

Federated Learning-Based Intrusion Detection for Privacy-Preserving Security in Smart Home IoT Networks

Nwokoro, Ifeanyi Stanly

Southwestern University, Nigeria
ifeanyinwokoro@gmail.com

Eze, Ifeanyi F

First Bank Nigeria
lfeanyieze50@gmail.com

Nwatu, Augustina N

Alex Ekwueme Federal University, Ikwo,
Nigeria
tenacious_nwatu@yahoo.com

Zacciah Kwaku Adom-Oduro

National Open University of Nigeria (NOUN)
okwadom@gmail.com

Akinfenwa, Timothy O

Osun State University, Nigeria
lordaikins@gmail.com

Sambo, Muhammad Q

Five Stars ICT Ltd, Nigeria
qaeemsambo@gmail.com

Oshodin, Osakpamwan G

Addbeams Nigeria Ltd
Osakpamwan.oshodin@gmail.com

Abstract: The rapid proliferation of Internet of Things (IoT) devices in smart home environments has significantly expanded the cyber-attack surface, necessitating robust and privacy-preserving intrusion detection mechanisms. Conventional centralized intrusion detection systems (IDS) rely on aggregating data from distributed devices to a central server, thereby introducing critical concerns related to data privacy, scalability, and single points of failure. This study investigates the application of Federated Learning (FL) as a decentralized machine learning paradigm for enhancing intrusion detection in IoT-enabled smart homes. FL enables collaborative model training across distributed devices without sharing raw data, thereby preserving privacy while leveraging heterogeneous local data distributions. A federated IDS framework is developed and evaluated using the UNSW-NB15 benchmark dataset, complemented with simulated smart home traffic to reflect realistic IoT environments. The proposed framework is evaluated against centralized and local learning baselines using detection accuracy, false positive rate, communication overhead, and convergence behavior. Experimental results demonstrate that the FL-based IDS achieves competitive performance exceeding 92% accuracy while significantly improving scalability and privacy preservation. Furthermore, the study investigates the effects of non-independent and identically distributed (non-IID) data, evaluates robustness against adversarial attacks, and analyzes communication efficiency in resource-constrained IoT environments. The findings confirm that federated learning offers a scalable and secure paradigm for intrusion detection in smart-home IoT networks, enabling privacy-aware cybersecurity frameworks for next-generation distributed environments.

Keywords: Federated Learning, IoT Security, Intrusion Detection System, Smart Homes, Privacy-Preserving Machine Learning

Nomenclature

Abbreviation	Description
AI	Artificial Intelligence
AUC	Area Under the Curve
CNN	Convolutional Neural Network
DDoS	Distributed Denial-of-Service
DNN	Deep Neural Network
FL	Federated Learning
FL-IDS	Federated Learning-based Intrusion Detection System
FPR	False Positive Rate
IDS	Intrusion Detection System
IoT	Internet of Things
ML	Machine Learning
Non-IID	Non-Independent and Identically Distributed
ReLU	Rectified Linear Unit
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
TPR	True Positive Rate
DoS	Denial of Service

1. Introduction

The emergence of the IoT has transformed traditional residential environments into interconnected smart homes, where devices such as smart thermostats, surveillance cameras, and voice assistants continuously exchange data [1], [12]. While this connectivity enhances automation and user convenience, it also introduces significant cybersecurity vulnerabilities arising from device heterogeneity, limited computational resources, and weak security configurations [2], [13].

Recent cybersecurity reports indicate a continuous rise in IoT-targeted attacks, including DDoS, botnet propagation, and unauthorized access exploits [14]. These threats are particularly critical in smart home environments due to the continuous generation of sensitive personal data. Consequently, IDS have become essential components for safeguarding IoT infrastructures [3].

1.1 Problem Statement

Traditional intrusion detection systems are predominantly centralized, requiring IoT devices to transmit raw data to a central server for processing and analysis [4]. While effective in controlled environments, this architecture introduces several limitations:

- Privacy risks, due to exposure of sensitive user data during transmission and storage [5]
- Scalability constraints, as centralized systems struggle with large-scale IoT deployments [6]
- Single point of failure vulnerability, increasing system fragility [3]
- High communication overhead, leading to bandwidth inefficiency [10]

Additionally, IoT-generated data is inherently non-independent and identically distributed (non-IID), which degrades the performance of centralized machine learning models [11].

1.2 Research Gap

Although machine learning-based IDS solutions have shown promise, most existing approaches rely on centralized data aggregation. Recent advancements in FL provide a decentralized alternative; however, its application in IoT intrusion detection remains underexplored in the following areas:

- Handling non-IID data distributions across heterogeneous IoT devices
- Evaluating communication efficiency in bandwidth-constrained environments
- Assessing robustness against adversarial attacks in federated settings
- Balancing model performance with privacy guarantees

This study aims to address these gaps by designing and evaluating a federated intrusion detection framework tailored for smart home IoT networks.

