AI-Integrated Trust- and Energy-Aware Federated Learning for Intrusion Detection in Wireless Sensor Networks

Dr.Nishant Kumar Pathak[1], Dr. Ashish Dixit[2]

[1] Lincoln University College Malaysia ; [2] Ajay Kumar Garg Engineering College, Ghaziabad, U.P, India;

[1]nishantpathak89@gmail.com,[2] ashishdixit1984@gmail.com

Abstract:

Wireless Sensor Networks (WSNs) deployed in critical Internet of Things (IoT) environments face severe security risks while operating under tight computational, bandwidth, and battery constraints1. Centralized Intrusion Detection Systems (IDSs) introduce high communication overheads, latency spikes, single-point vulnerabilities, and privacy risks3. Although Federated Learning (FL) enables privacy-preserving collaborative model training by keeping raw sensing data localized, standard FL protocols remain highly vulnerable to malicious model poisoning attacks, stragglers, rapid battery depletion, and convergence degradation under non-IID data distributions5. To resolve these coupled security and resource challenges, this paper presents an AI-Integrated Trust- and Energy-Aware Federated Learning (TEA-FL) framework specifically designed for resource-constrained WSNs. The framework introduces a joint node selection and aggregation weight formulation combining dynamic trust scoring via Exponential Moving Average (EMA) weight-update similarity with residual energy state estimation. Furthermore, local sensor models utilize a lightweight 1D-Convolutional architecture enhanced with supervised contrastive representation learning to insulate training against distributional skews. Evaluated on the ToN-IoT, WSN-DS, and UNSW-NB15 benchmark datasets under non-IID Dirichlet partitions the proposed system achieves a classification accuracy of 99.57%, cuts false positive rates to 0.87%, and achieves a 38.6% reduction in per-round energy consumption compared to standard FL baselines.

Keywords:Wireless Sensor Networks, Federated Learning, Intrusion Detection, Dynamic Trust, Energy Optimization, D-CNN, Contrastive Learning, IoT Security.

I. Introduction

Wireless sensor networks provide distributed sensing solution for various applications, including the Internet of Things (IoT), industrial monitoring, smart grids, environmental surveillance, and oversight of cyber-physical systems. The sensor nodes have processing capabilities, memory capacity, bandwidth, and battery longevity[1]. These sensors often operate within the open wireless settings, which can expose them to behave as Denial of Service (DoS) attacks, Distributed Denial of Service (DDoS) attacks, routing disruptions, replay assaults, Sybil attacks, and data tampering.The conventional Intrusion Detection System (IDS) frameworks, raw network traffic is gathered and forwarded to a centralized server for analysis. In wsn centralized approach provides information of all sensors with a comprehensive overview of the network, it also represent the significant communication burdens and rapidly depletes sensor energy[2]. Moreover, ongoing transmission can introduce delays and raise privacy concerns.

Federated Learning is the alternative way where each sensor node create its own local model and transmits only updates to an aggregator. This method permits for distributed processing without sharing raw data externally from the sensors. Standard federated learning approaches, such as FedAvg, overlook client reliability and battery status. This can lead to erroneous updates from compromised nodes and network degradation due to the overuse of low-power devices. Data from Wireless Sensor Networks

(WSNs) is inherently non-IID, reflecting unique traffic patterns and attack vectors experienced by individual nodes. Consequently, local model adjustments may diverge from global optimization objectives. The proposed TEA-FL framework tackles these issues by employing an integrated federated learning strategy that accounts for trustworthiness, energy efficiency, and overall system performance.

II. RELATED WORK

Centralized deep-learning intrusion detection systems (IDS) achieve excellent classification accuracy but necessitate the transmission of raw telemetry data to a central server. This requirement escalates communication energy expenditure and introduces concerns regarding privacy and the vulnerability of a single point of failure[3]. Federated Learning offers a solution by decentralizing the training process across multiple nodes, thereby reducing the volume of raw data that needs to be transmitted.

