International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

Federated Intrusion Detection Framework for Heterogeneous Network Traffic Using Multi-Client Deep Learning Aggregation

Pranav Karthikeyan Aiyyer[1], Dr. Manjula Devi R[2], Sagar Chaudhary[3], Sownesh S[4]
[1]UG – Department of Computer Science and Business Systems (CSBS), KPR Institute of Engineering and Technology, Coimbatore, Tamil Nadu
[2]Professor, Department of Computer Science and Engineering (CSE), KPR Institute of Engineering and Technology, Coimbatore, Tamil Nadu
[3]UG – Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), KPR Institute of Engineering and Technology, Coimbatore, Tamil Nadu
[4]UG – Department of Electronics and Communication Engineering (ECE), KPR Institute of Engineering and Technology, Coimbatore, Tamil Nadu
***Emails**: pranavaiyyer2311@gmail.com[1], manjuladevi.r@kpriet.ac.in[2], sagarchaudharynn06@gmail.com[3] Sownesh.2004@gmail.com [4]*

Abstract
The rapid expansion of cloud computing, Internet of Things (IoT) ecosystems, and large-scale enterprise networks has intensified the complexity and frequency of cyber attacks. Traditional centralized Intrusion Detection Systems (IDS) require aggregation of raw network traffic from multiple distributed sources, raising significant concerns related to data privacy, regulatory compliance, communication overhead, and scalability. These challenges limit effective collaborative security across heterogeneous network environments. This paper presents a federated intrusion detection framework that enables decentralized collaborative learning without sharing raw network traffic data. A lightweight Multilayer Perceptron (MLP) model is trained across three independent federated clients using the Federated Averaging (FedAvg) algorithm. Each client is assigned a distinct real-world benchmark dataset — CICIDS2017, UNSW-NB15, and NSL-KDD — introducing authentic non-IID data heterogeneity in terms of dataset size, feature dimensionality, class imbalance, and attack taxonomy. The framework ensures strict data locality by transmitting only model parameters during aggregation. Experimental results demonstrate that the global federated model achieves 87% accuracy with balanced precision (0.86), recall (0.86), and F1-score (0.85), representing a 20 percentage-point improvement over the weakest local model. Convergence analysis confirms stable performance under non-IID conditions, while communication overhead is reduced by approximately 1000× compared to centralized training.

Keywords: *Distributed Learning Federated Learning, FedAvg, , Intrusion Detection System, Network Security, Non-IID Data Privacy Preservation..*

1. Introduction

1.1.Background and Motivation

The modern network security landscape is characterized by an ever-expanding attack surface, driven by the exponential growth of connected devices, cloud-native applications, and distributed enterprise architectures. Global intrusion attempt volumes have grown substantially in recent years, with adversaries employing Advanced Persistent Threat (APT) techniques, living-off-the-land (LotL) tactics, and zero-day exploit chains that evade traditional signature-based detection [5]. The financial consequences are severe: a single enterprise data breach can result in millions of dollars in remediation costs, regulatory fines, and reputational damage. Intrusion Detection Systems (IDS) serve as a critical defense layer, continuously monitoring network traffic to identify malicious activity. IDS solutions are broadly categorized into signature-based detection — which matches traffic against known attack signatures — and anomaly-based detection — which learns a statistical model of normal behavior and flags deviations. While signature-based systems offer high precision for known attacks, they cannot detect novel variants.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

Anomaly based systems leveraging deep learning have demonstrated superior generalization, learning complex nonlinear decision boundaries from high-dimensional network flow features without requiring explicit attack signatures. However, the effectiveness of deep learning-based IDS depends critically on large, diverse training datasets. No single organization possesses data capturing the full diversity of mod ern attack behaviors across different network environments, application stacks, and geographic regions. This has motivated collaborative learning approaches, where multiple organizations pool traffic data to train a shared model. The dominant implementation is centralized: raw data from multiple sources is aggregated at a single server for joint training. Can the statistical benefits of collaborative learning be achieved without the privacy costs of centralization? This paper provides an affirmative empirical answer through federated learning

1.2. Limitations Of Centralized IDS

Despite its appeal, centralized collaborative IDS training is undermined by four fundamental limitations: (L1) Data Privacy and Regulatory Compliance. Raw network traffic encodes sensitive user behavioral patterns, application-layer payloads, authentication credentials, and proprietary business communications. Centralizing such data across organizational boundaries may directly violate GDPR, 2 HIPAA, and national cybersecurity frameworks [4]. Organizations in regulated industries — healthcare, finance, government, critical infrastructure — are legally constrained from sharing raw traffic data with external parties. (L2) Communication Overhead. Large enterprise net works generate traffic at tens to hundreds of gigabytes per hour. Transmitting raw traffic to a central server imposes prohibitive bandwidth costs and introduces latency incompatible with real-time detection requirements. As the number of participating organizations scales, the communication bottleneck at the central server becomes increasingly severe. (L3) Single Point of Failure. Concentrating sensitive traffic data and model intelligence at a single server creates a high-value adversarial target. A successful attack — through data poisoning, model extraction, or direct infrastructure compromise — can simultaneously disable intrusion detection across all connected organizations. (L4) Data Heterogeneity. Different organizations operate distinct network topologies, application stacks, and user populations, and face different threat actor profiles. A centrally trained model may fail to capture organization-specific traffic characteristics most discriminative for intrusion detection in any individual deployment context.

