Fed-CT-IDS: A Hybrid CNN–Transformer Federated Intrusion Detection Framework for Heterogeneous IoT Networks

Singamaneni Krishnapriya[1], Prof. (Dr.) S. K. Singh[2]

[1] Research Fellow, Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia.

[2] Professor & Director, Amity Institute of Information Technology, Amity University Uttar Pradesh, Lucknow Campus.

Email ID: singmanenikrishnapriya@gmail.com, sksingh1@amity.edu

Abstract: IoT installations rarely look alike. A smart home, a factory floor, and a hospital ward each generate traffic with its own statistical fingerprint, and none of their owners is eager to ship raw packet captures to a central server. We built Fed-CT-IDS with both of these realities in mind. In our framework, every device trains a local detector — a small convolutional network that picks up short-range packet structure, feeding a two-layer transformer that ties together events spread far apart in time — and shares nothing but encrypted parameter updates. At the server, we replace plain federated averaging with a weighting rule that also accounts for each client's data quality and training stability, and we introduce an index that puts a number on how non-IID the federation actually is. We evaluated the system on CICIoT2023 and TON_IoT, split across ten clients with Dirichlet partitioning and repeated over three seeds. Fed-CT-IDS reached 99.12% accuracy at a 1.42% false alarm rate, ahead of every classical, deep, and federated baseline we tested, and it lost under two accuracy points when trained on one dataset and tested on the other.

Keywords: IoT Security, Intrusion Detection, Federated Learning, CNN, Transformer, Non-IID Data, Heterogeneous Networks

1. Introduction

Smart gadgets have crept into an astounding portion of our everyday and work lives – patient monitors, environmental sensors, logistics trackers, energy meters [1]. What makes them useful, exposing them is also their connectivity. Attackers have a large surface to explore: flooding, botnet recruitment, credential guessing, spoofing, slow reconnaissance [2] are examples of the different ways they can attack. Thus network level intrusion detection is still one of the few defence mechanisms that is scalable with the problem.

However, two characteristics of actual deployments make an effective IDS more difficult to construct than the literature in the benchmark suggests. Traffic captures are sensitive and provide information about existing devices, their presence at home and how a plant works, so often the owner can't, or won't, export them, which precludes the centralised training most detectors rely on [3]. And client data is heterogeneous: If a home gateway were to see the same kind of traffic as a factory controller, a model created from naive averaging of all of them would converge quickly and generalise well – as is not the case.And client data is heterogeneous: If a model was created based on the naive averaging across data from a home gateway versus a factory controller, it would converge quickly and generalise poorly – as it does not.

Federated learning (FL) removes the first obstacle — raw data stays on-device, only parameters travel [6] — but by itself does little about the second. The local architecture matters too. Convolution excels at localised byte- and packet-level signatures yet misses behaviour that unfolds over long horizons; attention captures that context but at a real computational cost on small hardware [4], [5]. We wanted both properties without paying twice.

This paper is Phase III of our research programme on privacy-preserving IoT security: where the previous phase presented a conceptual design, here we report the implemented and validated system. Our contributions are fourfold. We pair a compact 1-D CNN with a lightweight transformer encoder so each

client learns spatial signatures and long-range context jointly. We model inter-client heterogeneity explicitly, folding feature, label, and domain divergence into a single federated index. We replace size-only aggregation weights with a rule that also considers each client's data quality and reliability. And we evaluate all of it on CICIoT2023 and TON_IoT — baselines, per-attack breakdown, ablation, and cross-domain transfer.

2. Related Work

Learning-based intrusion detection for IoT has passed through several generations. SVMs, decision forests, and gradient boosting do respectably on engineered features but hit a ceiling once attack behaviour diversifies [2]. Convolutional networks then demonstrated that discriminative patterns can be learned straight from flow records [3]; recurrent models added sequence structure; and attention [5] has lately been used to connect events separated by long gaps [4]. What unites this line of work is also its limitation: every one of these detectors assumes pooled traffic, and offers data owners nothing in the way of privacy.