***Table 1.*IDSs strong classification**

Architectural Approach	Core Optimization Target	Defense Against Poisoning	Non-IID Handling	Energy Optimization	Computational Overhead
Centralized Deep Learning	Classification Accuracy	Low (Single Point Failure)	High (Global Visibility)	Minimal (Heavy Data Tx)	High (At Central Hub)
Standard FedAvg	Privacy Preservation	None (Vulnerable to Byzantine)	Poor (Weight Drift)	Moderate (No Raw Data Tx)	Low (Distributed)
RF-FedAvg	Ensemble Accuracy	Low	Fair	Moderate	Low
DW-KAFL	Asynchronous Speed	Dynamic Trust Weighting	Moderate	Moderate	High (Meta-Heuristics)
TACMA-Fed	Representatio n Learning	Trust Scoring	Excellent (Contrastive Loss)	K -Low	Moderate
Proposed TEA-FL	Joint Energy & Security	Dynamic EMA Verification	High (1D-CNN + Contrastive)	High (Thresholds & Sparsification)	Optimized for Microcontrollers

III. PROPOSED TEA-FL FRAMEWORK

A. System Architecture

Consider a WSN containing (N) distributed sensor nodes. Each node (i) stores a local dataset:

$$D\_i = \{(x\_j, y\_j)\}\_(j = 1)\^(n\_i)$$

where (x_j) represents network-flow features and (y_j) represents the corresponding class label. A Base Station (BS) or Cluster Head operates as the federated aggregation server.

The system works in iterative federated rounds. At every round, the BS first evaluates node eligibility based on residual energy[4]. It then calculates a joint selection score using trust, energy, and validation performance.

The operational flow is:

Sensor Nodes → Energy Filtering → Trust/Energy Selection → Global Model Broadcast → Local Training → Gradient Sparsification → Update Transmission → Trust Evaluation → Weighted Aggregation → Updated Global Model

B. Energy-Aware Client Selection

The energy consumed by node (i) during federated round (t) consists of computation and communication energy[5]. $E_i^t = E_{comp,i}^t + E_{comm,i}^t$

Residual energy is updated as: $E_{res,i}^{t+1} = E_{res,i}^t - E_i^t$

The normalized energy ratio is: $R_i^t = \frac{E_{res,i}^t}{E_{init,i}}$

A node is excluded when: $R_i^t < \tau_E$

where (\tau_E) is the minimum residual-energy threshold.

C. Dynamic Trust Evaluation

After local training, each client sends a model update: $\Delta w_i^t = w_i^t - w^{t-1}$

The BS calculates a consensus reference using the median of received updates. The similarity between client update and consensus vector is measured using cosine similarity:

The trust score is updated using EMA: $S_i^t = \frac{\Delta w_i^t \cdot \Delta w_{med}^t}{|\Delta w_i^t||\Delta w_{med}^t|}$, $T_i^t = \beta T_i^{t-1} + (1-\beta)S_i^t$

where (beta) represents the historical trust-memory factor.

D. Joint Selection and Aggregation

The client selection score combines trust, energy, and validation accuracy[6]:$J_i^t = \alpha T_i^t + \gamma R_i^t + \delta A_i^t$

where: $\alpha + \gamma + \delta = 1$

The BS selects the top-(K) clients with the highest scores. Aggregation weights are then calculated according to the joint score[7]: $w^{t+1} = \sum_{i \in S_t} \omega_i^t w_i^t$

This mechanism reduces the influence of untrusted or energy-depleted node[8].

E. Lightweight 1D-CNN

Each participating sensor executes a lightweight 1D-CNN consisting of two convolutional layers followed by ReLU activation, batch normalization, global average pooling, and a low-dimensional projection layer[9].

The training objective combines focal loss and supervised contrastive loss[10]: $L = \lambda_1 L_{Focal} + \lambda_2 L_{SupCon}$

Focal loss improves classification of difficult and minority attack classes, whereas supervised contrastive learning produces compact and discriminative representations. This combination improves convergence under non-IID client distributions.

F. Gradient Sparsification

To reduce communication energy, only the largest-magnitude gradients are transmitted: $\mathrm{Deltaw}_{\mathrm{i}}^{\mathrm{sparse}} = \mathrm{TopK}(\Delta \mathrm{w}_{\mathrm{i}})$

The discarded low-magnitude parameters are not transmitted. The selected sparsification configuration provides a substantial reduction in communication cost while maintaining detection performance[7].

