Assumes allclients are trustworthy
Zhang et al. [8]	Attack Analysis in FL	N/A	Identifies vulnerabili-ties such as poisoning attacks	No defense mechanism proposed
Shone et al. [5]	Deep Learning IDS	NSL-KDD	High detection accuracy using autoencoders	Centralized approach; lacks privacy preservation
Alrawashdeh et al. [6]	Online Deep Learn- ing IDS	KDD-based datasets	Real-time detection capability	Centralized and resource-intensive

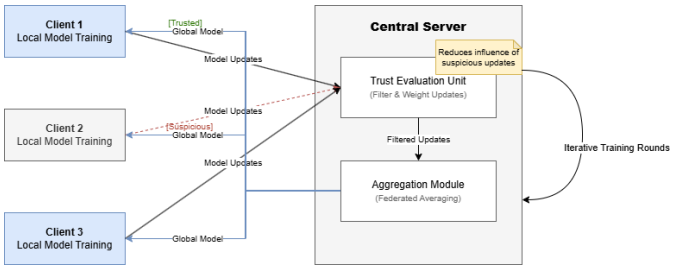

Fig. 1. Proposed architecture of the Federated Learning-based Intrusion Detection System with trust-aware aggregation.

based on its historical behavior, consistency of updates, and deviation from the global model. Let w_i represent the model update from the i^{th} client and T_i represent its corresponding trust score. The global model is computed as:

$$w_{global} = \sum_{i=1} T_i \cdot w_i \qquad (1)$$

where N is the total number of participating clients and

$$\Sigma T_i = 1.$$

Clients with higher trust scores contribute more signif- icantly to the global model, while suspicious or anoma- lous updates are assigned lower weights. This approach enhances robustness against poisoning and Byzantine at- tacks.

C. Workflow Description

The proposed system operates in the following steps:

1. **Initialization:** The central server initializes the global model and distributes it to all participating clients.
2. **Local Training:** Each client trains the model using its local IoT traffic data.
3. **Update Transmission:** Clients send their locally trained model parameters to the central server.
4. **Trust Evaluation:** The server evaluates each client based on update behavior and assigns trust scores.
5. **Aggregation:** The server aggregates model updates using the trust-aware weighted aggregation mecha- nism.
6. **Model Distribution:** The updated global model is sent back to clients for the next round of training.

This iterative process continues until the model con- verges or achieves desired performance.

D. Dataset Considerations

The effectiveness of the proposed system can be evalu- ated using benchmark intrusion detection datasets such as UNSW-NB15 and CICIoT2023. These datasets contain re- alistic network traffic patterns, including both normal and malicious activities, making them suitable for training and testing intrusion detection models in IoT environments.

E. Implementation Environment

The proposed system can be implemented using Python as the primary programming language, with machine learning frameworks such as TensorFlow or PyTorch for model development. Federated Learning simulations can be performed using frameworks such as TensorFlow Fed- erated or PySyft.

For visualization and user interaction, lightweight web frameworks such as Flask or Streamlit can be utilized. Development and experimentation can be carried out in environments such as

Jupyter Notebook or integrated development environments like Visual Studio Code.

The hardware requirements for implementing the sys- tem include a multi-core processor (Intel i5 or higher), a minimum of 8 GB RAM, and sufficient storage capacity. These resources are adequate for performing local training and federated aggregation in a simulated IoT environment.

V. Results and Discussion

This section presents an analytical discussion of the ex- pected performance of the proposed Federated Learning- based intrusion detection system with trust-aware aggre- gation. Since the framework is designed to address key lim- itations in existing FL-based IDS models, its effectiveness can be evaluated based on robustness, detection accuracy, and resilience to adversarial attacks.

A. Expected Performance Improvements

The integration of a trust-aware aggregation mechanism is expected to enhance the overall reliability of the global model. Unlike traditional Federated Learning approaches that treat all client updates equally, the proposed system assigns weights based on client trust scores. This reduces the influence of malicious or compromised clients, thereby improving model convergence and stability.

In comparison to standard Federated Averaging (Fe- dAvg), the proposed approach is expected to achieve higher detection accuracy, particularly in scenarios where a subset of clients behaves adversarially. By filtering or down-weighting suspicious updates, the model can main-tain consistent performance even in the presence of noisy or poisoned data.

B. Robustness Against Adversarial Attacks

One of the major challenges in Federated Learning is vulnerability to poisoning and Byzantine attacks. In such attacks, adversarial clients attempt to manipulate the global model by submitting malicious updates. The proposed trust-aware mechanism mitigates this risk by evaluating the consistency and deviation of each client's updates before aggregation.

As a result, the system is expected to exhibit improved robustness against:

- Model poisoning attacks
- Data poisoning attacks
- Byzantine failures

Clients that deviate significantly from the expected update patterns are assigned lower trust scores, thereby reducing their impact on the global model, compared to existing approaches

[4], [8]

C. Comparison with Existing Approaches

Compared to existing FL-based intrusion detection sys- tems, the proposed framework offers several advantages. Traditional systems focus primarily on privacy preserva- tion and distributed training, but often overlook security threats within the federated environment. By incorpo- rating trust evaluation into the aggregation process, the proposed system enhances both security and reliability.