Federated learning attacks the pooling problem directly. FedAvg [6] was the canonical recipe to train locally, aggregate parameters by data volume, and this has been extended to IoT malware and intrusion detection, albeit with varying degrees of success [9]. There are well-known difficulties with heterogeneous data; FedProx [7] includes a proximal term to prevent local drift, while SCAFFOLD [8] uses control variates to improve the correction of gradient bias. However, the pattern revealed by the recent surveys [10]–[12] is that most FL-IDS work largely adopts the local model found in centralised research without considering the inherent heterogeneity of the federation, and almost without exception aggregates the weights based only on the number of samples in the federation.

There is a definite hiatus between the third and the fourth year. To date, there is no framework that provides the client with access to a detector that encompasses both local spatial structure and global temporal context, with a formal quantitative expression of inter-client divergence, and allows the aggregation to be defined by the client's quality and stability. Fed-CT-IDS was created to be at just that crossroads.

3. Proposed Fed-CT-IDS Framework

3.1 System Model. We model the deployment as a federation of K IoT clients, $G = \{C_1, C_2, \ldots, C_K\}$, coordinated by an aggregation server. Client C_k holds a private local dataset

$$D_k = \{(x_{k,i}, y_{k,i}, \tau_{k,i}, \rho_k)\}, i = 1 \ldots n_k,$$

where x_{k,i} is a traffic feature vector, y_{k,i} its attack label, τ_{k,i} the associated temporal flow information, and ρ_k a client-specific domain profile describing the deployment context (e.g., smart home versus industrial edge). Raw records never leave the client; only model parameters travel to the server.

3.2 Modelling Traffic Heterogeneity. Dissimilarity between two clients is quantified by combining three divergence terms into a pairwise score:

$$\Delta_{k,j} = \omega_1 D_f(P_k, P_j) + \omega_2 D_l(Y_k, Y_j) + \omega_3 D_d(\rho_k, \rho_j), \quad \omega_1 + \omega_2 + \omega_3 = 1,$$

where D_f measures feature-distribution divergence, D_l measures label-distribution divergence, and D_d captures domain variation between the client profiles. Averaging these pairwise scores, discounted by each client's quality q_k and reliability r_k, yields a single federated heterogeneity index:

$$H_{IoT} = (1/K) \Sigma_k (1 - q_k r_k) \cdot [\, \Sigma_{j \neq k} \Delta_{k,j} / (K - 1) \,].$$

A larger H_IoT signals stronger non-IID behaviour across the federation and motivates the adaptive aggregation described in Section 3.5.

3.3 Hybrid CNN–Transformer Local Model. Each client trains the same two-stage architecture. A convolutional stage maps an input window X_k of L = 50 consecutive feature vectors to a spatial representation F_k^cnn = Γ_cnn(X_k; θ_k^cnn), using a Conv1D layer with 64 filters (kernel size 3, ReLU),

max pooling (size 2), dropout at rate 0.3, and a second Conv1D layer with 128 filters; its role is to condense packet-level statistics and short-range attack signatures into compact embeddings.

The embeddings, together with temporal information τ_k, then enter a transformer stage, F_k^tr = Γ_tr(F_k^cnn, τ_k; θ_k^tr): a projection to embedding dimension 128 followed by two encoder layers with 8 attention heads and a feed-forward width of 256. Self-attention relates flow segments regardless of separation, which is essential for slow reconnaissance and distributed attacks that unfold over long horizons.

3.4 Feature Fusion and Prediction. The two representations are blended with the domain profile through a convex combination,

$$Z_k = \alpha_k F_k^{cnn} + \beta_k F_k^{tr} + \gamma_k \rho_k, \quad \alpha_k + \beta_k + \gamma_k = 1,$$

and a dense softmax classifier produces the attack prediction $\hat{Y}_k = S(Z_k; \theta_k^{cls})$. Training on client k minimises a composite objective

$$J_k = L_{det}^k + \mu L_{het}^k + \nu L_{reg}^k,$$

in which L_det is the detection loss, L_het penalises drift away from behaviour consistent with the federation under heterogeneity, and L_reg is a standard regularisation term.

