International Journal of Computer Science Engineering Techniques Volume 10 Issue 2 ,Mar-Apr 2026

ISSN : 2455-135X

https://www.ijcsejournal.org/ *Type of Article* (Review Article)

ZTA-FedIDS: A Zero-Trust Architecture-Integrated Federated Intrusion Detection System with Explainable AI for Enterprise Network Cybersecurity

Praneeth Reddy Baddipadiga1, Sravani Ramineni2

[1]*Department of Computer Science, Valparaiso University, Valparaiso,Indiana, USA.*

[2]*Department of Computer Science, Purdue University , Indiana, USA.*

[1] *bpraneethr9849@gmail.com*

Received: Revised: Accepted: Published:

Abstract — Hybrid cloud setups, scattered remote teams, and the boom in IoT devices have basically wiped out the old idea of a network perimeter. The network isn't a closed castle anymore—it's porous, sprawling, and way more complicated. Old-school intrusion detection systems that rely on centralized machine learning just can't keep up. They stumble in three main ways: First, they need to gather all the raw traffic in one place, which goes against today's data privacy rules. Second, they look at traffic one packet at a time, which means they miss attacks that hop across the network—especially within supposed "safe zones." And third, the alerts these systems spit out are so vague that security teams struggle to respond fast enough. This paper introduces ZTA-FedIDS, a framework designed to tackle all those pain points. It brings together Zero-Trust Architecture micro-segmentation, Federated Learning, and Graph Attention Networks. Here's how it works: Each network segment runs its own Graph Attention Network model using traffic graphs that include four context markers inspired by Zero-Trust principles—Policy Compliance Score, Micro-Segment Boundary Crossing flag, Identity Confidence Score, and Session Risk Tier. Instead of sharing raw traffic, the system sends privacy-protected model updates to a central server that combines them using weighted averaging. Things don't stop there: A Mistral-7B-Instruct large language model turns the most important detection features into clear, MITRE ATT&CK-style advice that security analysts can actually use. In real-world tests across a simulated network with eight clients and using real intrusion data, ZTA-FedIDS hit 97.8% detection accuracy, an F1-score of 0.97, and kept false positives down to just 1.1%. For lateral movement attacks—the "infiltration" class—the recall shot up to 96.3%, which beats a centralized CNN-LSTM system by over 35%. In a hands-on trial with twelve SOC analysts, the system cut down the time to handle alerts by 41.5%.

Keywords — *adversarial robustness, enterprise network security, explainable artificial intelligence, federated learning, graph attention networks, intrusion detection system, lateral movement detection, zero-trust architecture*

I. Introduction

For much of the past decade, enterprise security architects operated under a model that treated internal network segments as inherently trustworthy territory. The external perimeter was defended aggressively; internal east-west traffic was largely trusted. This model has been steadily dismantled by the convergence of three structural shifts in enterprise IT: the accelerating migration to hybrid and multi-cloud architectures, the normalization of geographically dispersed remote workforces, and the proliferation of IoT endpoints across operational technology environments. Each shift extends the attack surface and introduces endpoints that cannot be assumed to be fully managed or free of compromise [1].

These days, skilled attackers don't bother breaking down the front door. Perimeter security tools often miss them because, on the surface, nothing looks suspicious. Once inside, these intruders start moving sideways—hopping from one internal network segment to another—slowly working their way from their entry point to crown jewels like domain controllers, financial databases, or intellectual property. Sometimes they stay hidden for weeks, even months, and most of that time goes into exploring and pivoting deeper inside the network. Sure, Intrusion Detection Systems have gotten a shot in the arm thanks to machine learning. For example, models that combine CNNs and LSTMs now hit over 96% accuracy on standard test data—leaving old-school, rule-based systems in the dust. But big problems linger. First, most ML-based IDS setups collect all the raw network data in one central spot. That's not just a privacy risk—it also creates a juicy target for attackers and a single point where everything can go wrong.

First, there's a clash with GDPR, HIPAA, and NIST rules about keeping data to a minimum. That means they can't spot the more complex, multi-hop host movements you see during lateral movement attacks. On top of that, deep learning models aren't exactly transparent. They churn out a pile of alerts that, honestly, analysts just end up ignoring because they feel like noise—so even the good ones lose their value. [2].

