Secure FedIDS - Privacy preserving IDS with ensemble deep learning approach

Poonguzhali V, *Srisakthi* Saravanan*

School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Chennai, India

Abstract. People are increasingly reliant on the internet for communication and various devices in their daily lives. The Internet of Things (IoT) has dramatically changed many industries by allowing more automation and easy data sharing, but it also has a high security risk because it exposes many entry points for hackers. An Intrusion Detection System (IDS) is therefore necessary to signal the occurrence of threats in this type of environment. With the advancements in machine learning and deep learning frameworks, these areas have attracted considerable attention in the field of network security. The current study introduces SecureFedIDS, a new approach to network security which utilizes a hybrid ensemble of CNN and LSTM. To solve the data privacy problem, SecureFedIDS implements Federated Learning through the Flower framework. Experimental results reveal that the methods are highly effective in terms of detection rates and precision, reaching 99.4% for binary classification and 97% for multiclass classification with a minimal number of false alarms.

1 Introduction

The Internet of Things (IoT) is a key factor in such a change, as it links an enormous variety of devices and machines, thus enabling improved automation, remote monitoring, and effortless data sharing. Nevertheless, the quick growth of these interlinked units generates a huge attack surface, thereby causing substantial security problems and the possibility of hostile attacks. Typically, such environments are safeguarded through active and passive measures that include network firewalls and Intrusion Detection Systems (IDS). With the advancement of machine learning (ML) and deep learning (DL) models, they have been extensively explored as a potential solution for next-generation intrusion detection systems [2].

*Corresponding author: srisakthi.saravanan@vit.ac.in

The biggest problem "Data Privacy" emerges in a way that ML and DL features require network data from intrusion-detected systems for their training, and hence if this data is shared, it may cause legal and operational violations. The sharing of network traffic from a hospital or a research center would be against the operating protocols and strict data privacy acts such as DPDPA in India, GDPR in the European Union, and HIPAA in the United States. To address these challenges, federated learning-based approaches have been proposed [1, 8,9]. In order to tackle these issues, SecureFedIDS is introduced by this study — a privacy-preserving IDS that is created by an ensemble of CNN and LSTM. Importantly, the system adopts a Federated Learning strategy through the Flower framework, permitting that IDS models are trained jointly over geographically spread IoT nodes without data transfer.

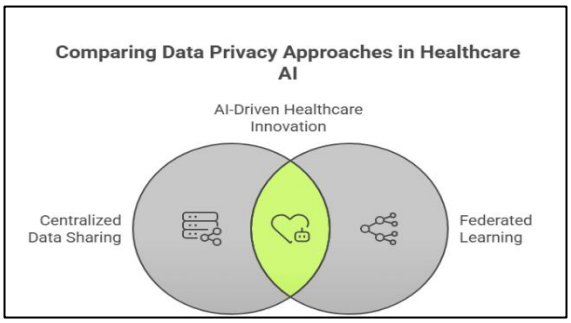

Fig. 1. Data Privacy Conflict.

The main contributions of this paper are as follows. We propose SecureFedIDS, a unified ensemble deep learning framework combining CNN and LSTM that handles both binary and multiclass intrusion detection within a single architecture — unlike existing approaches that address only one classification type. We integrate federated learning using the Flower framework, enabling distributed IDS nodes to train collaboratively without sharing raw network data, thereby preserving privacy. We evaluate the framework on two IoT datasets — IoT-DH and CIC IoT 2023 — achieving 99.4% accuracy for binary classification and 97.2% for multiclass classification.

2 Related works

The significant factor of IDS evolution has been the shift from straightforward rule-based systems to complex Machine Learning (ML) and Deep Learning (DL) structures. Present studies mainly revolve around the enhancement of detection accuracies for elaborate attack scenarios, as well as solving data privacy and class imbalance issues in IoT environments.

2.1 Conventional machine learning and ensemble approaches

Maseer et al. [3] also performed benchmarking on the CICIDS2017 dataset to shed light on the necessity of handling class imbalance and feature relevance for anomaly-based systems. Khan and Kim [11] created a highly effective IDS using Conv-AE that utilized heterogeneous datasets, on the other hand, Thakkar and Lohiya [12] presented a comprehensive study on the progression of intrusion detection datasets, emphasizing the transition from traditional datasets to the latest IoT-specific ones.