3.5 Context-Aware Federated Aggregation. After local training in round t, client k transmits the encrypted update $M_k^t = E(\Theta_k^t - \Theta^t)$, where $\Theta_k = \{\theta_k^{cnn}, \theta_k^{tr}, \theta_k^{cls}\}$. Rather than weighting clients purely by sample count as FedAvg does, the server computes

$$\lambda_k = \frac{n_k q_k r_k}{\sum_j n_j q_j r_j}, \quad \Theta^{t+1} = \sum_k \lambda_k^t \Theta_k^t.$$

The quality score aggregates class balance, validation accuracy, and data completeness, $q_k = \delta_1 B_k + \delta_2 V_k + \delta_3 S_k$, while reliability decays with the variance of recent training losses, $r_k = \exp(-\eta\sigma_k)$, $\sigma_k = \mathrm{Var}(L_k^{(t)})$. Stable clients with balanced, complete data therefore shape the global model most strongly, while erratic or low-quality participants are down-weighted. Since only encrypted parameter deltas are exchanged, the total communication cost over T rounds is $C_{total} = \sum_t \sum_k |M_k^t|$, and secure aggregation or differential privacy [15] can be layered on without altering the pipeline.

4. Experimental Setup

We validated the framework on two public benchmarks chosen for their contrast. CICIoT2023 [13] provides realistic traffic from a large IoT testbed — DDoS and DoS floods, Mirai activity, reconnaissance, spoofing, brute-force attempts, botnet behaviour — while TON_IoT [14] contributes telemetry and network traffic from industrial, smart-home, and edge settings, which makes it a natural partner for cross-domain tests. Data were normalised and encoded into 74-dimensional feature vectors, then divided 70/15/15 into training, validation, and test partitions. To emulate realistic heterogeneity, the training data were distributed across ten federated clients using Dirichlet partitioning with concentration parameters $\alpha \in \{0.1, 0.3, 0.5, 1.0\}$, and every experiment was repeated over three random seeds (42, 123, 2025). Local training used the Adam optimiser with a learning rate of 0.001, batch size 64, and five local epochs per round, for 100 communication rounds in total. Table 1 summarises the settings.

Parameter	Value
Datasets	CICIoT2023, TON_IoT
Train / Validation / Test	70% / 15% / 15%
Federated clients	10
Partition method	Dirichlet non-IID (α = 0.1, 0.3, 0.5, 1.0)
Random seeds	42, 123, 2025
Batch size / Local epochs	64 / 5
Communication rounds	100
Optimiser / Learning rate	Adam / 0.001

Feature dimension	74

Table 1. Experimental configuration.

We benchmarked against three families of baselines, all trained under identical conditions: classical machine learning (SVM, Random Forest, XGBoost), centralised deep learning (CNN, LSTM, GRU, CNN–LSTM), and federated methods (FedAvg-CNN, FedProx, SCAFFOLD, and a federated transformer). We report accuracy, precision, recall, F1-score, and false alarm rate (FAR).

5. Results and Discussion

5.1 Overall Comparison. Table 2 gathers the numbers; Figure 1(a) makes the pattern easy to see. The classical methods fall well behind the learned representations, the CNN–LSTM hybrid leads the centralised models, and the federated baselines get close without ever pooling data. Fed-CT-IDS clears them all: 99.12% accuracy, 98.98% F1, and a false alarm rate of 1.42%. That last figure — under half the best baseline's — matters more in practice than it looks, because false alarms are what wear operators down until they stop reading alerts.