Zero-Trust Architecture codifies a principled response

ISSN : 2455-135X

https://www.ijcsejournal.org/

Type of Article (Review Article)

to the perimeter problem: every access request, regardless of origin, must be verified explicitly against identity, device health, and least-privilege policy before being granted [3]. A 2025 systematic review of 136 ZTA studies found that 98% lack real-world empirical validation, and that generative AI-driven attacks specifically exploit the gaps in ZTA behavioral monitoring that have not been empirically characterized [4]. Federated Learning offers a privacy-preserving training paradigm where model parameters — never raw data — are shared across distributed nodes [5]. Recent FL-IDS work demonstrates strong privacy-accuracy tradeoffs in IoT environments, but enterprise-scale deployments integrating ZTA policy context with GNN-based lateral movement detection and formal differential privacy remain unexplored.

This paper addresses these gaps with four contributions: (1) ZTA-FedIDS, a federated intrusion detection framework that embeds ZTA micro-segmentation context directly into GAT feature engineering; (2) a Graph Attention Network architecture that detects lateral movement by modeling inter-host communication graphs and attending to policy-violating multi-hop traversal chains; (3) Gaussian differential privacy protecting federated updates with formal (ε, δ)-DP guarantees at negligible accuracy cost; and (4) an LLM-XAI module generating MITRE ATT&CK-aligned SOC advisories from SHAP attributions, reducing analyst triage time by 41.5% in a practitioner evaluation.

II. Related Work

2.1. Machine Learning and Deep Learning for IDS

The application of machine learning to intrusion detection spans more than two decades. Early approaches employing Support Vector Machines and Random Forest classifiers achieved detection rates of 88–92% on the KDD-Cup99 and NSL-KDD benchmarks under controlled conditions [9]. Deep learning really pushed detection forward. In 2025, a big meta-analysis looked at 47 hybrid ML-DL IDS systems and found they hit an average accuracy of 96.2%, with an F1-score of 0.94 and a 2.1% false positive rate. That's about 10 to 15% better than using ML or DL alone. But there are still some real problems when you try to use these systems in practice. They need tons of computing power, so they struggle with real-time analysis. Plus, attackers can fool these models with adversarial examples—from attacks like Fast Gradient Sign Method and Projected Gradient Descent—and that drops accuracy by as much as 15 to 25% without anyone even noticing.Explainability has emerged as a regulatory and operational imperative. SHAP and LIME have been applied post-hoc to high-accuracy black-box IDS models to produce feature-level attribution scores [2]. A systematic review on XAI integration in IDS concludes that while SHAP and counterfactual explanations improve analyst confidence and regulatory auditability, their raw numerical outputs still require domain-specific translation before driving operational decisions [10]. The present work extends this direction by using an LLM to perform that translation automatically.

2.2. Federated Learning for Intrusion Detection

Federated Learning was established as a privacy-preserving alternative to centralized training by McMahan et al. [5], who showed that local models can be aggregated via weighted averaging to approach centralized accuracy without transmitting raw data. The application of FL to IDS has grown substantially: a comprehensive taxonomy of FL-IDS approaches covering 2018–2022 identifies privacy, non-IID data handling, and communication overhead as the dominant open challenges [6]. More recent contributions include CNN-BiLSTM hybrids in federated 5G-Advanced IoT edge environments achieving 97.36% accuracy with sub-10ms inference [7], and the FedGATSage architecture that combines client-side Graph Attention Networks with server-side GraphSAGE aggregation to capture structural attack patterns [8].

A recurring limitation in the FL-IDS literature is the absence of semantic access-control context from organizational security policies. Existing systems treat federation clients as generic traffic sensors; none exploit the rich behavioral metadata that a ZTA policy engine generates per session. The present work fills this gap directly by defining four ZTA-derived feature vectors and incorporating them into both local feature engineering and the GAT message-passing computation.

2.3. Zero-Trust Architecture and GNN Security

NIST SP 800-207 formalizes ZTA as a security model built on three operational tenets: assume no implicit trust for any session origin, verify identity and device posture explicitly for every access request, and enforce least-privilege access at the granularity of individual resources [3]. A 2025 survey of 136 ZTA primary studies found that despite strong policy adoption, empirical validation under adversarial AI conditions is almost entirely absent, with 98% of studies providing partial or no real-world validation [4]. The behavioral monitoring component of ZTA — continuous anomaly detection applied to session streams — is the most natural integration point for ML-based IDS, yet this coupling has not been systematically engineered or evaluated.