2.2 Deep learning: CNN and LSTM architectures

Li et al. [6] presented a second line of defence for network intrusion in Industrial IoT using multi-CNN feature fusion. They showed that detection accuracy could be improved by comprehensive spatial feature extraction. Yet, single CNN-based methods miss out on temporal sequence modelling abilities. To overcome this, Zhang et al. [7] presented a deep learning-based intrusion detection system using multi-scale feature fusion for IoT networks, combining spatial and temporal features. Still, these models do not include full temporal sequence modelling, which limits their ability to understand long-term dependencies in network traffic. To detect IoT security attacks, Mothukuri et al. [10] came up with a federated-learning-based anomaly detection method. They proved that FL is an excellent approach to preserving privacy and at the same time, it's highly effective in detection tasks.

Recent developments in federated learning have opened up a new avenue for privacy-preserving intrusion detection. De Oliveira et al. [8] presented F-NIDS, a network intrusion detection system that uses federated learning concept. Similarly, Zhao et al. [9] came up with a semi-supervised federated learning technique for the detection of intrusions in IoT devices. Besides, Lyu et al. [13] focused on the problem of noisy labels in federated learning which is quite a big issue in case of IDS where labeling errors occur frequently.

3 Methodology

The methodology of SecureFedIDS is focused on two very important aspects of machine learning based Intrusion Detection System: performance improvement through Ensemble Deep Learning, and privacy preservation by Federated Learning.

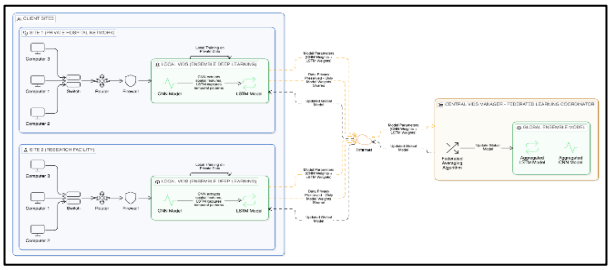

Fig. 2. SecureFedIDS system architecture.

3.1 Ensemble deep learning

The central part of SecureFedIDS is a hybrid ensemble system structurally aimed at identifying both spatial and temporal features of network traffic.

3.1.1 Datasets and feature selection

Two major datasets were used to train and test the model. The IoT-DH Dataset (2024) [4] represents a DDoS attack scenario via IoT honeypots, used for binary classification with 11 features selected for training. The CIC IoT 2023 Dataset [5] is the benchmark of real-time large-scale IoT attacks, used for multiclass classification with 38 features to distinguish between eight different attack types.

3.1.2 Data preprocessing

Standardization is essential for deep learning convergence. Rows with infinity values are dropped rather than replaced, as infinite values cannot be meaningfully normalized and would distort model training. Missing numerical fields are filled with zero, since in network traffic data, absent feature values typically indicate the absence of that network activity rather than unknown data. Feature values are scaled to a range between 0 and 1 using StandardScaler, as neural networks converge faster and more stably when input features share a common scale. Categorical labels are encoded to numerical format to make them compatible with the softmax output layer.

Class imbalance is addressed using oversampling, specifically to prevent the model from being biased toward majority classes and overlooking minority attack categories.

3.1.3 Ensemble model architecture

The presented progression leads from a simple baseline comparison to a complex ensemble methodology. CNN is selected for its ability to extract local spatial patterns from network traffic features through convolutional filters, while LSTM is chosen for its strength in capturing temporal dependencies and sequential patterns in network flow data. The ensemble of both architectures is motivated by the complementary nature of their capabilities — CNN alone lacks temporal modeling, and LSTM alone lacks spatial feature extraction. Together, they provide a more robust and comprehensive intrusion detection capability than either model individually.

The architecture consists of: (i) a 1D-Convolutional layer for spatial feature extraction and dimensionality reduction; (ii) a global average pooling layer; (iii) a single-layer unidirectional LSTM for temporal relationship modeling; and (iv) a fully connected layer with SoftMax for classification. The model operates with a batch size of 1024/2048, a learning rate of 0.0004, and the Adam optimizer for 10 to 30 epochs.

