

FADIoT: Federated GANomaly with Temporal NetFlow Features and Perturbation-Aware Training for Privacy-Preserving IoT DDoS Detection

Danni Yin^{*}, Aniket Mahanti^{*}, Kalpdrum Passi[†], Ranesh Naha[‡], Saqib Hakak[§]

^{*}University of Auckland, New Zealand; [†]Laurentian University, Canada

[‡]Queensland University of Technology, Australia; [§]University of New Brunswick, Canada

*dyin891@aucklanduni.ac.nz; *a.mahanti@auckland.ac.nz; †kpassi@laurentian.ca

‡ranesh.naha@qut.edu.au; §saqib.hakak@unb.ca

Abstract—Internet of Things (IoT) infrastructures remain vulnerable to distributed denial-of-service (DDoS) attacks. The conditions of real deployments, namely cross-domain heterogeneity, fast-evolving traffic, and privacy constraints, undermine the centralized and supervised detectors that dominate the literature. We present FADIoT, a federated anomaly detection framework that combines three components suited to these conditions: NetFlow-v3 temporal features, GANomaly-based one-class learning, and Fast Gradient Sign Method (FGSM)-based perturbation-aware training. Each client models benign traffic locally and shares only model parameters through Federated Averaging (FedAvg), so raw flows never leave their domain of origin. Perturbations are injected during local training as a client-side optimization mechanism rather than a test-time robustness guarantee. Across three heterogeneous IoT benchmark datasets (NF-CSE-CIC-IDS2018-v3, NF-BoT-IoT-v3, NF-ToN-IoT-v3), temporal descriptors raise F1-score by up to 0.28 on the benign-rich datasets. A single federated model then retains usable F1 across all three participating domains, with AUROC between 0.894 and 0.955, whereas single-domain models lose as much as 0.91 F1 once applied outside their training domain. Privacy-preserving IoT DDoS detection with label-free detector training is therefore attainable by combining temporal enrichment, federated aggregation, and perturbation-aware training, without centralized data collection.

Index Terms—Federated learning, Perturbation-aware training, IoT security, DDoS

I. INTRODUCTION

The Internet of Things (IoT) has become a critical component of modern infrastructures, enabling large-scale connectivity across domains such as healthcare, transportation, manufacturing, and smart homes. However, the increasing number of connected devices has also expanded the attack surface for distributed denial-of-service (DDoS) attacks, which remain one of the most prevalent threats against IoT environments [1].

Deep learning has advanced anomaly-based detection, which flags previously unseen attacks without predefined signatures [2]. Three obstacles remain before such methods are dependable in IoT deployments. First, most existing approaches assume centralized access to network traffic, raising privacy and deployment concerns. Second, federated IoT intrusion detection systems often rely on labeled attack samples or focus solely on collaborative training, leaving the impact of perturbation-aware optimization largely unexplored. Third,

temporal flow characteristics, despite their potential for capturing bursty DDoS behaviors, have not been systematically evaluated in federated one-class anomaly detection.

These challenges are coupled in practice. In realistic IoT deployments, attack labels are scarce or absent, traffic distributions differ sharply across operator domains, and raw flow records cannot be pooled for regulatory and bandwidth reasons. A practical detector must therefore meet three requirements simultaneously: learn from benign traffic alone, generalize across heterogeneous domains, and train without data exchange. One-class anomaly detection addresses the label scarcity, federated aggregation addresses the data-locality and cross-domain constraints, and temporal NetFlow descriptors supply the time-varying signal that distinguishes bursty DDoS flooding from benign fluctuation. We study these mechanisms on NF-CSE-CIC-IDS2018-v3, NF-BoT-IoT-v3, and NF-ToN-IoT-v3 [17]. These are the current standardized NetFlow benchmarks that expose the ten temporal descriptors absent from earlier v1 and v2 captures. They also span markedly different benign-to-attack ratios, which provides a controlled testbed for cross-domain federated evaluation.