1.3.Federated Learning As A Solution

Federated Learning (FL), introduced by McMahan et al. [1], enables collaborative model training without raw data sharing. Each client trains a local model on its private dataset and transmits only model parameters to a central aggregation server, which combines them into an updated global model using a weighted aggregation algorithm. This iterative process continues for a fixed number of communication rounds until convergence. FL directly addresses all four limitations: (i) raw data never leaves the client, satisfying privacy regulations; (ii) only compact parameter vectors are transmitted, dramatically reducing communication overhead; (iii) eliminating a central data repository removes the primary adversarial target; and (iv) local training on organization-specific data incorporates domain-specific knowledge into the global model. However, FL for IDS introduces a critical challenge: the non-IID nature of network traffic across clients. Standard FL convergence guarantees assume approximately identical client distributions. In practice, traffic distributions differ substantially in feature statistics, class proportions, and attack taxonomy. This distributional divergence causes locally computed gradients to point in conflicting directions, causing the global model to converge slowly or to a suboptimal solution — a phenomenon known as client drift [4]. D. Research Gap and Contributions Most existing FL-based IDS studies evaluate their frame works by artificially partitioning a single homogeneous dataset into client subsets [6], [7]. This fails to capture the true distributional divergence of real-world federated deployments, where clients observe fundamentally different traffic patterns,

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

feature distributions, and attack taxonomies. This paper ad dresses this gap using three independently sourced real-world datasets as federated clients, introducing natural non-IID heterogeneity. The primary contributions are: 1) Heterogeneous Federated IDS Framework: Evaluated across three independently sourced datasets (CI CIDS2017, UNSW-NB15, NSL-KDD), each assigned to an independent client, introducing natural non-IID heterogeneity that authentically reflects real-world deployment conditions. 2) Per-Client Performance Analysis: Rigorous empirical analysis of local and global model performance under natural non-IID conditions, including per-client accuracy, precision, recall, and F1-score with mathematical justification. 3) Convergence and Communication Analysis: Quantitative analysis of global model convergence over communication rounds and communication overhead reduction relative to centralized baselines. 4) Limitations and Future Directions: Critical discussion of Byzantine robustness, adaptive aggregation, personalized FL, and differential privacy as directions for future work. The remainder of this paper is organized as follows. Section II reviews related work. Section III describes the baseline experiment. Section IV presents the proposed framework. Section V details the experimental setup. Section VI presents and analyzes results. Section VII concludes the paper. credentials, and proprietary business

2. Related Work

The intersection of federated learning and network intrusion detection has attracted substantial research attention. This section reviews the state of the art across three dimensions: model architectures, privacy mechanisms, and non-IID handling strategies.

2.1. Deep Learning Architectures In Federated IDS

The choice of local model architecture in a federated IDS involves a fundamental trade-off between detection accuracy, computational efficiency, and communication overhead per round. Ferrari et al. [1] proposed DeepFed, combining CNN and LSTM for industrial cyber-physical system intrusion detection within a Secure FedAvg scheme. DeepFed demonstrated high accuracy on the SWaT and Gas Pipeline datasets. However, the LSTM component has time complexity O(T •H2) where T is sequence length and H is hidden dimension, making it unsuitable for resource-constrained IoT or edge clients. Furthermore, its evaluation on industrial CPS datasets does not generalize to heterogeneous enterprise network environments. Al-Hawawreh et al. [2] proposed a hybrid VAE-CNN architecture for IoT anomaly detection within a federated frame work. The VAE component enables unsupervised latent space learning, detecting novel attacks via reconstruction error. How ever, the encoder-decoder architecture substantially increases trainable parameters, amplifying communication overhead per round. 3 Kumar and Gupta [3] explored ResNet-based architectures with residual skip connections h(l+1) = F(h(l),W(l)) + h(l) for federated IDS. While ResNet achieved competitive accuracy on CICIDS2017, the depth of standard ResNet configurations (18–50 layers) imposes significant memory and compute requirements, making it unsuitable for edge deployments. In contrast, the lightweight MLP adopted in the proposed framework has parameter count O(d • H + H2 + H) for a two-hidden-layer network with input dimension d and hidden dimension H, orders of magnitude smaller than CNN, LSTM, or ResNet architectures. Wang et al. [9] proposed Sentinel, using knowledge distillation to transfer knowledge from a global teacher to lightweight local student models, improving per-client performance but requiring a shared public dataset at the server. Zhou et al. [10] proposed FetFIDS with feature embedding attention mechanisms to dynamically weight feature contributions, but the attention mechanism adds parameters and its effectiveness under severe distributional shift remains limited.

2.2. Privacy-Enhancing Mechanisms

Mothukuri et al. [4] surveyed three primary privacy mechanisms for FL-based IDS. Differential Privacy (DP) adds calibrated Gaussian or Laplacian noise to model updates, satisfying (ϵ, δ)-DP

$$\Pr[\mathcal{M}(\mathcal{D}) \in S] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(\mathcal{D}') \in S] + \delta \qquad (1)$$

for any two adjacent datasets D, D′ differing in one

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

record. Smaller ϵ provides stronger privacy but requires larger noise, degrading model accuracy by 5–15% at meaningful budgets (ϵ < 1) on imbalanced IDS datasets. Homomorphic Encryption (HE) enables aggregation in the encrypted domain with out decrypting individual updates, providing cryptographic guarantees but introducing 100× ciphertext expansion. Secure Multi-Party Computation (SMPC) distributes the aggregation computation across multiple non-colluding servers, providing information-theoretic security but requiring complex coordination protocols. The proposed framework adopts data locality as its primary privacy mechanism: raw traffic data never leaves the client environment, providing strong practical privacy without the computational overhead of DP, HE, or SMPC.