3.1.4 Mathematical formulation

The CNN convolutional operation extracts spatial features from input network traffic X using learned filters W and bias b:

$$h = ReLU(W * X + b) \quad (1)$$

where * denotes the convolution operation and ReLU is the activation function. The LSTM unit processes the CNN output sequentially. The forget gate f_t, input gate i_t, and output gate o_t are computed as:

$$f_t = \sigma(W_n \cdot [h_{t-1}, x_t] + b_n) \quad (2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4)$$

The cell state C_t and hidden state h_t are updated as:

$$C_t = f_t \odot C_{t-1} + i_t \odot tanh(W\_c \cdot [h_{t-1}, x_t] + b\_c) \quad (5)$$

$$h_t = o_t \odot tanh(C_t) \quad (6)$$

The final classification is performed using the Softmax function:

$$P(y = k \mid x) = e^{z_k} / \Sigma\, e^{z_j} \quad (7)$$

For federated learning, the global model is updated using Federated Averaging across N clients:

$$w\_global = \Sigma\ (n_k / n) \cdot w_k \quad (8)$$

where w_k are the local model weights of client k, n_k is the number of local samples, and n is the total number of samples across all clients.

3.2 Federated learning mechanism

SecureFedIDS implements privacy preservation with the help of the Flower framework, which is an open-source tool for decentralized machine learning. The Central IDS Manager (Server) manages the global model parameters and coordinates updates. Local IDS Systems (Clients) at different sites train models locally on their own private data. The interaction between the server and clients is limited to model parameters (weights) only — no raw network traffic is ever shared.

The federated averaging strategy used in SecureFedIDS slightly modifies the popular FL models for edge computing environments [2]. To make the framework work smoothly despite different data types and incorrectly labelled instances - an inevitable situation for distributed IDS systems, we decide to employ local self-regularization of the model [13]. Given that on-the-fly unlabelled normal network traffic data is very different from previously labelled attack data; our semi-supervised IDS using MLCNs can generally identify the two, which is one of the main reasons for the improved accuracy of the proposed model while at the same time privacy is ensured [14].

Table 1. Federated learning flow protocol.

Central IDS Manager (Server)	**Client IDS (Site 1 / Site 2)**
Load and extract global model attributes	Receive global model attributes and sync with local model
Propagate global attributes to clients	Load new local intrusion data for training
Wait for updates from clients	Perform local training on private data
Receive model parameters from clients	Extract parameters and send to Central Manager
Aggregate global model using Federated Averaging	(Wait for updated global model)

4 Experimental results and analysis

The evaluation of the SecureFedIDS framework was conducted on an Intel Core Ultra 5 processor (4.3 GHz, 12 Core) with 16 GB of RAM, running Python 3.11, TensorFlow 2.8, and Flower framework for federated learning orchestration.

4.1 Performance on IoT-DH dataset (binary classification)

As a binary classifier, the ensemble CNN-LSTM model was challenged to identify normal network traffic as well as spot malicious DDoS attacks. Both the training and test accuracy demonstrated quick convergence and the highest value of 99.4% was attained.

Table 2. Ensemble performance on IoT-DH dataset.

Model	Accuracy	Label	Precision	Recall	F1-Score
CNN-LSTM	99.4%	0 (Normal)	1.0000	0.9881	0.9940
CNN-LSTM		1 (Attack)	0.9882	1.0000	0.9940

The confusion matrix for the IoT-DH dataset indicates a False Positive rate of only 0.60% and a False Negative rate of 0.00%, proving that the model successfully captured all trained intrusions.

The binary classification accuracy of 99.4% achieved here goes beyond that of the recent federated learning-based IDS methods in the literature [2, 8] and is also able to keep very close performance in multiclass scenario. In addition, the extremely low false positive rate of 0.60% reveals the system's suitability for deployment in the real world and thus the previous concerns raised by IDS dataset research [12] are effectively sorted out.

4.2 Performance on CIC IoT 2023 dataset (multiclass classification)

For complex scenarios involving eight different attack types, the ensemble model achieved a high overall accuracy of 97.23%.

Table 3. Ensemble performance on CIC IoT 2023 dataset.