To address these limitations, we propose FADIoT, a privacy-preserving federated anomaly detection framework that combines temporal NetFlow-v3 features, GANomaly-based one-class learning, and Fast Gradient Sign Method (FGSM)-based perturbation-aware training. FADIoT enables collaborative learning across multiple IoT domains without sharing raw traffic while improving optimization stability and cross-domain generalization.

This work investigates the following research questions:

- **RQ1:** Do NetFlow-v3 temporal features improve detection performance in federated anomaly detection?
- **RQ2:** Does FGSM-based perturbation-aware training keep local optimization stable during training?
- **RQ3:** How does federated multi-domain training compare with single-domain training for cross-domain IoT DDoS detection?

The main contributions of this paper are summarized as follows:

- We show that temporal NetFlow-v3 descriptors improve one-class IoT DDoS detection across three heterogeneous datasets. The F1 gains reach 0.28 on the benign-rich CICIDS2018 benchmark and 0.24 on ToN-IoT, with a small F1 change on the already-saturated, attack-dominated BoT-IoT.
- We show that federated multi-domain training produces one model that covers all three participating domains with usable F1, where single-domain models lose as much as 0.91 F1 off-domain, while keeping all raw traffic local.
- We introduce FGSM-based perturbation-aware training into federated GANomaly learning and characterize its effect on local optimization behavior, deliberately framed as a training-time property rather than a robustness guarantee.
- We integrate these components into FADIoT, a privacy-preserving framework that trains without attack labels and requires no centralized traffic collection, and report each component as a controlled study with the other components held fixed. A full factorial ablation of all three components is identified as the principal item for future work.

The remainder of this paper is organized as follows. Section II reviews centralized, adversarially robust, and federated intrusion detection and positions FADIoT against the three identified gaps. Section III states the threat model and details the four-stage pipeline: temporal NetFlow-v3 features, one-class GANomaly training, FGSM perturbation-aware local updates, and Federated Averaging (FedAvg) aggregation. Section IV reports the three component studies (RQ1–RQ3) and the cross-domain analysis. Section V delimits the scope and validity of the evaluation, and Section VI concludes.

II. RELATED WORK

Work on IoT DDoS detection has advanced along three largely separate tracks: centralized intrusion detection, adversarially robust learning, and federated intrusion detection. Each addresses one facet of the deployment problem, but their assumptions rarely hold simultaneously. We review them in turn and locate FADIoT at their intersection.

A. IoT Intrusion Detection Systems

Machine learning and deep learning have improved IoT intrusion detection by automating feature extraction and anomaly detection [3], [4]. Recent approaches based on convolutional neural networks (CNNs), recurrent networks, transformers, and generative adversarial networks (GANs) have demonstrated strong performance on benchmark datasets [5]–[7]. However, most existing intrusion detection system (IDS) frameworks assume centralized access to network traffic, creating privacy, scalability, and deployment challenges in distributed IoT environments.

B. Adversarially Robust Learning

Deep learning-based IDS models are vulnerable to adversarial perturbations generated by methods such as FGSM [9],

projected gradient descent (PGD), and Carlini and Wagner (C&W) attacks [8]. Adversarial training, which exposes models to such perturbations during optimization, has been widely adopted to improve robustness and optimization stability [10]. Nevertheless, most existing approaches assume centralized training and have seen limited adoption in federated anomaly detection settings, particularly for generative models deployed in resource-constrained IoT environments [11], [12].

C. Federated Intrusion Detection

Federated learning (FL) enables collaborative model training without sharing raw traffic data and has emerged as a promising paradigm for privacy-preserving IDS [13]. Existing frameworks such as FL-DAD and Anomaly-Flow demonstrate improved cross-domain generalization and reduced data exposure [14], [15]. Anomaly-Flow in particular performs federated generative DDoS detection across domains, but it does not incorporate perturbation-aware training or the temporal NetFlow-v3 descriptors, and no prior federated GAN approach combines these elements within a single one-class framework.