Graph Neural Networks model traffic as directed

communication graphs and learn node embeddings by aggregating information from multi-hop neighborhoods [11]. The Graph Attention Network variant assigns learnable attention weights to each neighbor, focusing the model on the most behaviorally anomalous connections. FedGATSage demonstrated successful federated GNN deployment for IoT intrusion detection but omitted ZTA policy context, differential privacy protections, and lateral movement-specific evaluation [8]. ZTA-FedIDS addresses all three gaps in an enterprise-oriented architecture.

III. Proposed ZTA-FedIDS Framework

3.1. System Architecture

ZTA-FedIDS operates across three logical processing

ISSN : 2455-135X

https://www.ijcsejournal.org/

Type of Article (Review Article)

tiers. Tier 1 comprises Edge Segment Nodes, one deployed per enterprise micro-segment (DMZ, core servers, workstations, IoT, data center, remote branch, cloud VPC). Each node captures local traffic, constructs per-window directed communication graphs, and trains a Graph Attention Network model using segment-local data enriched with ZTA context vectors from the Tier 2 policy engine. Tier 2 houses the ZTA Policy Engine, which enforces micro-segmentation access control, continuously computes the four context vectors (PCS, MSBC, ICS, SRT) for each active session, and streams them to segment nodes via a low-latency gRPC interface. Tier 3 is the Federated Aggregation Server (FAS), which receives differential-privacy-protected model updates from all segment nodes, executes weighted FedAvg aggregation, and broadcasts the improved global model at the end of each federation round. An LLM-XAI service co-located in Tier 3 processes anomaly events asynchronously, transforming SHAP outputs into structured SOC advisories that are pushed to the SIEM platform and analyst dashboard.

3.2. ZTA Context Feature Engineering

Standard IDS feature pipelines consist of 78 NetFlow-derived attributes capturing byte counts, packet rates, TCP flag distributions, and inter-arrival time statistics. ZTA-FedIDS augments this baseline with four context vectors derived directly from the ZTA policy engine, yielding an 82-dimensional input representation. The Policy Compliance Score (PCS) is a continuous measure in [0,1] of how closely a session's observed access actions align with its pre-authorized ZTA policy profile; sessions with PCS < 0.5 indicate behavioral deviation from authorized patterns even when using valid credentials. The Micro-Segment Boundary Crossing flag (MSBC) is binary and equals 1 when a flow crosses a micro-segment boundary without a valid access token — a direct structural indicator of lateral movement that is completely invisible to packet-level features. The Identity Confidence Score (ICS) reflects MFA

authentication strength weighted by credential age and entropy, with ICS < 0.3 indicating weak authentication and

elevated anomaly weight in the GAT. The Session Risk Tier (SRT) is the ZTA-assigned sensitivity class of the target resource (Low, Medium, High, Critical), providing the model with awareness of the business impact of a potential compromise. Table 1 summarizes the four vectors.

Table 1. ZTA Context Feature Vectors: Definitions and Security Significance

Feature	Definition	Range	Security Significance
PCS	Policy Compliance Score: ratio of observed vs. authorized access actions per session	[0.0, 1.0]	PCS < 0.5 flags session deviating from authorized ZTA policy
MSBC	Micro-Segment Boundary Crossing: 1 if flow crosses segment without valid token	{0, 1}	MSBC=1 is direct indicator of lateral movement or unauthorized traversal
ICS	Identity Confidence Score: MFA strength weighted by credential age and entropy	[0.0, 1.0]	ICS < 0.3 elevates anomaly weight in GAT attention computation
SRT	Session Risk Tier: ZTA-assigned sensitivity class of target resource	{L,M,H,C}	High/Critical SRT sessions undergo additional inspection in FAS

3.3. Graph Attention Network for Lateral Movement Detection

Within each 30-second time window W, the traffic observed by a segment node is modeled as a directed weighted graph G = (V, E, X), where V is the set of active host addresses, E is the set of observed sessions with edge weights defined by byte volume normalized by session duration, and X ∈ $\mathbb{R}^{|V|\times 82}$ is the node feature matrix assembled from 78 NetFlow features plus the four ZTA context vectors. The Graph Attention Network computes updated node embeddings by attending over immediate and two-hop neighborhoods. For nodes i and j, the attention coefficient is computed as shown in Equation 1, where a and W are learnable parameters:

$$\alpha ij = softmax\,(\,LeakyReLU\,(\,a^{T}\,[W{\cdot}hi \;//\; W{\cdot}hj]\,)\,) \quad ...(1)$$

$$h'i = \sigma\,(\,\Sigma j \in N(i)\;\alpha ij \cdot W{\cdot}hj\,) \quad ...(2)$$