Attack Type	Precision	Recall	F1-Score
BENIGN	0.9999	1.0000	1.0000
DDOS-ICMP_FLOOD	0.9986	0.9993	0.9989
DDOS-UDP_FLOOD	0.9697	0.9986	0.9839
DDOS-TCP_FLOOD	0.9995	0.9996	0.9995
DDOS-PSHACK_FLOOD	0.9998	0.9984	0.9991
DDOS-RSTFINFLOOD	0.9959	0.9994	0.9976
DDOS-SYNONYMOUSIP_FLOOD	0.5200	0.9103	0.6619
DDOS-SYN_FLOOD	0.7491	0.2443	0.3685
Overall Accuracy	97.23%		

The system showed remarkable performance for the majority of flood categories. The DDOS-SYN_FLOOD category showed lower F1-scores (Precision: 0.7491, Recall: 0.2443, F1: 0.3685), marking an area for future optimization. This underperformance can be attributed to two key factors. First, SYN flood attacks share significant feature overlap with other TCP-based flood attacks such as DDOS-TCP_FLOOD, making it difficult for the model to distinguish between them based on packet-level features alone. Second, despite oversampling being applied to address class imbalance, the SYN flood traffic patterns may still be insufficiently diverse in the training data, causing the model to miss a large proportion of true SYN flood instances as reflected in the low recall score. Future work should explore

additional distinguishing features specific to SYN floods, such as TCP handshake incompletion rates, to improve classification of this attack category.

4.3 Federated learning efficiency and privacy

Collaborative training across decentralized nodes was successfully enabled by the federated approach via the Flower framework without the need for moving raw data.
Data Privacy: There were no network traffic datasets transferred between the Central IDS manager and local site clients. Aggregation: The Central Manager upgraded the global model with the Federated Average of the attributes received from local clients. Compliance: The staged parameter sharing upheld privacy regulations such as GDPR and DPDPA.

Privacy Analysis: The primary privacy guarantee in SecureFedIDS is achieved through federated learning itself — only trained model parameters (gradients) are shared between clients and the central server, never raw network traffic data. While the current implementation does not incorporate formal differential privacy mechanisms such as noise injection or privacy budget (ε) bounds, the federated parameter-sharing approach significantly reduces the risk of data exposure compared to centralized training. Incorporating differential privacy guarantees such as DP-SGD remains an important direction for future work to provide mathematically provable privacy bounds.

Communication Overhead and Scalability: In the current implementation, federated learning was tested with two client nodes exchanging model parameters with the central server across multiple rounds. The model size transferred per round consists only of the CNN-LSTM weight parameters, which is significantly smaller than transferring the full training datasets. The CIC IoT 2023 dataset contains over 150,000 samples — transferring raw data of this scale would be impractical and privacy-violating, whereas only model weights are exchanged in SecureFedIDS. Training time per federated round averaged approximately 2-5 minutes per client. Scalability to larger numbers of clients is supported by the Flower framework's federated averaging strategy. A full scalability analysis across larger client counts remains an important direction for future work.

5 Summary and future works

In this paper, we proposed and implemented SecureFedIDS, a novel privacy-preserving intrusion detection system developed as part of this research. Our system combines an ensemble of CNN and LSTM to detect network intrusions in IoT environments, handling both binary classification (normal vs. attack) and multiclass classification (eight specific DDoS attack types) within a single unified architecture. To address the critical challenge of data privacy, we implemented federated learning using the Flower framework, enabling multiple distributed IDS nodes to collaboratively train the global model by sharing only model parameters — never raw network data. Our experimental evaluation demonstrated 99.4% accuracy for binary classification and 97.2% for multiclass classification, confirming that SecureFedIDS successfully balances high detection performance with strict data privacy compliance, making it a practical solution for real-world IoT security deployments.

Future work will focus on: (i) Model Optimization — system refinement to detect more challenging attack types such as SYN floods; (ii) Cloud Integration — improving model aggregation and deployment for extensive cloud environments; (iii) Deployment Complexity: Deploying and maintaining a federated ensemble IDS in real operational environments introduces practical challenges including orchestrating model synchronization across geographically distributed nodes, handling client dropouts or unreliable network connections during federated rounds, and ensuring local client hardware meets minimum computational requirements for running CNN-LSTM training locally. Real-world deployment would require careful orchestration tools, monitoring dashboards, and fault-tolerance mechanisms to handle these complexities at scale. (iv) Operational Monitoring:

Future work will explore automated monitoring pipelines that can detect model drift over time and trigger retraining rounds across federated clients automatically when new attack types emerge.