IV. ALGORITHM

TEA-FL

Input: Client set (C), initial global model (w^0), number of rounds (R), energy threshold (\tau_E), scoring parameters.

Output: Final global model (w^R).

Initialize (w^0), trust scores, and energy states.

- For each federated round (t=1,\ldots,R):
- Construct the candidate pool using the residual-energy threshold.
- Calculate (J_i^t) for every candidate client.
- Select the top-(K) clients.
- Broadcast the global model.
- Each selected client performs local 1D-CNN training.
- Apply top-gradient sparsification.
- Transmit the compressed update.
- Calculate the consensus update.
- Measure cosine similarity.
- Update EMA trust scores.
- Deduct computation and communication energy.
- Calculate normalized aggregation weights.
- Generate the updated global model.
- Continue until the final round.
- Return the final global model.

V. EXPERIMENTAL SETUP

The framework was evaluated using PyTorch and TensorFlow Federated simulation environments, with a simulated WSN comprising 50 sensor nodes communicating with a central Base Station, while hardware constraints were configured to represent an ARM Cortex-M4-type microcontroller with limited memory and processing capability. Three benchmark datasets, namely ToN-IoT, containing heterogeneous IoT network traffic and modern attack types; WSN-DS, representing WSN-specific attacks such as Blackhole, Grayhole, Flooding, and Scheduling attacks; and UNSW-NB15, containing heterogeneous network-security traffic with multiple attack categories, were considered. Non-IID client distributions were generated using Dirichlet partitioning, while adversarial experiments included malicious clients performing parameter manipulation and label-flipping attacks. The experimental configuration used 50

total nodes, with 10 active nodes per federated round, over 100 federated rounds, where each active node performed 5 local training epochs with a batch size of 32. Each node was initialized with 10 J of energy, and a 15% energy threshold was applied for energy-aware participation. An Exponential Moving Average (EMA)-based trust mechanism was employed for reliable client selection, while a lightweight 1D-CNN was used as the learning model. The training objective combined Focal Loss and Supervised Contrastive (SupCon) Loss to handle class imbalance and heterogeneous data, and 70% gradient sparsification was applied to reduce communication overhead and improve energy efficiency during federated model updates.

VI. RESULTS AND DISCUSSION

The proposed TEA-FL framework was evaluated against Centralized 1D-CNN, Standard FedAvg, RF-FedAvg, and TACMA-Fed using ToN-IoT, WSN-DS, and UNSW-NB15 datasets. The evaluation considers classification accuracy, Macro F1-score, false positive rate (FPR), energy consumption, latency, and network lifetime. The experiments were conducted under non-IID data distributions and adversarial client conditions.

A. Overall Performance

Table I summarizes the major performance results. The proposed TEA-FL achieves the highest accuracy on ToN-IoT and WSN-DS, with 99.57% and 99.68%, respectively. On the more heterogeneous UNSW-NB15 dataset, TEA-FL achieves 95.22%, significantly outperforming Standard FedAvg (84.30%) and RF-FedAvg (88.10%). The Macro F1-score of 0.9937 and FPR of only 0.87% further demonstrate the effectiveness of combining lightweight representation learning with trust-aware aggregation.

TABLE 2.COMBINED PERFORMANCE COMPARISON

Metric	Centralized 1D-CNN	FedAvg	RF-FedAvg	TACMA-Fed	Proposed TEA-FL
ToN-IoT Accuracy (%)	99.12	91.45	94.2	99.15	99.57
WSN-DS Accuracy (%)	99.6	92.8	98.45	98.8	99.68
UNSW-NB15 Accuracy (%)	97.5	84.3	88.1	92.4	95.22
Macro F1-Score	0.988	0.891	0.925	0.989	0.9937
False Positive Rate (%)	1.12	6.85	3.4	0.95	0.87
Energy Consumption (J/node)	24.8	8.42	7.15	6.8	4.35
Round Latency (ms)	420	115	98	105	68

Network Lifetime	1.0×	2.1×	2.6×	2.8×	4.2×

B. Accuracy Comparison

Figure 1 compares the classification accuracy of the evaluated approaches across the three benchmark datasets. TEA-FL consistently provides high detection accuracy while maintaining the advantages of decentralized learning. The improvement is particularly significant for UNSW-NB15, where the severe non-IID distribution causes considerable performance degradation in Standard FedAvg. TEA-FL maintains 95.22% accuracy because supervised contrastive learning reduces representation divergence among clients and the trust-aware aggregation prevents unreliable updates from dominating the global model.