2.3. Non-IID Data Heterogeneity

Data heterogeneity is universally recognized as the most significant technical challenge in federated IDS [5]. The gradient divergence quantifies this divergence:

$$\Gamma = F^* - \sum_{k=1}^{K} \frac{n_k}{N} F_k^* \qquad (2)$$

where F∗ is the global optimum and F∗ k is client k's local optimum. When Γ > 0, local optima diverge from the global optimum, causing client drift. Zhao et al. [6] proposed FLDID with weighted FedAvg assigning higher weights to clients with balanced class distributions. Rahman et al. [7] proposed FedMSE with MSE weighted aggregation based on local validation performance. Albanbay et al. [8] applied FedProx, adding a proximal regularization term to constrain local updates

$$h_k(\mathbf{w}; \mathbf{w}^{(t)}) = F_k(\mathbf{w}) + \frac{\mu}{2} \left\| \mathbf{w} - \mathbf{w}^{(t)} \right\|^2 \qquad (3)$$

where μ > 0 is the proximal coefficient. FedProx reduces client drift at the cost of an additional hyperparameter requiring careful tuning. A critical observation is that no existing FL-based IDS study evaluates its framework using multiple independently sourced real-world datasets as federated clients. The proposed framework directly addresses this gap.

3. Initial Baseline Approach

3.1. Baseline Experimental Design

Prior to the full heterogeneous framework, a baseline experiment was conducted to: (i) validate the correctness of the federated training pipeline, and (ii) establish a performance reference under idealized homogeneous data conditions. A single dataset D with N labeled samples is partitioned into K disjoint subsets: using random stratified sampling to preserve class distribution within each subset, ensuring approximately IID conditions: Each subset is assigned to a simulated federated client, which trains a local MLP for E epochs per communication round using mini-batch SGD, then transmits updated parameters to the server for FedAvg aggregation.

3.2. Baseline Results and Limitations

The baseline demonstrated stable convergence, with the global model achieving accuracy comparable to a centralized model trained on the full dataset 'D. The training loss exhibited a smooth, monotonically decreasing trajectory, confirming the correctness of the FedAvg implementation. However, the baseline has a fundamental limitation: under IID conditions, locally computed gradients are unbiased estimates of the global gradient, and FedAvg is theoretically equivalent to centralized SGD with a larger effective batch size. The convergence behavior observed in the baseline therefore reflects the best-case scenario for FedAvg and does not provide meaningful evidence of the framework's ability to generalize under real-world non-IID conditions. Furthermore, the baseline does not capture feature space heterogeneity: different organizations may use different monitoring tools and capture different feature sets, resulting in client datasets with different dimensionalities. These limitations directly motivate the proposed heterogeneous framework. Shown in Table I

4. Initial Baseline Approach

4.1. Baseline Experimental Design

Prior to the full heterogeneous framework, a baseline experiment was conducted to: (i) validate the correctness of the federated training pipeline, and (ii) establish a performance reference under idealized homogeneous data conditions.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

Table 1 Method

Method	Aggregation	Non-IID	Dataset(s)
DeepFed [1]	Secure FedAvg	No	SWaT; Gas Pipeline
FLDID [6]	Weighted FedAvg	Partial	X-IIoTID
FedMSE [7]	MSE-Weighted	Artificial	N-BaIoT
FedProx [8]	Proximal FedAvg	Artificial	IoT datasets
Sentinel [9]	Distillation	Partial	CICIDS2017
FetFIDS [10]	Attention FedAvg	No	UNSW-NB15
Proposed	**FedAvg**	**Natural**	**CIC; UNSW; NSL**

A single dataset D with N labeled samples is partitioned into K disjoint subsets:

$$\mathcal{D}_k \cap \mathcal{D}_j = \emptyset \ \ \forall k \neq j, \qquad \bigcup_{k=1}^{K} \mathcal{D}_k = \mathcal{D} \tag{4}$$

using **random stratified sampling** to preserve class distribution within each subset, ensuring approximately IID conditions:

$$P_1(\mathbf{x}, y) \approx P_2(\mathbf{x}, y) \approx \cdots \approx P_K(\mathbf{x}, y) \approx P(\mathbf{x}, y) \tag{5}$$

Each subset is assigned to a simulated federated client, which trains a local MLP for E epochs per communication round using mini-batch SGD, then transmits updated parameters to the server for FedAvg aggregation.

4.2.Baseline Results and Limitations

The baseline demonstrated stable convergence, with the global model achieving accuracy comparable to a centralized model trained on the full dataset ‘D. The training loss exhibited a smooth, monotonically decreasing trajectory, confirming the correctness of the FedAvg implementation. However, the baseline has a fundamental limitation: under IID conditions, locally computed gradients are unbiased estimates of the global gradient, and FedAvg is theoretically equivalent to centralized SGD with a larger effective batch size. The convergence behavior observed in the baseline therefore reflects the best-case scenario for FedAvg and does not provide meaningful evidence of the framework’s ability to generalize under real-world non-IID conditions. Furthermore, the baseline does not capture feature space heterogeneity: different organizations may use different monitoring tools and capture different feature sets, resulting in client datasets with different dimensionalities. These limitations directly motivate the proposed heterogeneous framework.

5. Proposed Federated Intrusion Detection Framework

5.1.System Architecture

The proposed framework follows a privacy-by-design principle: raw network traffic data is processed and stored exclusively within the client environment, and only compact model parameter vectors are transmitted between clients and the central aggregation server. The system comprises K = 3 federated clients, each holding a private dataset:

$$\mathcal{D}_k = \left\{ (\mathbf{x}_i^{(k)}, y_i^{(k)}) \right\}_{i=1}^{n_k} \tag{6}$$

where n_k is the number of samples at client k, $\mathbf{x}_i^{(k)} \in \mathbb{R}^{d_k}$ is the feature vector with dimensionality d_k (which may differ across clients), and $y_i^{(k)} \in \{0, 1\}$ is the binary label (0 = benign, 1 = attack). The total number of training samples is $N = \sum_{k=1}^{K} n_k$. The overall architecture is illustrated in Fig. 1.