References

1. O. Aouedi, K. Piamrat, G. Muller, K. Singh, Federated semisupervised learning for attack detection in industrial internet of things, IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 286-295, (2022). doi: 10.1109/TII.2022.3149902
2. M. H. Bhavsar, Y. B. Bekele, K. Roy, J. C. Kelly and D. Limbrick, FL-IDS: Federated Learning-Based Intrusion Detection System Using Edge Devices for Transportation IoT, IEEE Access, vol. 12, pp. 52215-52226, (2024).
3. Z. K. Maseer, R. Yusof, N. Bahaman, S. A. Mostafa and C. F. M. Foozy, Benchmarking of Machine Learning for Anomaly Based IDS in the CICIDS2017 Dataset, IEEE Access, vol. 9, pp. 22351-22370, (2021).
4. S. Saif, W. Widyawan, R. Ferdiana, IoT-DH dataset for classification, identification, and detection DDoS attack in IoT, Data in Brief, vol. 54, 110496, (2024).
5. E. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu and A. Ghorbani, CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment, (2023). doi: 10.20944/preprints202305.0443.v1.
6. Y. Li, Y. Xu, Z. Liu, H. Hou, Y. Zheng, Y. Xin, Y. Zhao and L. Cui, Robust Detection for Network Intrusion of Industrial IoT Based on Multi-CNN Feature Fusion, IEEE Transactions on Industrial Informatics, vol. 18, no. 5, pp. 3424-3434, (2022). doi: 10.1109/TII.2021.3106012
7. H. Zhang, J. Yu, C. Tian, F. Xu, Y. Li, S. Yang and Y. Liu, Deep Learning-Based Intrusion Detection for IoT Networks: A Multi-Scale Feature Fusion Approach, Sensors, vol. 21, no. 3, p. 998, (2021). doi: 10.3390/s21030998
8. J. A. de Oliveira, V. P. Gonçalves, R. I. Meneguette, R. T. de Sousa Jr, D. L. Guidoni, J. C. Oliveira, G. P. Rocha Filho, F-NIDS—A Network Intrusion Detection System based on federated learning, Computer Networks, vol. 236, 110010, (2023). doi: 10.1016/j.comnet.2023.110010
9. R. Zhao, Y. Wang, Z. Xue, T. Ohtsuki, B. Adebisi, G. Gui, Semi-Supervised Federated Learning Based Intrusion Detection Method for Internet of Things, IEEE Internet of Things Journal, vol. 10, no. 10, pp. 8645-8657, (2023). doi: 10.1109/JIOT.2023.3240331
10. V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriyeh, A. Dehghantanha, G. Srivastava, Federated-learning-based anomaly detection for IoT security attacks, IEEE Internet of Things Journal, vol. 9, no. 4, pp. 2545-2554, (2022). doi: 10.1109/JIOT.2021.3077803
11. M. A. Khan and J. Kim, Toward Developing Efficient Conv-AE-Based Intrusion Detection System Using Heterogeneous Dataset, Electronics, vol. 9, no. 11, p. 1771, (2020).
12. A. Thakkar and R. Lohiya, A Review of the Advancement in Intrusion Detection Datasets, Procedia Comput. Sci., vol. 167, pp. 636-645, (2020).
13. C. Lyu, K. Shine, P. Jacquet and W. Dabbous, Towards Federated Learning Against Noisy Labels via Local Self-Regularization, IEEE Trans. Neural Netw. Learn. Syst., vol. 35, no. 3, pp. 3324-3337, (2024).
14. R. Zhao, Y. Wang, Z. Xue, T. Ohtsuki, B. Adebisi and G. Gui, Semi-Supervised Federated Learning Based Intrusion Detection Method for Internet of Things, IEEE Internet Things J., vol. 10, no. 10, pp. 8645-8657, (2023).