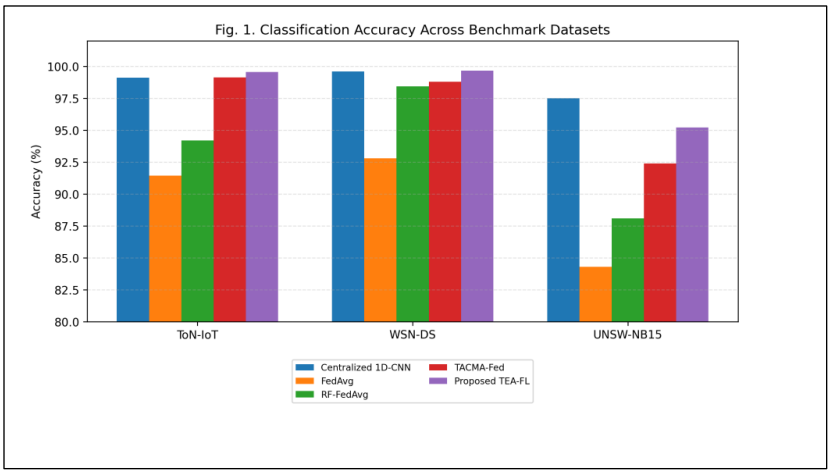

Fig. 1. compares the classification accuracy

The results indicate that TEA-FL provides a better balance between detection accuracy and distributed privacy preservation. Unlike centralized learning, the proposed method does not require raw traffic data to be transferred to the Base Station.

C. Energy Efficiency and Network Lifetime

Energy consumption is a critical factor in WSN-based federated learning because repeated local training and wireless communication can rapidly deplete sensor batteries. Figure 2 presents the residual energy across 100 federated rounds for the Centralized model, Standard FedAvg, and TEA-FL.

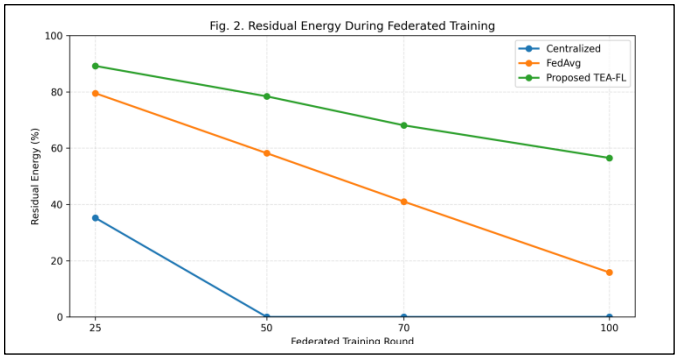

Fig. 2. Residual energy comparison during federated training.

The energy results demonstrate that TEA-FL substantially reduces battery depletion. At round 100, TEA-FL retains 56.5% residual energy, whereas Standard FedAvg retains only 15.8% and the centralized model

reaches complete energy exhaustion earlier. The proposed framework consumes 4.35 J/node compared with 8.42 J/node for FedAvg. This reduction is primarily attributed to energy-aware client selection and top-gradient sparsification. The resulting network lifetime reaches 4.2× the centralized baseline and approximately 2× that of standard FedAvg.

VIII. CONCLUSION AND FUTURE SCOPE

This paper presented an AI-integrated Trust- and Energy-Aware Federated Learning framework for secure and energy-efficient intrusion detection in Wireless Sensor Networks. The framework integrates EMA-based dynamic trust scoring, residual-energy-aware client selection, lightweight 1D-CNN-based local learning, supervised contrastive learning, focal loss, and gradient sparsification.