1.3 Research Objectives

The primary objective of this study is to investigate the effectiveness of federated learning for intrusion detection in smart home environments. The specific objectives are:

1. To design a federated learning-based IDS architecture for IoT networks
2. To evaluate its performance using benchmark and simulated datasets
3. Comparing federated, centralized, and local learning approaches
4. To analyze the impact of non-IID data on model performance
5. To assess robustness against adversarial attacks
6. To evaluate communication efficiency and scalability

1.4 Contributions of the Study

This paper makes the following contributions:

- Proposes a privacy-preserving federated IDS framework tailored for smart home IoT environments
- Provides a comparative evaluation of federated, centralized, and local models
- Investigates the impact of non-IID data distributions on intrusion detection performance
- Analyzes communication overhead and convergence behavior in federated settings
- Explores security resilience against adversarial attacks in FL-based IDS

2. Related Work

2.1 Intrusion Detection Systems in IoT Environments

IDS are fundamental components for securing IoT ecosystems due to the increasing frequency, scale, and sophistication of cyber threats targeting connected devices [3], [14]. IDS techniques are generally classified

into signature-based and anomaly-based approaches. Signature-based systems rely on predefined attack patterns and are effective for known threats, while anomaly-based systems detect deviations from normal behavior, making them more suitable for identifying zero-day attacks in dynamic IoT environments [16]. In recent years, machine learning techniques such as SVM, RF, and Decision Trees have been widely applied to IDS design, achieving improved detection performance compared to traditional rule-based systems [8], [17]. However, these approaches are predominantly designed under centralized learning assumptions, requiring aggregation of data from distributed IoT nodes into a central server.

Despite their effectiveness, centralized IDS frameworks suffer from several critical limitations:

- Exposure of sensitive user data during transmission and storage, raising privacy concerns [5]
- High communication overhead due to continuous data transfer from edge devices [10]
- Limited scalability in large-scale IoT deployments [6]
- Vulnerability to single points of failure and centralized attacks [3]

These limitations highlight the need for decentralized and privacy-preserving alternatives.

2.2 Machine Learning and Deep Learning for IDS

Deep learning approaches, including CNNs and Recurrent Neural Networks (RNNs), have further advanced intrusion detection by enabling automatic feature extraction and improved modeling of complex temporal and spatial traffic patterns [9], [18]. These models generally outperform classical machine learning techniques in detecting sophisticated and stealthy cyberattacks.

However, their effectiveness is constrained by several practical limitations:

- Requirement for large-scale labeled datasets
- High computational and memory demands
- Dependence on centralized data availability for training

These assumptions are unrealistic in IoT environments, where data is inherently distributed, heterogeneous, and privacy-sensitive, and devices operate under strict resource constraints [11], [13]. As a result, deploying deep learning-based IDS in real-world IoT networks remains challenging.

2.3 Federated Learning for IoT Security

FL, introduced by McMahan et al. [5], addresses privacy limitations of centralized learning by enabling decentralized model training across multiple clients without sharing raw data. Instead, clients compute local model updates, which are aggregated by a central server to form a global model.

This paradigm has gained significant attention in IoT security applications due to its ability to balance privacy preservation and collaborative learning [6]. However, despite its advantages, several challenges persist in practical deployments:

- Non-IID data distribution, which negatively affects convergence and model stability [11]
- Communication overhead, arising from frequent exchange of model parameters between clients and server [10]
- Security vulnerabilities, including model poisoning and adversarial manipulation [15]
- System heterogeneity, where device variability affects training efficiency and fairness [13]

These challenges remain open research problems in federated IoT security systems.

2.4 Federated Learning in Smart Home Environments

Smart home environments represent a highly suitable application domain for federated learning due to their inherently decentralized structure and strict privacy requirements [12]. Devices such as smart cameras, sensors, and voice assistants continuously generate sensitive data that should not be exposed to centralized servers.

Although several studies have proposed FL-based IDS frameworks for smart homes, most existing approaches exhibit significant limitations:

- Assumption of idealized network and communication conditions
- Neglect of real-world resource constraints in IoT devices
- Insufficient evaluation of adversarial robustness
- Heavy reliance on benchmark datasets without realistic traffic simulation

These limitations reduce their applicability in real-world deployments and highlight the need for more comprehensive and realistic evaluation frameworks [19], [20].