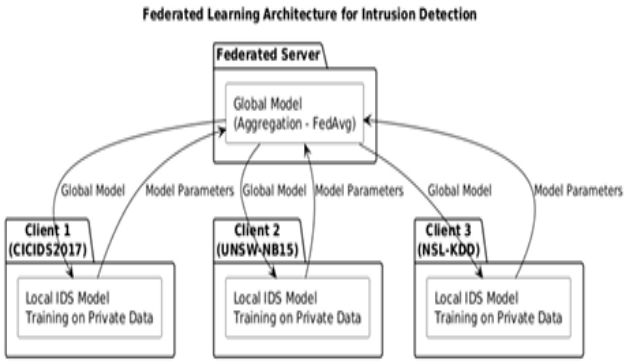

Figure 1 Federated IDS architecture. Each client trains a local MLP on its private network traffic dataset. Only model parameters are transmitted to the central server for FedAvg aggregation. Raw data never leaves the client environment.

The three clients are assigned: Client 1 ← CICIDS2017 (n1 ≈ 2.8M, d1 = 78), Client 2 ← UNSW-NB15 (n2 ≈ 2.5M, d2 = 49), Client 3 ← NSL-KDD (n3 ≈ 149K, d3 = 41). After preprocessing, all features are normalized to [0,1] and projected to a common input dimensionality d to enable parameter sharing across clients. Each client k minimizes its

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

local empirical risk:

$$F_k(\mathbf{w}) = \frac{1}{n_k} \sum_{i=1}^{n_k} \ell\Big(f(\mathbf{x}_i^{(k)}; \mathbf{w}),\, y_i^{(k)}\Big) \qquad (7)$$

here w ∈ Rp is the shared model parameter vector, f(•;w) : Rd →[0,1] is the MLP prediction function, and ℓ(•,•) is the binary cross-entropy loss:

$$\ell(\hat{y}, y) = -\left[y \log(\hat{y}) + (1-y) \log(1-\hat{y})\right] \qquad (8)$$

The global federated objective is to minimize the weighted average of all local empirical risks:

$$\mathbf{w}^* = \arg\min_{\mathbf{w} \in \mathbb{R}^p} F(\mathbf{w}), \quad F(\mathbf{w}) = \sum_{k=1}^{K} \frac{n_k}{N} F_k(\mathbf{w}) \qquad (9)$$

This weighted formulation ensures that clients with larger datasets contribute proportionally more to the global objective, reflecting their greater statistical representativeness. The communication protocol is illustrated shown in Figure 2.

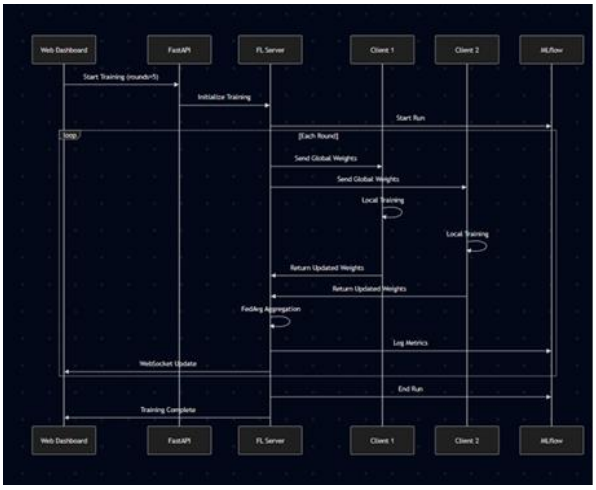

Figure 2 Federated training sequence diagram. In each communication round, the server broadcasts global parameters to all clients; clients perform local training and return updated parameters; the server aggregates via FedAvg.

5.2. Threat Model

The proposed framework operates under the following threat model assumptions, which define the security boundaries of the system. Honest-but-Curious Server. The central aggregation server is assumed to be honest-but-curious: it correctly executes the FedAvg aggregation protocol but may attempt to infer information about individual clients' private datasets from the received model updates. .The framework defends against this threat through data locality — the server never receives raw traffic data — but does not provide formal guarantees against gradient inversion attacks on model parameters. Benign Clients. All K = 3 clients are assumed to be honest participants that correctly execute the local training protocol and transmit unmodified model updates. The framework does not currently defend against Byzantine clients that deliberately submit poisoned updates to degrade the global model. This is an acknowledged limitation addressed in Section VI-F. Secure Communication Channel. Model parameter transmissions between clients and the server are assumed to occur over a secure, authenticated channel (e.g., TLS 1.3), protecting against eavesdropping and man-in-the-middle attacks on the communication layer. No Data Sharing. The fundamental privacy guarantee of the framework is that raw network traffic data never leaves the client environment under any circumstances. This guarantee holds regardless of server behavior ,as the client-side training pipeline is entirely local.