In summary, three gaps remain in the literature: (1) most IDS frameworks rely on centralized data collection; (2) perturbation-aware training is rarely explored in federated generative anomaly detection; and (3) temporal NetFlow descriptors remain underutilized in one-class IoT DDoS detection. FADIoT integrates temporal NetFlow-v3 features, federated GANomaly learning, and FGSM-based perturbation-aware training within a single privacy-preserving framework, and evaluates each as a controlled component study.

III. METHODOLOGY

FADIoT is a federated anomaly detection framework with perturbation-aware local training. Built around GANomaly [16], it enables collaborative training across IoT domains without exposing raw traffic. The pipeline consists of four steps: (1) temporal feature extraction from NetFlow-v3 [17]; (2) local one-class training on benign samples with FGSM perturbation injection; (3) weight upload and FedAvg aggregation; (4) cross-domain evaluation.

A. Threat Model

The aggregation server is honest-but-curious: it follows FedAvg but may inspect the updates it receives, which leaves them open to gradient inversion [22]. The privacy guarantee is architectural: the server never observes raw traffic. Malicious servers, model-poisoning clients, and Sybil attacks lie outside this scope. The FGSM perturbations applied during training ($\epsilon \in \{0.1, 0.2, 0.3\}$) are an optimization mechanism acting on benign inputs at training time. The framework makes no claim of robustness against adaptive or stronger test-time adversaries, and Section V delineates this boundary.

B. Datasets and Setup

We use NF-CSE-CIC-IDS2018-v3, NF-ToN-IoT-v3, and NF-BoT-IoT-v3 [17], each extending its v2 version with ten temporal descriptors (flow duration, inter-arrival time, burst

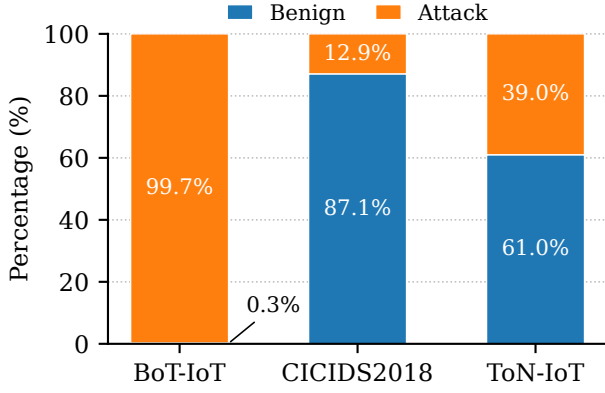


Fig. 1. Distribution of benign and attack samples across datasets, highlighting class imbalance.

metrics). Datasets were preprocessed by removing IP identifiers, one-hot encoding protocols, and z-score normalization. Training (80%), validation (2%), and testing (18%) splits were used, with benign-only training. Fig. 1 shows the benign-to-attack distribution across the three datasets, highlighting the severe and heterogeneous class imbalance.

Federated training involves three silos (CICIDS2018, ToN-IoT, BoT-IoT), each training GANomaly for 50 local epochs per round, with FedAvg aggregation over 10 rounds. FGSM perturbation-aware training uses $\epsilon \in \{0.1, 0.2, 0.3\}$ and a 1:1 clean-to-perturbed batch ratio. Baselines include single-domain GANomaly (non-FL) and supervised classifiers: Random Forest (RF), Logistic Regression (LR), XGBoost, and a multilayer perceptron (MLP).

C. Metrics

We report the area under the receiver operating characteristic curve (AUROC) and F1-score. Under the extreme class imbalance of these datasets, F1 is the primary operating-point metric and AUROC summarizes ranking quality across thresholds. Training uses benign flows only; the detection threshold is selected with the Youden index on a small held-out validation split that contains both benign and attack samples.

D. Local One-Class Training with GANomaly

Fig. 2 illustrates the GANomaly architecture. FADIoT adopts benign-only training: labeled attacks are often unavailable in real deployments, and one-class learning avoids overfitting to domain-specific attack signatures, supporting cross-domain generalization.