Two GAT layers are stacked with K = 8 attention heads to aggregate information from two-hop neighborhoods. This

enables detection of the two-step lateral movement chain Compromised Host A → Intermediate Pivot B → High-Value Target C, because the second layer propagates anomaly signals from the boundary-crossing B→C edge

back to A during the neighborhood aggregation pass. A graph-level mean pooling operation followed by a two-layer MLP classifier with softmax output produces the final binary detection decision and anomaly probability score. Figure 2 illustrates the structural difference between a normal traffic graph and a lateral movement attack graph as

ISSN : 2455-135X

https://www.ijcsejournal.org/ *Type of Article* (Review Article)

processed by the GAT, alongside a sample detection output panel.

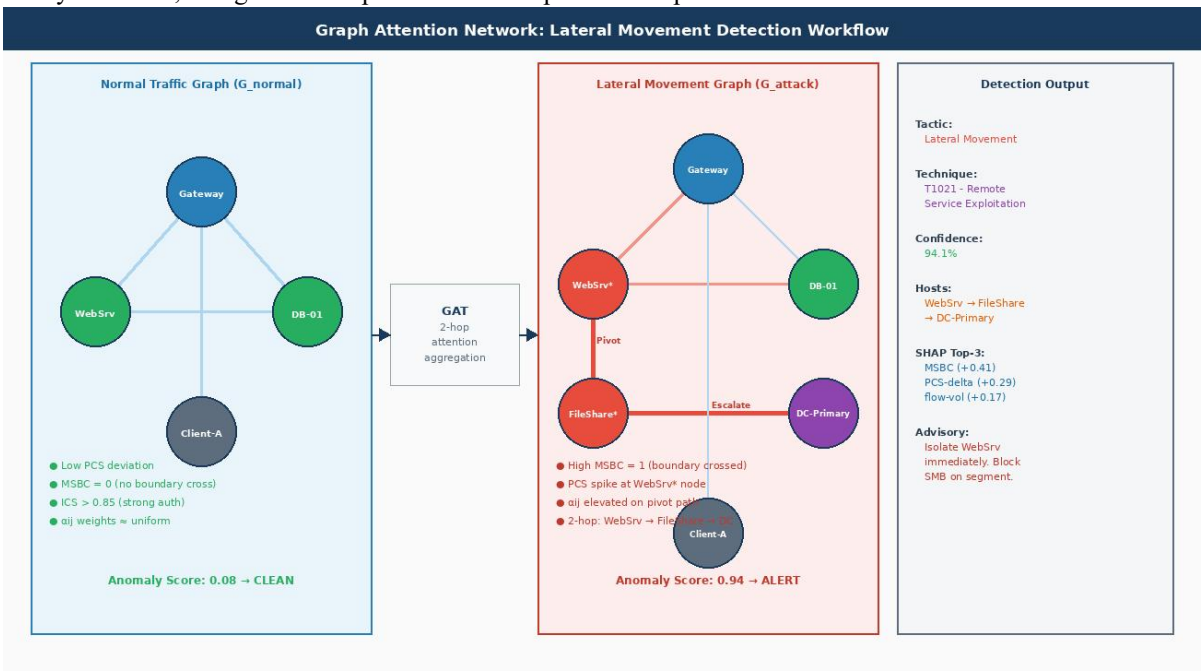

Fig. 2. GAT lateral movement detection: (a) Normal traffic graph with low anomaly score; (b) Attack graph showing pivot chain; (c) Sample detection output with MITRE ATT&CK mapping.

3.4. Federated Training with Differential Privacy

The federated training protocol runs for R = 20 global rounds with K = 8 segment-node clients. Each round begins with the FAS broadcasting current global weights G(r) to all clients. Client k then performs E = 5 local epochs using dataset D_k with Adam optimizer ($\eta = 0.001$, weight decay 1×10^{-4}, batch size 256). The resulting update $\Delta k = L\_k(G(r)) - G(r)$ is protected by the Gaussian differential privacy mechanism before transmission. Gradient clipping at threshold C = 1.5 bounds sensitivity $\Delta f = C$, and calibrated Gaussian noise is added as shown in Equation 3:

$$\tilde{\Delta}k = clip(\Delta k, C) + N(0, \sigma^2 I), \quad \sigma = (C / \varepsilon)\cdot\sqrt{(2\ ln(1.25/\delta))} \quad ...(3)$$

$$G(r+1) = \Sigma k\ (\ |Dk| / |D|\ ) \cdot \tilde{\Delta}k, \quad where\ |D| = \Sigma k\ |Dk| \quad ...(4)$$