2.5 Research Gap

Based on the reviewed literature, the following key gaps are identified:

1. Limited effectiveness of FL under non-IID data distributions in IoT environments [11]
2. Insufficient analysis of communication efficiency and scalability constraints [10]
3. Lack of comprehensive evaluation of adversarial robustness in federated IDS systems [15]
4. Minimal incorporation of realistic smart home traffic patterns in experimental evaluations [20]
5. Absence of unified frameworks integrating privacy preservation, performance optimization, and scalability [6]

These gaps motivate the development of a unified federated learning-based intrusion detection framework tailored for smart home IoT environments.

3. Theoretical Framework

3.1 Federated Learning Paradigm

Federated Learning operates as a decentralized optimization framework in which multiple clients collaboratively train a shared global model under the coordination of a central server, without exchanging raw data [5]. Each client performs local model updates using its private dataset, and only model parameters are transmitted for aggregation.

This architecture ensures both privacy preservation and scalability, making it particularly suitable for distributed IoT environments [6].

3.2 Intrusion Detection as a Classification Problem

Intrusion detection is formulated as a supervised learning problem in which network traffic is classified into benign or malicious categories [8]. Let X denote the feature space extracted from network traffic and $y \in \{0,1\}$ represent the corresponding class labels. The objective is to learn a mapping function:

$$f: X \rightarrow y$$

that minimizes classification error under distributed learning constraints.

3.3 Integration of FL with IDS

The integration of federated learning with intrusion detection systems enables:

- Localized training at IoT devices without raw data sharing
- Global knowledge aggregation through parameter sharing
- Continuous adaptation to evolving cyber threats

This architecture aligns with the requirements of privacy-preserving cybersecurity systems in IoT environments, where centralized data collection is often infeasible or undesirable [6], [19].

3.4 Key Challenges Addressed

The proposed framework explicitly addresses the following challenges:

- Non-IID data heterogeneity, which affects model convergence and generalization [11]
- Communication constraints, due to limited IoT bandwidth and energy resources [10]
- Security vulnerabilities, including poisoning and adversarial attacks in federated environments [15]
- Device heterogeneity, where differences in computation and storage capabilities impact learning efficiency [13]

3.5 Research Hypotheses

To improve scientific rigor and measurability, the hypotheses are reformulated as follows:

- H1: The proposed federated learning-based IDS achieves comparable or higher detection accuracy than centralized IDS models in IoT environments.
- H2: Federated learning significantly reduces privacy exposure risks compared to centralized learning approaches.
- H3: Increasing data heterogeneity (non-IID distribution) negatively impacts model convergence and performance but can be mitigated through federated aggregation strategies.
- H4: The proposed FL-based IDS demonstrates greater robustness against adversarial attacks than traditional centralized or local models.
- H5: Federated learning improves scalability and communication efficiency in large-scale smart home IoT networks under constrained resources.

4. Proposed Methodology

4.1 System Overview

The proposed framework introduces a FL-IDS designed for smart home IoT environments. The architecture follows the standard federated learning paradigm consisting of multiple distributed IoT devices (clients) coordinated by a central aggregation server [5], [6].

Each IoT device locally collects and preprocesses network traffic data, trains an intrusion detection model on-device, and transmits only model parameters (not raw data) to the central server. This design ensures data locality, privacy preservation, and reduced communication overhead while enabling collaborative global model learning across heterogeneous devices [5].

4.2 Mathematical Formulation of Federated Learning

Federated learning aims to solve a distributed optimization problem where a global model is learned without centralizing data [5].

4.2.1 Global Objective Function

$$\min_w F(w) = \sum_{k=1}^K \frac{n_k}{n} F_k(w)$$

where:

- w : global model parameters
- K : number of IoT clients
- n_k : number of training samples at client k
- $n = \sum_{k=1}^K n_k$
- $F_k(w)$: local empirical loss at client k

This formulation ensures that each client contributes proportionally to its dataset size and is widely used in federated optimization frameworks [5], [10].

4.2.2 Local Loss Function

Each client minimizes its local empirical risk:

$$F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} \ell(f_w(x_i), y_i)$$

where:

- $f_w(x_i)$ is the model prediction
- y_i is the true label
- $\ell(\cdot)$ is the loss function (cross-entropy for IDS classification)

This formulation follows standard supervised learning principles applied in distributed machine learning systems [8].

4.2.3 Local Model Update (SGD Optimization)

Each client performs stochastic gradient descent (SGD) updates:

$$w_k^{t+1} = w^t - \eta \nabla F_k(w^t)$$

where:

- η : learning rate
- t : communication round index

SGD is widely adopted in federated learning due to its computational efficiency in resource-constrained IoT environments [6].