Additionally, centralized IDS models, although effective in detecting complex attack patterns, suffer from privacy and scalability limitations. The proposed approach over- comes these issues by combining decentralized learning with trust-aware mechanisms, making it more suitable for real-world IoT deployments.

D. DISCUSSION

The proposed framework demonstrates a balanced ap- proach to intrusion detection by integrating privacy preservation, scalability, and security. While the absence of experimental validation is a limitation, the analytical insights and comparison with existing methods indicate that trust-aware aggregation can significantly improve the robustness of Federated Learning systems.

Future implementation and experimental evaluation us- ing benchmark datasets such as UNSW-NB15 and CI- CIoT2023 can further validate the effectiveness of the proposed approach. Metrics such as accuracy, precision, recall, and F1-score can be used to quantify performance improvements.

Overall, the proposed system provides a promising di- rection for developing secure and reliable intrusion detec-tion solutions in distributed IoT environments.

VI. FUTURE SCOPE

The proposed trust-aware Federated Learning-based in- trusion detection framework opens several directions for future research and development. One potential extension is the integration of advanced trust evaluation mecha- nisms using statistical or learning-based approaches to dy- namically assess client behavior over time. Incorporating blockchain technology can further enhance transparency and security by providing a decentralized and tamper- proof record of client interactions and model updates.

Additionally, future work can explore the implemen- tation of secure aggregation techniques to prevent infor- mation leakage during model update transmission. The use of lightweight

models and optimization techniques can also improve the feasibility of deploying the system on resource-constrained IoT devices.

Experimental validation using real-world datasets such as UNSW-NB15 and CICIoT2023, along with performance evaluation using metrics like accuracy, precision, recall, and F1-score, can further strengthen the effectiveness of the proposed approach. Moreover, extending the frame- work to support real-time intrusion detection in large- scale IoT networks remains an important area for future investigation.

VII. CONCLUSION

This paper presented a comprehensive literature review of intrusion detection systems in IoT environments, with a particular focus on Federated Learning-based approaches. While Federated Learning offers significant advantages in terms of privacy preservation and decentralized model training, existing systems often assume that all partici- pating clients are trustworthy, which is not practical in real-world scenarios.

To address this limitation, a trust-aware Federated Learning-based intrusion detection framework was pro- posed. The proposed system incorporates a trust evalua- tion mechanism within the aggregation process to mitigate the impact of malicious or compromised clients. By assign- ing weights to client updates based on their reliability, the framework enhances robustness, improves model stability, and strengthens resistance against adversarial attacks.

The analytical discussion indicates that integrating trust-aware mechanisms into Federated Learning can sig- nificantly improve the reliability and security of intrusion detection systems in distributed IoT environments. The proposed approach provides a promising foundation for developing scalable, privacy-preserving, and resilient cy- bersecurity solutions.

REFERENCES

1. H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.
2. P. Kairouz, H. B. McMahan, B. Avent *et al.*, "Advances and open problems in federated learning," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1–44, 2021.
3. D. C. Nguyen, M. Ding, P. N. Pathirana *et al.*, "Federated learning for internet of things: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 8, no. 21, pp.

16 224–16 247, 2021.

4. T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.
5. N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, 2018.
6. K. Alrawashdeh and C. Purdy, "Toward an online intrusion detection system using deep learning," in *IEEE International Conference on Machine Learning and Applications*, 2016, pp. 195–200.
7. N. Albanbay, Y. Tursynbek, K. Graffi, R. Uskenbayeva, Z. Kalpeyeva, Z. Abilkaiyr, and Y. Ayapov, "Federated learning- based intrusion detection in iot networks: Performance evalua- tion and data scaling study," *IEEE Access*, vol. 11, pp. 1–15, 2023.
8. J. Zhang, C. Xie, L. Chen *et al.*, "Poisoning attacks and defenses in federated learning: A survey," *IEEE Access*, vol. 9, pp. 123 456–123 478, 2021.
9. N. Moustafa and J. Slay, "Unsw-nb15: A comprehensive data set for network intrusion detection systems," in *Military Commu- nications and Information Systems Conference (MilCIS)*, 2015, pp. 1–6.
10. M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Federated deep learning for cyber security in the internet of things: Concepts, applications, and experimental analysis," *IEEE Access*, vol. 9, pp. 138 537–138 561, 2021.
11. E. M. Campos *et al.*, "Evaluating federated learning for intru- sion detection in internet of things: Review and challenges," *Computer Networks*, vol. 203, p. 108661, 2022.
12. M. A. Ferrag *et al.*, "Edge-iiotset: A new comprehensive realistic cyber security dataset of iot and iiot applications," *IEEE Access*, vol. 10, pp. 40 281–40 306, 2022.
13. T. D. Nguyen *et al.*, "Dïot: A federated self-learning anomaly detection system for iot," in *IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2019, pp. 756–767.
14. A. N. Bhagoji, S. Chakraborty, P. Mittal, and S. Calo, "Ana- lyzing federated learning through an adversarial lens," in *Inter- national Conference on Machine Learning (ICML)*, 2019, pp. 634–643.