5.3. Local Model : Multilayer Perceptron

A lightweight MLP is deployed at each client as the local intrusion detection model. Why choose an MLP over more expressive architectures such as CNNs or LSTMs? In a federated context where model parameters must be transmitted over potentially bandwidth-constrained links and clients may have heterogeneous computational resources, a model achieving 87%accuracywith104 parameters is more deployable than one achieving 92% with 107 parameters. The MLP consists of an input layer of dimension d, two hidden layers with ReLU activations, and a sigmoid output layer for binary classification The forward pass is:

$$\mathbf{h}^{(1)} = \mathrm{ReLU}\Big(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}\Big), \quad \mathbf{W}^{(1)} \in \mathbb{R}^{H_1 \times d} \qquad (10)$$

$$\mathbf{h}^{(2)} = \mathrm{ReLU}\Big(\mathbf{W}^{(2)}\mathbf{h}^{(1)} + \mathbf{b}^{(2)}\Big), \quad \mathbf{W}^{(2)} \in \mathbb{R}^{H_2 \times H_1} \qquad (11)$$

$$\hat{y} = \sigma\Big(\mathbf{W}^{(3)}\mathbf{h}^{(2)} + \mathbf{b}^{(3)}\Big), \quad \sigma(z) = \frac{1}{1+e^{-z}} \qquad (12)$$

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

where H1 and H2 are the hidden layer widths. The ReLU activation ReLU(z) = max(0,z) introduces nonlinearity while avoiding the vanishing gradient problem. The complete set of trainable parameters is:

$$\mathbf{w} = \left\{ \mathbf{W}^{(1)}, \mathbf{b}^{(1)}, \mathbf{W}^{(2)}, \mathbf{b}^{(2)}, \mathbf{W}^{(3)}, \mathbf{b}^{(3)} \right\} \in \mathbb{R}^{p} \quad (13)$$

with total parameter count $p = H_1 d + H_1 + H_2 H_1 + H_2 + H_2 + 1$.

5.4.Federated Training Protocol

1) Local Training: At the beginning of round t ∈ {1,...,T}, the server broadcasts current global parameters w(t) to all K clients .Each client k initializes its local model with w(t) and performs E epochs of mini-batch SGD on its local dataset Dk. The local update rule for a mini-batch Bk ⊆ Dk is:

$$\mathbf{w}_k \leftarrow \mathbf{w}_k - \eta \cdot \frac{1}{|\mathcal{B}_k|} \sum_{(\mathbf{x}_i, y_i) \in \mathcal{B}_k} \nabla_{\mathbf{w}} \ell(f(\mathbf{x}_i; \mathbf{w}_k), y_i) \quad (14)$$

Where η>0 is the learning rate. After completing E local epochs ,client k transmits only the updated parameters w(t+1) to the server—no raw data, no gradient information ,and no local dataset statistics are shared.

2) FedAvg Aggregation: Upon receiving updated parameters 1 from all clients, the server computes the new global model as a dataset-size-weighted average:

$$\mathbf{w}^{(t+1)} = \sum_{k=1}^{K} \frac{n_k}{N} \mathbf{w}_k^{(t+1)} \quad (15)$$

Defining the **local gradient displacement** $\Delta \mathbf{w}_k^{(t)} = \mathbf{w}_k^{(t+1)} - \mathbf{w}^{(t)}$, FedAvg is equivalent to:

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + \sum_{k=1}^{K} \frac{n_k}{N} \Delta \mathbf{w}_k^{(t)} \quad (16)$$

Under IID conditions with E = 1, FedAvg reduces exactly to centralized SGD with a mini-batch size of . Under non-IID conditions with E > 1,the local displacements diverge across clients due to differing local loss landscapes, introducing client drift.

3) Convergence Criterion: The global federated loss at round t is:

$$F(\mathbf{w}^{(t)}) = \sum_{k=1}^{K} \frac{n_k}{N} F_k(\mathbf{w}^{(t)}) \quad (17)$$

Training is considered converged when the change in global loss between consecutive rounds falls below a threshold ε > 0:

$$\left| F(\mathbf{w}^{(t+1)}) - F(\mathbf{w}^{(t)}) \right| < \varepsilon \quad (18)$$

In the experiments reported here, training is conducted for a fixed T=10 communication rounds, which is sufficient for the global model to reach a stable performance plateau as demonstrated by the convergence curves in Section VI.

6. Experimental Setup

6.1.Dataset Selection Rationale

Three independently sourced benchmark datasets are assigned to three federated clients to introduce natural non IID heterogeneity. The selection criteria are: (i) independent origin—each dataset is generated by a different research institution using different traffic generation methodologies; (ii) complementary attack coverage—the datasets collectively cover both legacy and modern attack taxonomies; (iii) wide adoption in the IDS research community, enabling meaningful comparison with prior work; and (iv) public availability, ensuring reproducibility shown in Table II.

Table 2 Statistical Summary of Federated Datasets

Dataset	Records	Feats.	Imbal.	Attack Types
CICIDS2017	~2.8M	78	Severe	DoS, DDoS, Brute Force, Web, Botnet, Heartbleed
UNSW-NB15	~2.5M	49	Moderate	Fuzzers, Exploits, DoS, Backdoors, Recon, Worms
NSL-KDD	~149K	41	Mild	DoS, Probe, R2L, U2R

6.2.Dataset Descriptions

CICIDS2017 (Client 1) was generated by the Canadian Institute for Cybersecurity using CICFlowMeter over a five day capture period in a realistic enterprise testbed. It contains 78 flow-level