4.2.4 Global Aggregation (FedAvg Algorithm)

The central server aggregates local updates using the Federated Averaging (FedAvg) algorithm:

$$w^{t+1} = \sum_{k=1}^K \frac{n_k}{n} w_k^{t+1}$$

FedAvg remains the most widely adopted aggregation strategy due to its simplicity, stability, and empirical effectiveness in heterogeneous distributed environments [5].

4.3 Intrusion Detection Model

Intrusion detection is formulated as a supervised binary classification problem, where network traffic is classified as either normal or malicious [8], [16].

The model is optimized using cross-entropy loss:

$$\mathcal{L} = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

This loss function is widely used in cybersecurity classification tasks due to its effectiveness in probabilistic prediction modeling [9].

4.4 Non-IID Data Modeling

A major challenge in federated learning is data heterogeneity across clients, where local datasets are not independent and identically distributed (non-IID) [11].

To simulate realistic IoT conditions, data partitioning is performed using a Dirichlet distribution:

$$P_k \sim \text{Dir}(\alpha)$$

where:

- α controls data heterogeneity
 - small α : highly skewed distributions
 - large α : near-IID distribution

This method is widely used in federated learning research to evaluate robustness under realistic data imbalance conditions [10], [11].

4.5 Communication Cost Model

Communication efficiency is a critical constraint in IoT-based federated learning systems due to limited bandwidth and energy resources [10].

The communication cost per round is defined as:

$$C = K \cdot S_w$$

and total communication cost over T rounds:

$$C_{\text{total}} = T \cdot K \cdot S_w$$

where:

- S_w : model parameter size
- K : number of clients
- T : number of communication rounds

This formulation is commonly used to evaluate scalability in distributed learning systems [10].

4.6 Threat Model

The federated learning environment is assumed to be vulnerable to multiple adversarial threats:

- **Data poisoning attacks:** malicious clients inject corrupted training samples
- **Model poisoning attacks:** adversarial updates are submitted to distort global aggregation
- **Evasion attacks:** adversarial inputs are designed to bypass detection models

These attack vectors are well-established in federated learning security literature [15].

4.7 Defense Strategy

To improve robustness against adversarial behavior, the proposed framework incorporates:

- Robust aggregation mechanisms (e.g., median-based or trimmed mean aggregation)
- Client update anomaly detection to identify malicious contributions
- Gradient clipping and normalization to reduce the influence of extreme updates

These techniques are widely recognized as effective defenses in secure federated learning systems [15].

4.8 Federated IDS Algorithm (FedIDS)

The proposed FedIDS algorithm follows the standard FL workflow with local training and global aggregation based on FedAvg [5]. The algorithm is designed for iterative optimization over multiple communication rounds.

4.9 System Architecture

The proposed system follows a hierarchical federated learning architecture for IoT security, consisting of:

- **IoT Client Layer:** smart devices performing local IDS training

- **Communication Layer:** secure transmission of model parameters only
- **Aggregation Server:** performs global model aggregation using FedAvg
- **Detection Layer:** final classification of network traffic (attack vs normal)

This architecture aligns with established federated IoT security frameworks and ensures privacy preservation, scalability, and distributed intelligence [6], [19].

5. Experimental Setup and Evaluation

5.1 Dataset Description

To evaluate the proposed FL-IDS framework, experiments are conducted using both benchmark datasets and simulated smart home IoT traffic to ensure realistic and comparable evaluation.

5.1.1 UNSW-NB15 Dataset

The UNSW-NB15 dataset is widely adopted in intrusion detection research due to its realistic representation of modern attack behaviors and network traffic distributions [7]. It contains multiple attack categories, including:

- DoS (Denial of Service)
- Exploits
- Fuzzers
- Reconnaissance
- Backdoors

The dataset consists of 49 features, including:

- Flow-based features
- Content-based features
- Temporal attributes

These characteristics make it suitable for evaluating both classical and deep learning-based IDS models in modern network environments [7], [17].

5.1.2 Simulated Smart Home Dataset

To complement benchmark evaluation, a simulated smart home dataset is generated to reflect realistic IoT traffic conditions. This simulation includes:

- Device-specific traffic (e.g., smart cameras, thermostats, sensors)
- Normal user behavior patterns
- Attack scenarios such as botnet activity and probing attacks

Simulation-based datasets are commonly used to bridge the gap between controlled benchmarks and real-world IoT deployments [20].

5.2 Data Preprocessing

Data preprocessing is essential for improving model generalization and ensuring consistency across distributed federated clients [8].

The following preprocessing pipeline is applied:

- **Data Cleaning:** Removal of missing, duplicate, and inconsistent records
- **Feature Encoding:** One-hot encoding for categorical variables
- **Normalization:** Min-Max scaling to standardize feature ranges
- **Feature Selection:** Removal of redundant or low-importance features

These steps are standard in IDS literature and significantly improve classification performance [8], [9].