Let $G = E \circ Dec$ denote the generator (encoder E + decoder Dec), and D the discriminator. Given input x , the encoder produces latent code $z = E(x)$; the generator reconstructs $\hat{x} = G(x)$, and a second encoder re-embeds the reconstruction as $\hat{z} = E(\hat{x})$. The training minimizes:

$$\mathcal{L} = \lambda_{adv} \mathcal{L}_{adv} + \lambda_{rec} \|x - \hat{x}\|_1 + \lambda_{enc} \|z - \hat{z}\|_2^2.$$

At inference, the anomaly score is the latent discrepancy:

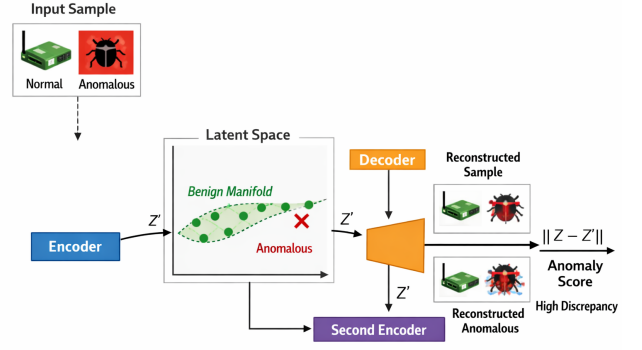


Fig. 2. GANomaly architecture and detection principle.

$$s(x) = \|z - \hat{z}\|_2.$$

A threshold τ (selected via the Youden index on a small held-out validation split with both benign and attack samples) flags x as anomalous if $s(x) > \tau$.

E. Perturbation-Aware Training at the Input Layer

To improve optimization stability, we inject FGSM perturbations during local training. For benign input x and loss $\mathcal{L}$:

$$x' = x + \epsilon \text{sign}(\nabla_x \mathcal{L}), \quad \epsilon \in \{0.1, 0.2, 0.3\}.$$

Clean and perturbed samples are mixed at a 1:1 ratio per mini-batch to stabilize the learned decision boundary. FGSM is chosen over PGD because it requires a single gradient step per batch rather than several; training time, memory, and energy overhead are not measured here.

F. Federated Optimization

Fig. 3 illustrates the FADIoT pipeline. Clients train locally for E epochs and upload weights for FedAvg aggregation [18]:

$$w^{(t+1)} = \sum_{k=1}^K \frac{n_k}{\sum_j n_j} w_k^{(t)},$$

where $w_k^{(t)}$ is client k 's model at round t , and n_k is its benign sample count. Only model parameters are exchanged; raw flows never leave any client. All three silos participate in every round.

G. Evaluation Protocol

We evaluate two settings: (S1) single-domain training and testing within the same dataset, and (S2) cross-dataset generalization, training on one dataset and testing on the disjoint datasets. Federated training uses three silos, one per dataset, aggregated with FedAvg over 10 rounds with all silos participating each round. We compare FL-GANomaly against single-domain GANomaly (non-FL, per-domain models) and supervised classifiers (RF, LR, XGBoost, MLP) trained on labeled source-domain data, where labels are available. Primary

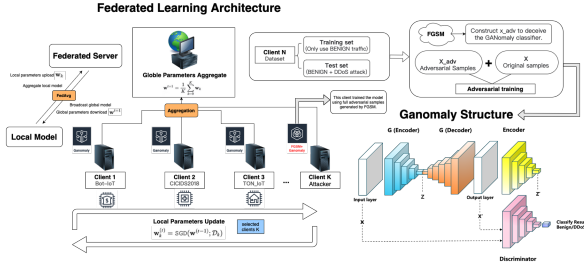


Fig. 3. Federated optimization workflow: clients train locally and upload weights; the server aggregates via FedAvg and broadcasts the global model.

TABLE I
PERFORMANCE COMPARISON OF v2 VS. v3 UNDER FL-GANOMALY.