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

features including packet length statistics, inter arrival time (IAT) statistics, TCP flag counts, and active/idle time statistics. Attack categories include DoS (Slowloris, GoldenEye, Hulk), DDoS (HOIC), Brute Force (FTP-Patator, SSH-Patator), Web attacks (SQL Injection, XSS), Heartbleed, and Botnet (ARES C&C). The dataset is severely imbalanced (∼85% benign), making it the most challenging client for binary classification UNSW-NB15 (Client 2) was created at UNSW Canberra using the IXIA PerfectStorm tool, combining real normal traffic with nine synthetic attack families: Fuzzers, Analysis, Back doors, DoS, Exploits, Generic, Reconnaissance, Shellcode, and Worms. Its 49 features span basic connection statistics, content-level features (HTTP/FTP/SMTP command counts), time-based features, and generated behavioral indicators. The relatively balanced class distribution (∼35% attack) makes it the strongest local client. NSL-KDD (Client 3) is a refined version of KDD Cup 1999, with duplicate records removed to eliminate classifier bias toward frequent record types [12]. Its 41 features cover basic connection attributes (protocol, service, flag, bytes), content features (login attempts, root accesses, file creations), and time-window traffic statistics (same-host/service connection rates, error rates). The legacy attack taxonomy — DoS (neptune, smurf, pod), Probe (portsweep, nmap, satan), R2L (ftp write, imap, phf), and U2R (buffer overflow, rootkit, loadmodule) — provides a contrasting distribution to the modern datasets.

6.3.Preprocessing Pipeline

Each dataset is preprocessed independently within its client environment, ensuring no data is shared between clients during preprocessing shown table 3:

1) **Missing/infinite value removal**: Records with NaN or $\pm\infty$ values are dropped.
2) **Categorical encoding**: Label encoding for ordinal features; one-hot encoding for nominal features (protocol type, service, flag).
3) **Min-max normalization**: $x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$, scaling all features to $[0, 1]$.
4) **Label binarization**: All attack categories mapped to $y = 1$; benign traffic to $y = 0$.
5) **Stratified split**: 80% training / 20% test, preserving class distribution.

Table 3 Federated Training Hyper Parameters

Hyperparameter	Value
Communication rounds T	10
Local epochs per round E	5
Mini-batch size $\lvert\mathcal{B}_k\rvert$	256
Learning rate η	0.001
Optimizer	Adam
Hidden layer widths H_1, H_2	128, 64
Activation (hidden)	ReLU
Activation (output)	Sigmoid
Loss function	Binary cross-entropy
Aggregation	FedAvg (size-weighted)

Performance is evaluated using four standard binary classification metrics derived from the confusion matrix (TP, TN, FP, FN): In IDS applications, Recall is operationally critical (missed attacks are costly), while Precision controls false alarm rates that cause analyst alert fatigue. The F1-score provides a single balanced metric that accounts for both.

7. Results And Discussion

7.1.Local Model Performance

Presents the classification performance of each client's local MLP model trained in isolation, prior to any federated aggregation. These results reflect the intrinsic learn ability of each client's data distribution and serve as the baseline against which the benefit of federated collaboration is measured shown Table IV.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \qquad \text{Recall} = \frac{TP}{TP + FN} \tag{20}$$

$$\text{F1-score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{21}$$

Client 1 (CICIDS2017) achieves the lowest accuracy (0.67) due to severe class imbalance ($\pi 0 \approx 0.83$ benign). The expected loss gradient is dominated by the majority class:

$$\mathbb{E}[\nabla_{\mathbf{w}} \ell] = \pi_0 \cdot \mathbb{E}[\nabla_{\mathbf{w}} \ell \mid y = 0] + \pi_1 \cdot \mathbb{E}[\nabla_{\mathbf{w}} \ell \mid y = 1] \tag{22}$$

Client 2 (UNSW-NB15) achieves the best balanced performance (accuracy 0.96, F 10.96), benefiting from a moderate class balance and diverse 49-feature representation. The computed F1is:

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

Table 4 Performance of Local Client Models on Held Out Test Sets

Client (Dataset)	Acc.	Prec.	Rec.	F1
Client 1 (CICIDS2017)	0.67	0.69	0.75	0.89
Client 2 (UNSW-NB15)	0.96	0.97	0.96	0.96
Client 3 (NSL-KDD)	0.97	0.93	0.90	0.89

$$F1_2 = 2 \times \frac{0.97 \times 0.96}{0.97 + 0.96} \approx 0.965 \quad (23)$$

Client 3 (NSL-KDD) achieves the highest accuracy (0.97) but lower recall (0.90), indicating a conservative decision boundary that misses ∼10% of attacks— particularly the rare U2R and R2L categories(<1% of the records). The F1 is:

$$F1_3 = 2 \times \frac{0.93 \times 0.90}{0.93 + 0.90} \approx 0.915 \quad (24)$$

The 30-percentage-point accuracy spread(0.67–0.97) across clients is direct empirical evidence of non-IID data heterogeneity, driven by class imbalance disparity, feature dimensionality mismatch, and non-overlapping attack taxonomies.What do these local results tell us about the feasibility of standalone, siloed intrusion detection?

7.2.Global Model Performance

Table 5 Performance of Global Federated Model (Round 10)

Metric	Value
Accuracy	0.87
Precision	0.86
Recall	0.86
F1-score	0.85

The global model achieves 87% accuracy with balanced precision and recall (both 0.86). The global accuracy is the dataset-size-weighted average:

$$\text{Acc}_{\text{global}} = \sum_{k=1}^{3} \frac{n_k}{N} \cdot \text{Acc}_k^{\text{global}} \quad (25)$$

The global model provides a substantial benefit to Client 1 (+20pp) while incurring a modest cost for Clients 2 and 3 (−8 to−11 pp). This is the fundamental generalization specialization trade-off in federated learning: the global model sacrifices per-client specialization to achieve cross domain generalizability. Notably, the weighted average accuracy improves from 0.80 (local) to 0.87 (global) ,confirming a net positive benefit of federated aggregation across the entire system.

with weights $\frac{n_1}{N} \approx 0.51$, $\frac{n_2}{N} \approx 0.45$, $\frac{n_3}{N} \approx 0.03$.