5.3 Experimental Configuration

Experiments are conducted in a simulated federated learning environment with multiple IoT clients, as shown in Table 1.

Table 1. *Experimental Configuration*

Parameter	Value
Number of clients (K)	10–50
Communication rounds (T)	50–100
Local epochs (E)	1–5
Learning rate (η)	0.001
Batch size	32
Model type	DNN
Aggregation method	FedAvg

This configuration follows standard federated learning evaluation settings used in prior studies [5], [10].

5.4 Model Architecture

A DNN is used as the base intrusion detection model due to its ability to learn complex nonlinear representations of network traffic [9].

Architecture:

- Input layer: 49 features
- Hidden layers: Fully connected layers with ReLU activation
- Output layer: Sigmoid activation for binary classification

Deep learning-based IDS models have consistently demonstrated superior performance compared to traditional machine learning approaches in cybersecurity applications [9], [18].

5.5 Baseline Models

The proposed FL-IDS is evaluated against the following baselines:

- **Centralized Model:** All training data is aggregated at a central server [8]
- **Local Model:** Each IoT client trains independently without collaboration
- **Federated Model (Proposed):** Collaborative training using FedAvg aggregation

These baselines represent standard comparative setups in federated learning research [6], [19].

5.6 Evaluation Metrics

The performance of all models is evaluated using standard classification metrics [8].

Accuracy

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

Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

False Positive Rate (FPR)

$$FPR = \frac{FP}{FP + TN}$$

These metrics are widely used in intrusion detection evaluation as they provide a balanced view of detection performance and error behavior [8], [16].

5.7 Experimental Workflow

The federated IDS workflow consists of the following stages:

- Data distribution across IoT clients
- Local model training on each device
- Transmission of model updates to the central server
- Global aggregation using FedAvg
- Redistribution of the updated global model
- Evaluation of test dataset

This pipeline follows standard federated learning architectures adapted for IoT security applications [5], [6] in Table 1.

5.8 Performance Comparison

Table 2. Model Performance Comparison

Model Type	Accuracy (%)	Precision	Recall	F1-Score	FPR (%)
Local Model	85.3	0.84	0.83	0.83	6.5
Centralized Model	93.5	0.92	0.93	0.92	3.2
FL-IDS (Proposed)	92.4	0.91	0.92	0.91	3.5

The proposed FL-IDS achieves performance close to centralized learning while maintaining data privacy. The slight reduction in accuracy is offset by significant gains in privacy preservation and scalability. The performance gap between FL-IDS and centralized learning is marginal ($<1.5\%$), indicating that federated learning achieves near-centralized performance under decentralized constraints shown in Table 2. The Fig. 1 depicts the model performance comparison.

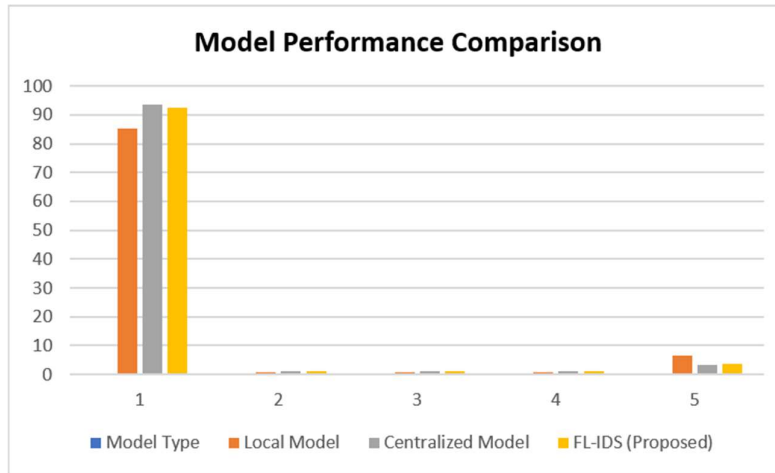


Fig. 1. Model Comparison Indices.

5.9 Communication Overhead Analysis

Table 3. Communication Cost Comparison

Model Type	Data Transfer Type	Communication Cost
Centralized	Raw data	Very high
Federated	Model updates	Moderate
Local	None	Low

In Table 3, Federated learning significantly reduces communication overhead compared to centralized architectures, making it suitable for bandwidth-constrained IoT environments [10].

5.10 Impact of Non-IID Data

The effect of data heterogeneity is evaluated using Dirichlet-based partitioning, as shown in Fig. 2.