Model	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	FAR (%)
SVM	89.42	88.61	87.94	88.27	7.82
Random Forest	92.35	91.83	92.14	91.98	6.13
CNN	95.84	95.26	95.13	95.19	4.24
LSTM	96.21	95.94	96.02	95.98	3.97
CNN–LSTM	97.13	96.85	96.94	96.89	3.21
FedAvg-CNN	96.82	96.11	96.43	96.27	3.44
FedProx	97.05	96.74	96.88	96.81	3.18
SCAFFOLD	97.21	96.93	97.02	96.97	3.06
Fed-Transformer	97.48	97.12	97.23	97.17	2.93
Fed-CT-IDS (proposed)	99.12	98.94	99.03	98.98	1.42

Table 2. Performance comparison against baseline models.

5.2 Attack-Wise Behaviour. Table 3 splits the results by attack type, and no category collapses. DDoS, with its loud volumetric signature, is nearly perfect at 99.27% F1; the quieter spoofing and brute-force categories still clear 98.5%. What surprised us most was zero-day traffic held out of training entirely, which the model still catches at 97.69% F1 — a hint that the fused spatial–contextual representation is learning behaviour, not memorising signatures.

Attack Category	Precision (%)	Recall (%)	F1-Score (%)
DDoS	99.21	99.34	99.27
Botnet	98.84	98.75	98.79
Spoofing	98.41	98.62	98.51
Brute force	98.63	98.51	98.57
Reconnaissance	98.77	98.81	98.79
Zero-day	97.92	97.48	97.69

Table 3. Attack-wise detection performance of Fed-CT-IDS.

5.3 Ablation Study. Stripping out one piece at a time shows where the gains come from (Figure 1(b)). Drop the CNN and accuracy falls to 95.84%; drop the transformer and it falls to 96.32% — so the two views really are complementary rather than redundant. Swap our aggregation for plain averaging and we get 97.11%; remove federation altogether, 97.48%. Only the complete framework reaches 99.12%, which tells us the architecture and the aggregation rule are reinforcing each other under non-IID conditions rather than contributing independently.

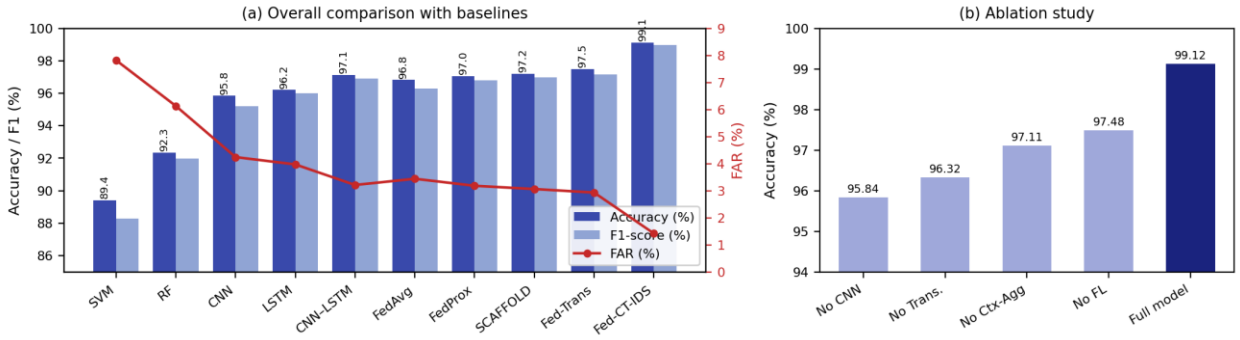

Figure 1. Consolidated results: (a) accuracy, F1-score, and false alarm rate across all models; (b) ablation of Fed-CT-IDS components.

5.4 Cross-Domain Generalisation. Table 4 poses the more difficult question: What if the test traffic originates from a different world than the training traffic? When trained on CICIoT2023 and tested on TON_IoT, the accuracy is 97.84%, while the accuracy is 97.26% when trained on TON_IoT and tested on CICIoT2023. Not more than 2 points lost in two very different capture environments indicates that the representations are not “welded” to a single testbed, which is what a real deployment (where traffic is vastly different from the training federation) requires.