Version	CICIDS2018		ToN-IoT		BoT-IoT	
	AUROC	F1	AUROC	F1	AUROC	F1
v2	0.926	0.493	0.784	0.482	0.994	0.968
v3	0.955	0.776	0.894	0.718	0.927	0.998

metrics are F1 and AUROC. AUROC trajectories over local training are reported for FGSM $\epsilon \in \{0.1, 0.2, 0.3\}$ under clean and matched- ϵ perturbed evaluation.

IV. RESULTS

We evaluate each component of FADIoT with the others held fixed: temporal features (RQ1), perturbation-aware training (RQ2), and federated training (RQ3).

A. Temporal Feature Impact (RQ1)

Table I compares FL-GANomaly on v2 vs. v3 datasets. The v3 temporal descriptors yield clear gains: on CICIDS2018, F1 rises from 0.493 to 0.776 and AUROC from 0.926 to 0.955; on ToN-IoT, F1 improves by 0.24 and AUROC by 0.11. These results indicate that temporal features capture DDoS traffic dynamics that the static v2 representation misses.

The v3 descriptors add flow duration, inter-arrival, and burst statistics that the v2 representation does not carry. The gains coincide with their introduction: a 57% relative F1 gain on CICIDS2018, and AUROC 0.894 with F1 0.718 on ToN-IoT. Because v2 and v3 feature sets are exchanged as a block, this design does not attribute the gains to individual descriptors.

BoT-IoT AUROC Regression. Despite general improvement, BoT-IoT shows an AUROC drop (0.994→0.927) while F1 rises (0.968→0.998). Benign flows are 0.3% of BoT-IoT, so the ROC estimate rests on a small benign sample. AUROC is threshold-independent, so the decline is not accounted for by the operating point, and the fixed-seed protocol used here cannot establish whether it is systematic. We report it as an open observation.

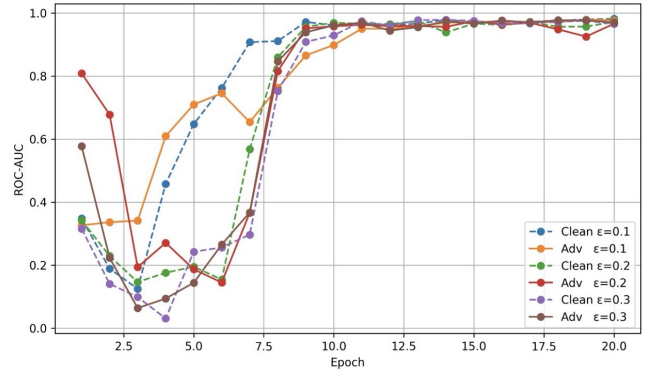


Fig. 4. AUROC over local training epochs for FGSM-hardened models at $\epsilon = 0.1, 0.2, 0.3$, evaluated on clean inputs and on FGSM-perturbed inputs at the matched ϵ .

Takeaway – RQ1 Temporal enrichment gives the largest single-component effect measured here: the ten NetFlow-v3 descriptors raise F1 by 0.24–0.28 over v2 on the benign-rich datasets, consistent with burst and inter-arrival structure carrying separability that static flow averages do not.

B. Perturbation-Aware Training Dynamics (RQ2)

Fig. 4 tracks AUROC over local training for FGSM-hardened models at $\epsilon = 0.1, 0.2, 0.3$, evaluated both on clean inputs and on FGSM-perturbed inputs at the matched ϵ . The two evaluations differ sharply over the first epochs, converge by roughly epoch 10, and then stay flat and near-saturated for the remainder of the plotted range at every ϵ .

The perturbed evaluation uses the same single-step attack and magnitude seen during training. Fig. 4 therefore characterizes local optimization behavior under the training perturbation budget and does not evaluate adaptive or stronger adversaries.

Takeaway – RQ2 Under FGSM-hardened training, clean and matched- ϵ perturbed evaluation converge by roughly epoch 10 and remain stable thereafter at every ϵ examined, so the perturbation budget does not destabilize local optimization.