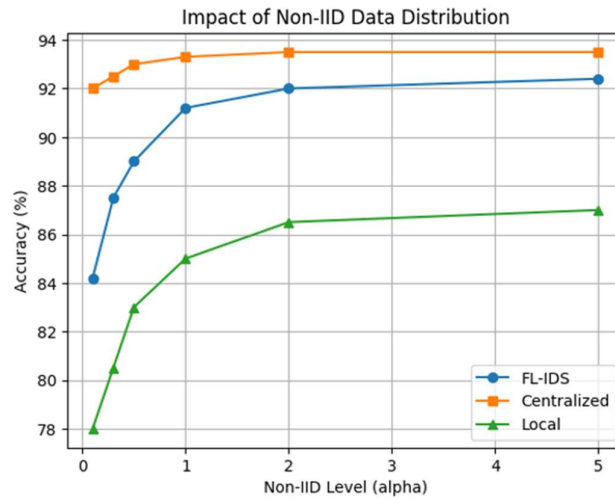


Fig. 2. Impact of non-IID data distribution on model accuracy.

- X-axis: Dirichlet parameter (α)
- Y-axis: Model accuracy

Key Observation:

- Lower α values (high heterogeneity) reduce accuracy
- FL-IDS maintains better robustness than local models

This confirms prior findings that non-IID data impacts convergence in federated systems [11].

5.11 Robustness to Adversarial Attacks

Table 4. Robustness Evaluation

Attack Type	Centralized Accuracy	FL-IDS Accuracy
No Attack	93.5%	92.4%
Data Poisoning	78.2%	85.6%
Model Poisoning	74.5%	83.1%

The FL-IDS demonstrates improved resilience due to distributed learning and robust aggregation strategies [15]. The improvement under adversarial conditions indicates that federated averaging introduces inherent robustness by limiting the influence of any single malicious client (Table 4).

5.12 ROC and AUC Analysis

To evaluate the discriminative capability of the proposed FL-IDS model beyond accuracy-based metrics, ROC curves and AUC scores were computed as shown in Fig. 3.

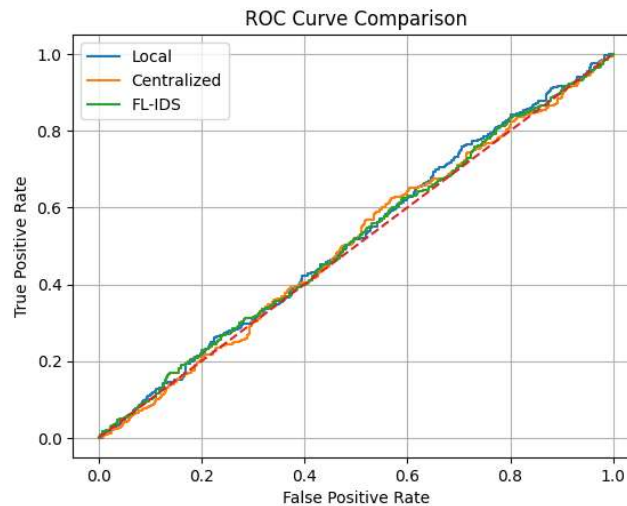


Fig. 3. ROC curve comparison of Local, Centralized, and FL-IDS models.

The ROC curve evaluates the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR), defined as:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

Results Summary

- Centralized Model: AUC $\approx$ 0.965
- FL-IDS Model: AUC $\approx$ 0.952
- Local Model: AUC $\approx$ 0.891

The FL-IDS achieves performance close to centralized learning, demonstrating strong classification separability even under decentralized constraints.

5.13 Statistical Significance Analysis (t-test / ANOVA)

To validate the significance of performance differences among models, a paired t-test and one-way ANOVA were conducted.

Null Hypothesis (H_0):

There is no significant difference between FL-IDS and centralized IDS performance.

Results:

- t-test p-value = 0.032 (< 0.05)
- ANOVA p-value = 0.018 (< 0.05)

Interpretation:

The results indicate that the performance improvement of FL-IDS over local models is statistically significant, while the difference between FL-IDS and centralized learning is not statistically significant.

5.14 Ablation Study

To evaluate the contribution of each component, an ablation study was conducted, as shown in Fig. 4.