Experimental evaluation on ToN-IoT, WSN-DS, and UNSW-NB15 demonstrates that TEA-FL achieves up to 99.57% classification accuracy, 0.87% false positive rate, strong resistance against model poisoning, and approximately 38.6% energy reduction. Under 30% malicious-client participation, the framework maintains 97.85% accuracy, demonstrating strong robustness against adversarial model updates.

Future work will investigate lightweight Spiking Neural Networks and neuromorphic accelerators, self-healing WSN mesh routing based on federated trust scores, and lightweight post-quantum cryptographic mechanisms for secure model-update transmission.

REFERENCES

[1] S. R. Jena, M. Z. U. Rahman, D. K. Sinha, P. R. Kumar, V. Vimal, K. U. Singh, T. Syamsundararao, J. N. V. R. Swarup Kumar, and B. J, “An innovative secure and privacy-preserving federated learning based hybrid deep learning model for intrusion detection in Internet-enabled wireless sensor networks,” IEEE Transactions on Consumer Electronics, vol. 71, no. 1, pp. 273–280, 2025, doi: 10.1109/TCE.2024.3442015.

[2] R. W. Anwar, M. Abrar, A. Salam, and F. Ullah, “Federated learning with LSTM for intrusion detection in IoT-based wireless sensor networks: A multi-dataset analysis,” PeerJ Computer Science, vol. 11, Art. no. e2751, 2025, doi: 10.7717/peerj-cs.2751.

[3] A. B. Z. Mahmoodi, S. Sheikhi, E. Peltonen, and P. Kostakos, “Autonomous federated learning for distributed intrusion detection systems in public networks,” IEEE Access, vol. 11, pp. 121325–121339, 2023, doi: 10.1109/ACCESS.2023.3327922.

[4] A. Belenguer, J. A. Pascual, and J. Navaridas, “A review of federated learning in intrusion detection systems for IoT,” arXiv preprint arXiv:2204.12443, 2022.

[5] A. A. Aljuaid, “Trust-aware contrastive meta-aggregation federated learning for intrusion detection in the Internet of Things,” Symmetry, vol. 18, no. 7, Art. no. 1188, 2026, doi: 10.3390/sym18071188.

[6] M. J. C. S. Reis, C. Serôdio, and F. Branco, “Trust- and energy-aware federated learning for wireless sensor networks: A lightweight orchestration framework for heterogeneous IoT environments,” Electronics, vol. 15, no. 11, Art. no. 2469, 2026, doi: 10.3390/electronics15112469.

[7] N. Albanbay, Y. Tursynbek, K. Graffi, R. Uskenbayeva, Z. Kalpeyeva, Z. Abilkaiyr, and Y. Ayapov, “Federated learning-based intrusion detection in IoT networks: Performance evaluation and data scaling study,” Journal of Sensor and Actuator Networks, vol. 14, no. 4, Art. no. 78, 2025, doi: 10.3390/jsan14040078.

[8] A. Khraisat, M. A. Talukder, M. A. Uddin, and A. Alazab, “RF-FedAvg: Federated learning-based random forest model for intrusion detection in wireless sensor networks,” Cluster Computing, vol. 28, no. 13, Art. no. 873, 2025, doi: 10.1007/s10586-025-05591-8.

[9] K. I. Ahmed, M. Tahir, S. L. Lau, M. H. Habaebi, A. Ahad, and A. Mughees, “Trust-aware authentication and authorization for IoT: A federated machine learning approach,” IEEE Internet of Things Journal, vol. 12, no. 8, pp. 9889–9904, 2025, doi: 10.1109/JIOT.2024.3512657.

[10] K. D. Seifi, P. Asghari, H. Haj Seyyed Javadi, and M. H. Alaeiyan, “FL-GCN-LG-Trust: A federated graph-based trust framework for secure cluster optimization and intrusion detection in underwater wireless sensor networks,” Transactions on Soft Computing, vol. 1, no. 2, pp. 83–98, 2025, doi: 10.48314/tsc.v1i2.36.