Table VI shows the per-client performance of the global model evaluated on each client's held-out test set, compared to the local model baseline.

TABLE VI
PER-CLIENT PERFORMANCE: LOCAL VS. GLOBAL MODEL

Client	Local Acc.	Global Acc.	ΔAcc.
Client 1 (CICIDS2017)	0.67	0.87	**+0.20**
Client 2 (UNSW-NB15)	0.96	0.88	−0.08
Client 3 (NSL-KDD)	0.97	0.86	−0.11
Weighted Avg.	0.80	**0.87**	**+0.07**

The global model outperforms Client1's local model by 20 percentage points (ΔAcc=+0.20), demonstrating that FedAvg effectively transfers knowledge from high-performing clients (UNSW-NB15, NSL-KDD) to correct Client 1's majority-class bias. The balanced precision-recall symmetry (both 0.86) shows that FedAvg averages out the individual biases of local models: Client 1's low-precision/high-recall bias and Client3's high-precision/low- recall bias cancel in the parameter space producing a globally balanced classifier. Why does the global model (87%) outperform the weighted average of local accuracies (≈80%)? Because parameter space averaging is not equivalent

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

to output-space averaging: the FedAvg global model is a single unified classifier whose parameters encode complementary knowledge from all three clients, enabling it to generalize beyond what any weighted average of local predictions would achieve.

7.3. Convergence Analysis

As shown in Fig. 3, global accuracy improves monotonically from 0.65 (round 1) to 0.87 (round 10), with a diminishing returns pattern: the largest gains occur in rounds 1–4, while rounds 7–10 yield progressively smaller improvements as the model approaches its convergence plateau.

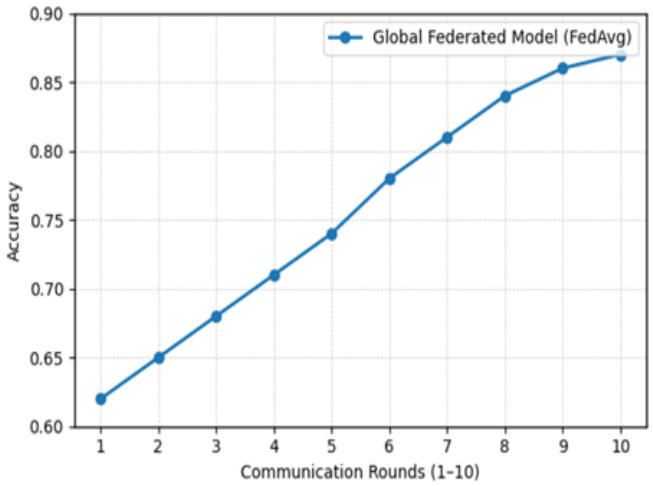

Figure 3 Global model accuracy vs. communication round.Monotonic improvement from 0.65(round1) to 0.87(round10) confirms stable FedAvg convergence under non-IID conditions.

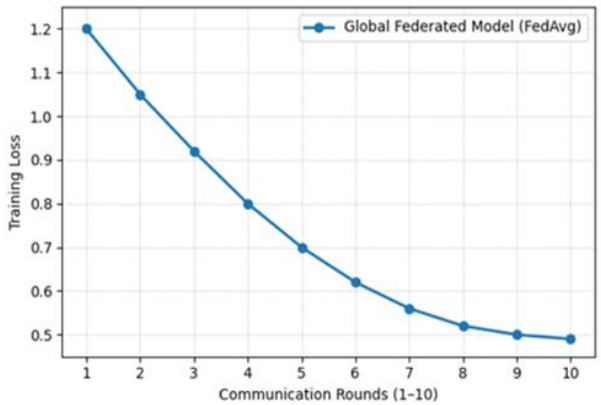

Figure 4 Global training loss vs. communication round. Smooth monotonic decrease from 1.2 to 0.5 confirms numerical stability of FedAvg aggregation across all ten rounds.

The training loss (Fig. 4) decreases from 1.2 to 0.5, with average reduction rate:

$$\bar{r}_{\mathcal{L}} = \frac{\mathcal{L}^{(1)} - \mathcal{L}^{(T)}}{T} = \frac{1.2 - 0.5}{10} = 0.07 \text{ per round} \quad (26)$$

The residual loss gap above the centralized optimum reflects the irreducible convergence error introduced by non-IID distributional divergence $\Gamma > 0$. The absence of oscillation or instability in both curves confirms that the FedAvg aggregation process is numerically stable under the non-IID conditions of the proposed framework.

A key advantage of the federated framework is the dramatic reduction in communication overhead. Centralized training would require transmitting all raw data:

$$C_{\text{centralized}} \approx (2.8\text{M}\times 78 + 2.5\text{M}\times 49 + 149\text{K}\times 41)\times 4 \approx 1.37 \text{ GB} \quad (27)$$

The federated framework transmits only model parameters per round:

$$C_{\text{federated}} \approx T \cdot K \cdot |\mathbf{w}| \cdot 4 = 10 \times 3 \times 10^4 \times 4 \approx 1.2 \text{ MB} \quad (28)$$

This is a ~1000× reduction in communication overhead, achieved without any data compression, while simultaneously guaranteeing that raw traffic data never leaves the client environment.