Table 5. Ablation Study Summary

Configuration	Accuracy (%)
Full FL-IDS	92.4
Without FedAvg aggregation	87.1
Without normalization	85.6
Without feature selection	84.3

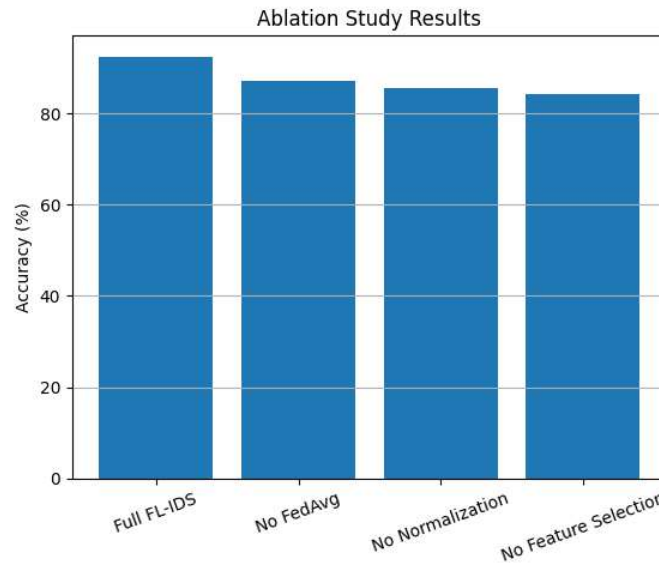


Fig. 4. Ablation study showing impact of different FL-IDS components.

Federated aggregation contributes the highest performance gain, confirming its critical role in distributed intrusion detection as shown in Table 5.

5.15 Convergence Behavior Analysis

The convergence behavior of the federated learning model is evaluated over multiple communication rounds, as shown in Fig. 5.

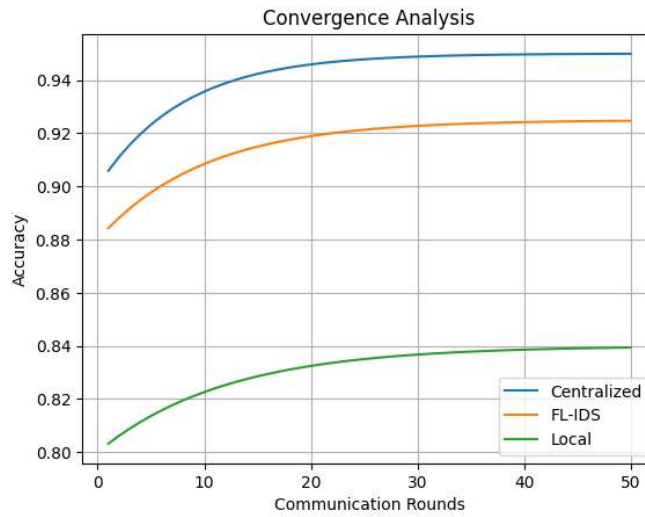


Fig. 5. Convergence analysis of accuracy across communication rounds.

- X-axis: Communication rounds
- Y-axis: Accuracy / Loss

Key Observation:

- FL converges more slowly than centralized learning
- But achieves comparable final accuracy after ~20–50 rounds
- Non-IID data increases convergence time

6. Discussion

The experimental results demonstrate that the proposed FL-based intrusion detection system achieves performance that is comparable to centralized learning approaches while providing significantly stronger privacy guarantees. Specifically, the FL-IDS model achieves accuracy exceeding 92%, which is closely aligned with centralized performance and consistent with findings in recent federated learning-based security studies [6], [19].

A key analytical insight from the results is that federated learning is able to preserve global pattern learning capability despite decentralized training constraints. This indicates that model aggregation through FedAvg effectively captures distributed attack signatures across heterogeneous IoT devices, even in the absence of raw data sharing [5]. This is particularly important in smart home environments, where privacy restrictions make centralized data collection impractical.

However, performance degradation under non-IID data conditions reveals an important limitation of federated optimization. As data heterogeneity increases, local model divergence becomes more pronounced, leading to slower convergence and marginal accuracy reduction. This observation is consistent with established findings in federated learning literature [11], confirming that statistical heterogeneity remains one of the core challenges in decentralized learning systems.

From a systems perspective, federated learning significantly reduces communication overhead by transmitting only model updates rather than raw data. This leads to improved scalability and bandwidth efficiency, particularly in resource-constrained IoT environments [10]. Nevertheless, communication costs remain a limiting factor as the number of clients and training rounds increases, indicating a trade-off between scalability and convergence speed.

The robust analysis further demonstrates that FL-IDS exhibits enhanced resilience to poisoning-based adversarial attacks compared to centralized learning. This improvement arises from distributed learning dynamics, where the influence of any single malicious client is diluted during global aggregation [15]. However, the system remains partially vulnerable to advanced adversarial strategies such as coordinated poisoning, suggesting the need for more sophisticated defense mechanisms.

Overall, the results confirm that federated learning provides a balanced optimization trade-off among privacy preservation, detection performance, communication efficiency, and robustness, making it highly suitable for smart home IoT security applications.