C. Cross-Domain Generalization (RQ3)

This comparison holds the GANomaly backbone fixed and varies whether training is federated across the three silos or confined to a single domain. Federated training also sees three domains rather than one, so aggregation and broader domain coverage are not separated here. Table II reports the result. The federated model retains usable F1 across all three participating domains: 0.776 on CICIDS2018, 0.998 on BoT-IoT, and 0.718 on ToN-IoT. Single-domain models degrade off-domain by as much as 0.91 F1: the BoT-IoT model falls from 0.995 in-domain to 0.081 on CICIDS2018, and the CICIDS2018 model from 0.721 to 0.498 on ToN-IoT.

Discussion. By aggregating non-independent and identically distributed (non-IID) statistics from all three silos, federated training yields one model covering all three NetFlow-v3 domains without moving raw data. For space, Table II reports

TABLE II
CROSS-DOMAIN EVALUATION: FEDERATED LEARNING (FL) VS.
SINGLE-DOMAIN GANomaly.

Model	Evaluation dataset	AUROC	F1-score
GANomaly (CICIDS2018)	CICIDS2018	0.955	0.721
	ToN-IoT	0.684	0.498
GANomaly (BoT-IoT)	BoT-IoT	0.931	0.995
	CICIDS2018	0.452	0.081
GANomaly (ToN-IoT)	ToN-IoT	0.894	0.703
	CICIDS2018	0.782	0.239
FL-GANomaly (v3)	CICIDS2018	0.955	0.776
	BoT-IoT	0.927	0.998
	ToN-IoT	0.894	0.718

one off-domain target per single-domain source rather than the full source–target matrix.

Takeaway – RQ3 FL-GANomaly retains usable F1 (0.776, 0.998, 0.718) across all three participating domains while single-domain models fall off-domain (0.995 to 0.081), with no raw flow leaving its silo. The federated backbone is identical to the single-domain one, but federated training also covers three domains, so this comparison does not separate aggregation from domain coverage.

D. Cross-Domain DDoS Detection Performance Analysis

This comparison weighs attack-specific supervised learning against label-free anomaly detection under domain shift. Supervised models trained on labeled source-domain attacks overfit to domain-specific signatures and degrade on unseen domains. The one-class GANomaly model learns only benign traffic and avoids that signature overfitting, though its absolute cross-domain scores are lower.

Under the held-out cross-dataset setting (train on one dataset, test on the disjoint datasets and average), FL-GANomaly (v3) attains F1 0.182 against 0.060–0.176 for the supervised models (RF, LR, XGBoost, MLP). In this matched setting the federated model, trained without attack labels, is comparable to the supervised baselines, though all methods score low in absolute terms under maximum domain shift. For reference, FL-GANomaly’s average in-distribution per-client F1 is 0.831 (the mean of 0.776, 0.998, and 0.718 in Table II), a different operating regime reported in text rather than plotted, so Fig. 5 shows the matched cross-dataset comparison alone. Federated GANomaly trained without attack labels therefore remains a meaningful baseline in this regime.

E. Summary

Across the three component studies, temporal NetFlow-v3 enrichment raises detection accuracy on the benign-rich benchmarks, FGSM perturbation-aware training leaves local optimization stable across the perturbation magnitudes examined, and federated multi-domain training produces one model that covers all participating domains while keeping raw traffic local, within the scope set out in Section V.

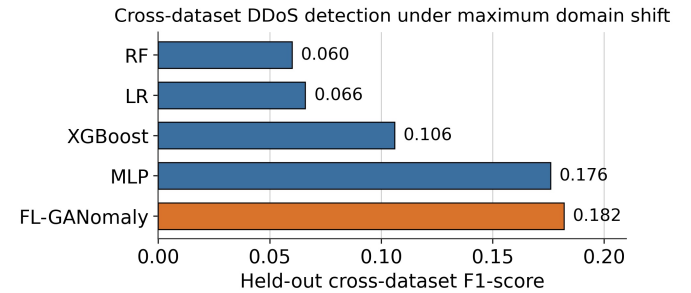


Fig. 5. Average F1-score under the held-out cross-dataset setting, training on one dataset and averaging over the disjoint held-out datasets. FL-GANomaly (v3) attains 0.182 against supervised baselines (RF, LR, XGBoost, MLP) at 0.060–0.176, so the federated model, trained without attack labels, is comparable to the supervised classifiers, though all methods score low in absolute terms. The in-distribution per-client mean (0.831) indexes a different regime and is reported in the text, not plotted here.