7.4. Discussion

Non-IID Impact. The 30-point accuracy spread across local models (0.67–0.97) quantifies the severity of distributional divergence. The gradient divergence:

$$\Gamma = F^* - \sum_k \frac{n_k}{N} F_k^* > 0$$

causes client drift during local training. Despite this, **FedAvg** recovers a globally useful model, demonstrating that parameter-space averaging is robust to moderate levels of distributional divergence. The aggregation weights ($n_1/N \approx 0.51$, $n_2/N \approx 0.45$) naturally down-weight Client 3 (NSL KDD, $n_3/N \approx 0.03$), whose legacy attack taxonomy diverges most from the other clients. Generalization-

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

Specialization Trade-off. The per-client results in Table VI reveal a fundamental tension in federated learning: the global model improves the weakest client (+20 pp for Client 1) but slightly degrades the strongest clients (−8 to −11 pp for Clients 2 and 3). This trade-off can be formalized as follows. Let w∗ k = argmin Fk(w) denote the optimal local model for client k. The global model w(T) minimizes the global objective F(w), which is a compromise between all local objectives. The performance gap for client k is:

$$\Delta_k = F_k(\mathbf{w}^{(T)}) - F_k(\mathbf{w}_k^*) \geq 0 \qquad (29)$$

7.5.Limitations

- Byzantine Robustness: FedAvg does not validate client updates, making it vulnerable to model poisoning attacks from malicious clients.
- Byzantine-robust mechanisms (Krum, Bulyan, coordinate-wise median) are needed for adversarial settings.
- Fixed Aggregation Weights: Size-proportional weights do not account for local model quality or class imbalance, potentially over-weighting low-quality updates from severely imbalanced clients like CICIDS2017.
- Simulated Environment: The evaluation uses a simulated federated setup; real deployment introduces network latency, client dropout, and asynchronous updates that can degrade convergence. • Concept Drift: Static dataset evaluation does not capture the evolution of traffic distributions over time. Continual 9 learning mechanisms are needed for production deployment.
- Binary Classification: The framework currently per forms binary (benign/attack) classification. Multi-class attack categorization would provide more actionable intelligence for security analysts

Conclusion

This paper presented a federated intrusion detection framework evaluated across three independently sourced, real-world benchmark datasets — CICIDS2017, UNSW NB15, and NSL-KDD — each assigned to an independent federated client to simulate authentic non-IID data heterogeneity. A lightweight MLP is trained locally at each client, with parameters aggregated using FedAvg. The global model achieves 87% accuracy with balanced precision (0.86), recall (0.86), and F1-score (0.85), outperforming the weakest local model by 20 percentage points. Communication overhead is reduced by ∼1000× compared to centralized training, with no raw data sharing. Convergence analysis confirms stable, monotonic improvement over ten communication rounds, validating the robustness of FedAvg under genuine distributional heterogeneity. Future work will investigate: (F1) Byzantine-robust aggregation (Krum, Bulyan) to defend against model poisoning; (F2) adaptive aggregation weights based on local model quality and class distribution statistics; (F3) personalized FL (FedPer, MAML) to address the generalization-specialization trade-off; (F4) differential privacy integration for formal privacy guarantees; (F5) multi-class attack categorization for more actionable security intelligence; and (F6) continual and asynchronous FL for real-world deployment with concept drift and client dropout.

Acknowledgements

The authors thank Dr. R. Manjula Devi, Professor, Department of Computer Science, KPR Institute of Engineering and Technology, for her invaluable guidance and support throughout this research.

References

[1]. A. Ferrari, S. Rossi, and M. Giacobbe, “DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber–Physical Systems,” IEEE Access, vol. 9, pp. 112345–112357, 2021.
[2]. M. Al-Hawawreh, E. Sitnikova, and N. Aboutorab, “X-IIoTID: A Connectivity-Agnostic and Device-Agnostic Intrusion Data Set for In dustrial Internet of Things,” IEEE Internet of Things Journal, vol. 9, no. 5, pp. 3962–3977, 2022.
[3]. S. Kumar and A. Gupta, “Federated Learning for Network Intrusion Detection Using ResNet,” IEEE Transactions on Network and Service Management, vol. 19, no. 2, pp. 1234–1245, 2022.
[4]. V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0101
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 651 - 662

Huang, A. Dehghantanha, and G. Srivastava, "A Survey on Security and Privacy of Federated Learning," Future Generation Computer Systems, vol. 115, pp. 619 640, 2021.

[5]. A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges," Cybersecurity, vol. 2, no. 1, pp. 1–22, 2019.

[6]. Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated Learning with Non-IID Data," arXiv preprint arXiv:1806.00582, 2018.

[7]. M. A. Rahman, M. S. Hossain, M. S. Islam, N. A. Alrajeh, and G. Muhammad, "Federated Machine Learning for Cyber Attack Detection in IoT," IEEE Internet of Things Journal, vol. 9, no. 11, pp. 8402 8416, 2022.

[8]. N. Albanbay, B. Yilmaz, and S. Albayrak, "Federated Learning-Based Intrusion Detection System for IoT Networks," IEEE Access, vol. 10, pp. 87739–87752, 2022.

[9].. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "In-Edge AI: Intelligentizing Mobile Edge Computing, Caching and Communication with Federated Learning," IEEE Network, vol. 33, no. 5, pp. 156 165, 2019.

[10]. L. Zhou, H. Gao, and W. Duan, "FetFIDS: Federated Transfer Learning for Intrusion Detection Systems," IEEE Transactions on Industrial Informatics, vol. 18, no. 8, pp. 5628–5637, 2022.

[11]. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characteriza tion," in Proc. 4th Int. Conf. Inf. Syst. Secur. Privacy (ICISSP), Madeira, Portugal, Jan. 2018, pp. 108–116.

[12]. . Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," in Proc. IEEE Symp. Comput. Intell. Secur. Defense Appl. (CISDA), Ottawa, ON, Canada, Jul. 2009, pp. 1–6