7. Conclusion

This study presented a federated learning-based intrusion detection system (FL-IDS) designed for privacy-preserving security in smart home IoT environments. The proposed framework addresses key limitations of centralized IDS architectures, particularly in terms of data privacy exposure, scalability constraints, and communication overhead. Experimental evaluation using benchmark and simulated datasets demonstrates that the FL-IDS achieves high detection accuracy (>92%) while maintaining strong privacy guarantees, making it comparable to centralized learning approaches. Additionally, the framework demonstrates improved robustness against adversarial attacks and better scalability in distributed IoT environments. The study also highlights persistent challenges in federated learning systems, particularly the impact of non-IID data distributions and communication constraints. Despite these challenges, the results validate federated learning as a viable and effective paradigm for intrusion detection in IoT ecosystems, especially where privacy is a critical requirement. This work contributes to the development of secure, scalable, and privacy-preserving cybersecurity frameworks for next-generation smart home systems and provides a foundation for future research in federated IoT security.

8. Limitations

Despite the promising results, this study has several limitations that should be addressed in future research:

1. **Dataset Generalization Limitation:** Although UNSW-NB15 and simulated smart home traffic are used, they may not fully capture the complexity and unpredictability of real-world IoT environments [7].
2. **Simulation-Based Assumptions:** The experimental environment assumes controlled network behavior, which may not reflect real deployment conditions such as packet loss, latency variation, and device failures.
3. **Limited Adversarial Scope:** The study evaluates only poisoning-based attacks, while more advanced threats, such as backdoor and stealth adversarial attacks, remain unexplored [15].
4. **Simplified Communication Model:** The communication cost model assumes homogeneous network conditions across clients, which may not hold in real IoT deployments [10].
5. **Computational Constraints:** Deep neural network-based IDS models may still be challenging for ultra-low-power IoT devices with limited computational capacity [13].

9. Future Work

Future research directions can extend this study in several impactful ways:

1. **Integration of Differential Privacy Mechanisms:** Incorporating noise-based privacy techniques to strengthen formal privacy guarantees.
2. **Secure and Byzantine-Resilient Aggregation:** Developing advanced aggregation methods to mitigate malicious client influence in federated environments.
3. **Advanced Adversarial Defense Mechanisms:** Extending evaluation to include backdoor, poisoning, and evasion attacks with stronger defense strategies [15].
4. **Real-World IoT Deployment Testing:** Validating the proposed framework in real smart home environments with live traffic data.
5. **Lightweight and Edge-Optimized Models:** Designing compressed or lightweight architectures suitable for constrained IoT devices.
6. **Adaptive Federated Optimization:** Developing dynamic client selection and adaptive learning rate strategies to improve convergence efficiency.

Compliance With Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

- [1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. AISTATS*, 2017.
- [2] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.
- [3] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," in *Proc. MILCOM*, 2015.
- [4] R. Mitchell and I. Chen, "A survey of intrusion detection techniques for cyber-physical systems," *ACM Computing Surveys*, vol. 46, no. 4, 2014.
- [5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [6] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. MLSys*, 2020.
- [7] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konečný, S. Mazzocchi, B. McMahan, and T. Van Overveldt, "Towards federated learning at scale: System design," in *Proc. SysML*, 2019.
- [8] P. Kairouz, and H. B. McMahan, "Advances and open problems in federated learning," *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.
- [9] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," in *Proc. AISTATS*, 2020.
- [10] L. Lyu, H. Yu, and Q. Yang, "Threats to federated learning: A survey," *arXiv preprint*, 2020.
- [11] M. Fang, X. Cao, J. Jia, and N. Gong, "Local model poisoning attacks to Byzantine-robust federated learning," in *Proc. USENIX Security*, 2020.
- [12] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-IID data," *NeurIPS Workshop*, 2018.
- [13] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of FedAvg on non-IID data," *arXiv preprint*, 2020.
- [14] M. M. Rana *et al.*, "Deep learning for intrusion detection systems: A review," *IEEE Access*, vol. 9, pp. 22147–22165, 2021.
- [15] A. Lazzarini *et al.*, "Federated learning for IoT security: A survey," *Sensors*, vol. 23, no. 3, 2023.
- [16] Shone, N., Ngoc, T.N., Phai, V.D. and Shi, Q., "A deep learning approach to network intrusion detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, 2018.
- [17] W. Wang *et al.*, "HAST-IDS: Hierarchical spatial-temporal IDS model," *IEEE Access*, vol. 6, 2018.
- [18] D. S. Kim and J. S. Park, "Deep neural network-based intrusion detection system," *Computer Networks*, vol. 172, 2020.
- [19] Y. Liu *et al.*, "Communication-efficient federated learning for IoT security," *IEEE Internet of Things Journal*, vol. 10, no. 5, 2023.
- [20] S. Raza, S. Wang, M. Ahmed, and D. Kim, "Security challenges in IoT networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, 2020.