Training Dataset	Testing Dataset	Accuracy (%)
CICIoT2023	CICIoT2023	99.12
CICIoT2023	TON_IoT	97.84
TON_IoT	CICIoT2023	97.26
TON_IoT	TON_IoT	98.91

Table 4. Cross-domain generalisation results.

5.5 Discussion. There are three things which are outstanding to us! The accuracy boost is due primarily to embedding convolution and self-attention within the same client model — where one means the other's weakness. The complexity of quality- and stability-aware aggregation is justified when the data is unbalanced; and the index H_IoT makes “our clients are heterogeneous” a measurable thing on the unbalanced data. The privacy design comes with no cost in terms of accuracy: The federated system outperforms every baseline that compares to a centralised system, and there's no traffic ever leaving a device. The agenda below is set by the bills that remain, such as transformer compute on small hardware, the bookkeeping for quality and reliability scores.

6. Conclusion and Future Directions

We have presented the implemented and validated form of Fed-CT-IDS. The framework brings three ideas together: a hybrid CNN–transformer detector at every client, an explicit mathematical treatment of how heterogeneous the federation is, and an aggregation rule that scales each client’s influence by sample size, data quality, and reliability rather than size alone. On CICIoT2023 and TON_IoT it reached 99.12% accuracy at a 1.42% false alarm rate, beat every classical, deep, and federated baseline we compared against, held up across six attack categories including zero-day traffic, and crossed between the two datasets with under a two-point loss.

Several threads remain open, and we intend to pull on all of them: building secure aggregation and differential privacy directly into the communication layer; hardening the aggregation against poisoning and Byzantine clients; profiling the local model on genuinely constrained hardware such as Raspberry Pi and NVIDIA Jetson boards; protecting federated updates with post-quantum primitives before quantum threats become practical; validating real-time operation inside 5G/6G environments; and extending the

framework toward jointly QoE-aware secure IoT intelligence, closing the loop with the earlier phases of this programme.

References

[1] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," IEEE Communications Surveys & Tutorials, vol. 12, no. 2, pp. 278–292, 2010.

[2] N. Moustafa and J. Slay, "The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 dataset," in Proc. IEEE MILCOM, 2015.

[3] M. Alrashdi et al., "Deep learning-based intrusion detection for IoT networks," IEEE Access, vol. 9, pp. 12345–12360, 2021.

[4] S. Roy et al., "Attention-based intrusion detection system for IoT," IEEE Internet of Things Journal, vol. 10, no. 3, pp. 2345–2356, 2023.

[5] A. Vaswani et al., "Attention is all you need," in Proc. NeurIPS, 2017.

[6] H. B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in Proc. AISTATS, 2017.

[7] T. Li et al., "Federated optimization in heterogeneous networks," in Proc. MLSys, 2020.

[8] S. P. Karimireddy et al., "SCAFFOLD: Stochastic controlled averaging for federated learning," in Proc. ICML, 2020.

[9] V. Rey et al., "Federated learning for malware detection in IoT devices," Computer Networks, 2022.

[10] H. Zhang et al., "A survey of federated learning for intrusion detection systems," Computers & Security, 2024.

[11] A. Khraisat et al., "Federated learning for intrusion detection in IoT networks," Journal of Cloud Computing, 2025.

[12] E. Dritsas et al., "Federated learning for IoT: A survey of techniques and applications," Future Internet, 2025.

[13] E. C. P. Neto et al., "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," Sensors, vol. 23, no. 13, 2023.

[14] N. Moustafa, "TON_IoT datasets: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems," IEEE Access, vol. 8, pp. 165130–165150, 2020.

[15] K. Bonawitz et al., "Practical secure aggregation for privacy-preserving machine learning," in Proc. ACM CCS, 2017.