V. SCOPE AND VALIDITY

Several boundaries of this evaluation are deliberate, chosen to keep each component’s effect interpretable. We state them here so the claims are read at their intended scope, and this section collects every caveat in one place.

A. Internal Validity

RQ1 and RQ3 are controlled single-component comparisons, each holding FGSM out so that temporal features and federated aggregation are not confounded with one another or with perturbation-aware training. The combined three-way ablation is left to future work. RQ2 reports FGSM-hardened runs only; no clean-trained control was run, so no comparison of convergence speed against clean training is made. Results come from fixed-seed runs, so we report consistent rankings and effect magnitudes across the configurations examined and not significance intervals. A multi-seed variance and significance analysis is the natural complement to the full ablation.

B. Construct Validity and Claim Scope

The FGSM evidence covers local optimization behavior and evaluation under the same single-step attack and magnitude used during training; it supports no conclusion about adaptive or stronger adversaries, which a held-out adversarial evaluation would settle. FADIoT trains the detector without attack labels, but the operating threshold is calibrated on a small held-out validation split that includes attack samples, so the label-free property applies to detector training rather than to threshold selection; fully unsupervised thresholding, for example a benign-score quantile or extreme-value calibration, is a direct next step. The privacy guarantee is architectural: the server observes only aggregated updates, which matches the standard federated threat model. Without differential privacy or secure aggregation [21], FedAvg updates remain exposed to gradient inversion under a curious server [22], so stronger privacy mechanisms are a direct next step.

C. External Validity

Communication cost, latency, and memory footprint sit outside the present evaluation, and the federated setting spans three silos and not a large client population. These dimensions govern behavior at deployment scale and are taken up as future work in the conclusion.

VI. CONCLUSION

FADIoT shows that reliable IoT DDoS detection is achievable without centralized data collection or attack labels for detector training. The framework combines NetFlow-v3 temporal feature enrichment, GANomaly-based one-class learning, and FGSM-based perturbation-aware training within a federated architecture. Temporal enrichment is the dominant source of detection gain, and federated multi-domain training produces a single model that retains usable F1 across all participating domains, where single-domain models degrade once applied off-domain. All raw traffic stays within client boundaries, in line with data-minimization principles for real IoT deployments. Four directions follow from this work: a full factorial ablation paired with multi-seed variance analysis; stronger privacy through differential privacy or secure aggregation; a held-out adversarial evaluation of FGSM-trained models against adaptive attackers; and a scale-out study under many-client, strongly non-IID partitions with robust aggregators such as FedProx [19] or SCAFFOLD [20].

ACKNOWLEDGMENT

The authors used Anthropic's Claude (Claude Opus 4.8) for editorial assistance, including restructuring, consistency editing, and refinement of wording across the manuscript. The authors thank Feng Xiong for his contributions during the initial phase of the project. All technical content, experimental design, results, and conclusions are the authors' own. The authors have verified all text and references and take full responsibility for the accuracy and integrity of the paper.

REFERENCES

- [1] A. Heidari and M. A. Jabrael Jamali, "Internet of things intrusion detection systems: a comprehensive review and future directions," *Cluster Computing*, vol. 26, pp. 3753–3780, Dec. 2023.
- [2] A. Aldweesh, A. Derhab, and A. Z. Emam, "Deep learning approaches for anomaly-based intrusion detection systems: a survey, taxonomy, and open issues," *Knowledge-Based Systems*, vol. 189, p. 105124, Feb. 2020.
- [3] S. Hizal, A. F. M. S. Akhter, U. Cavusoglu, and D. Akgun, "Blockchain-based IoT security solutions for IDS research centers," *Internet of Things*, vol. 27, p. 101307, Oct. 2024.
- [4] M. Abbasi, A. Shahraki, and A. Taherkordi, "Deep learning for network traffic monitoring and analysis (NTMA): a survey," *Computer Communications*, vol. 170, pp. 19–41, Mar. 2021.
- [5] I. A. Fares and M. Abd Elaziz, "Explainable TabNet transformer-based on Google Vizier optimizer for anomaly intrusion detection system," *Knowledge-Based Systems*, vol. 316, p. 113351, May 2025.
- [6] J. Gao, Y. Lu, Y. He, M. Fan, D. Han, and Y. Qiao, "Tokenization representation and deep-learning-based intrusion detection in internet of vehicles," *IEEE Internet of Things Journal*, vol. 11, no. 23, pp. 37974–37987, Dec. 2024.
- [7] M. G. Constantin, D. C. Stanciu, L. D. Stefan, M. Dogariu, D. Mihăilescu, G. Ciobanu, M. Bergeron, W. Liu, K. Belov, O. Radu, and B. Ionescu, "Exploring generative adversarial networks for augmenting network intrusion detection tasks," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 21, no. 1, pp. 16:1–16:19, Dec. 2024.
- [8] C. Zhang, S. Yu, Z. Tian, and J. J. Q. Yu, "Generative adversarial networks: a survey on attack and defense perspective," *ACM Computing Surveys*, vol. 56, no. 4, pp. 91:1–91:35, Dec. 2023.
- [9] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2015.
- [10] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2018.
- [11] Y. Luo, T. Zhu, Z. Liu, T. Mao, Z. Chen, H. Pi, and Y. Lin, "GANFAT: robust federated adversarial learning with label distribution skew," *Future Generation Computer Systems*, vol. 160, pp. 711–723, Nov. 2024.
- [12] S. Utomo, A. Rouniyar, H.-C. Hsu, and P.-A. Hsiung, "Federated adversarial training strategies for achieving privacy and security in sustainable smart city applications," *Future Internet*, vol. 15, no. 11, p. 371, Nov. 2023.
- [13] T. A. Ahanger, A. Aldaej, M. Atiquzzaman, I. Ullah, and M. Yousufudin, "Federated learning-inspired technique for attack classification in IoT networks," *Mathematics*, vol. 10, no. 12, p. 2141, June 2022.
- [14] Y. Alhasawi and S. Alghamdi, "Federated learning for decentralized DDoS attack detection in IoT networks," *IEEE Access*, vol. 12, pp. 42357–42368, 2024.
- [15] L. H. de Melo, G. C. Bertoli, M. Nogueira, A. L. dos Santos, and L. A. Pereira Junior, "Anomaly-flow: a multi-domain federated generative adversarial network for distributed denial-of-service detection," *IEEE Network*, vol. 39, pp. 45–53, Mar. 2025.
- [16] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "GANomaly: semi-supervised anomaly detection via adversarial training," in *Proc. Asian Conf. Computer Vision (ACCV)*, 2018, pp. 622–637.
- [17] M. Luay, S. Layeghy, S. Hosseinnoorbin, M. Sarhan, N. Moustafa, and M. Portmann, "Temporal analysis of NetFlow datasets for network intrusion detection systems," *arXiv preprint arXiv:2503.04404*, Mar. 2025.
- [18] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agueray Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, 2017, pp. 1273–1282.
- [19] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Machine Learning and Systems (MLSys)*, vol. 2, 2020.
- [20] S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, and A. T. Suresh, "SCAFFOLD: stochastic controlled averaging for federated learning," in *Proc. 37th Int. Conf. Machine Learning (ICML)*, 2020, pp. 5132–5143.
- [21] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *Proc. ACM SIGSAC Conf. Computer and Communications Security (CCS)*, 2017, pp. 1175–1191.
- [22] J. Geiping, H. Bauermeister, H. Droege, and M. Moeller, "Inverting gradients – how easy is it to break privacy in federated learning?," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 16937–16947.