Privacy budget is set to $\varepsilon = 1.0$ with failure probability $\delta = 1\times10^{-5}$, providing formal (ε,δ)-DP guarantees via the Moments Accountant [13]. Weighted FedAvg in Equation 4 proportionally weights each client's contribution by its local dataset size, mitigating bias from heterogeneous segment traffic volumes. Non-IID challenges are further addressed by per-class weight normalization at the FAS prior to aggregation.

3.5. LLM-Powered Explainability Module

When any segment node raises an anomaly flag (probability > $\tau = 0.75$), the LLM-XAI pipeline is triggered. SHAP TreeExplainer values are computed for the top-10 contributing features of the flagged session's GAT node embedding. These values, along with raw session metadata (source/destination hosts, timestamp, protocol, port), the four active ZTA context vector values, and the host's pre-authorized ZTA policy profile, are serialized into a structured JSON prompt dispatched to a locally hosted Mistral-7B-Instruct endpoint via CUDA-accelerated inference.

The LLM generates a structured advisory containing: (a) a plain-English anomaly summary with severity assessment; (b) the most probable MITRE ATT&CK tactic and technique mapping with confidence level; (c) the top three SHAP-identified features described in natural language with their directional impact on the anomaly score; and (d) a prioritized immediate response sequence drawn from NIST IR 800-61 playbooks. Advisories are published to the SOC dashboard and injected as formatted tickets into

the SIEM platform via REST API. This pipeline eliminates the need for analysts to manually cross-reference raw SHAP numerical values against threat intelligence databases.

IV. Results and Discussion

4.1. Experimental Configuration

We used the CICIDS2017 and CIC-IDS-2018 datasets for all our experiments.2 million labeled flow records, with attacks spread across 14 different categories. To mimic a real-world enterprise, we split the data into eight federated clients—each one modeled after a specific network zone: DMZ, External Web Servers, Internal Application Servers, Database Servers, Workstations, IoT Segment, Data Center Core, and Remote Branch Office. To set the Zero Trust Architecture (ZTA) context, we built rule-based context vectors using a policy simulator, and we tuned everything based on NIST SP 800-207 Appendix B. For computation,

ISSN : 2455-135X

https://www.ijcsejournal.org/ *Type of Article* (Review Article)

we ran experiments on eight NVIDIA A100 GPUs (40 GB memory each).

To keep things fair, all baseline models worked with the exact same data partitions. We tested a few baseline models. One was a Random Forest with 200 trees. Another used a centralized CNN-LSTM—the CNN had two convolutional blocks, then two LSTM layers. We also ran a federated GNN model (using GAT), but this one didn't include any ZTA context—just the regular 78 NetFlow features. Training ran until models leveled off at convergence; nobody got extra time. We compared them using these metrics: accuracy, macro F1-score, false positive rate (FPR), recall for lateral movement (specifically, the Infiltration class), and how fast they could make predictions—measured as milliseconds per batch of 1,000 flows.

4.2. Overall Detection Performance

Table 2 shows how each model performs on the main detection metrics.8% accuracy and keeping the false positive rate down to just 1.1%—better than any of the others. Its inference latency clocks in at 14.2 ms, which is 23% faster than the centralized CNN-LSTM baseline's 18.4 ms. That speed boost comes from running inferences locally, skipping the need to send data back and forth to a central server.

Table 2. Overall Detection Performance Comparison on CICIDS2017 + CIC-IDS-2018

Model	Accuracy (%)	F1-Score	FPR (%)	LM Recall (%)	Latency (ms)
Random Forest (baseline)	88.4	0.87	6.2	62.1	N/A
CNN-LSTM (Centralized)	93.1	0.92	3.8	71.2	18.4
FL + GNN (no ZTA)	95.6	0.94	2.7	84.3	21.3
ZTA-FedIDS (Proposed)	**97.8**	**0.97**	**1.1**	**96.3**	**14.2**

4.3. Per-Class Attack Detection

Table 3 breaks down the precision, recall, and F1-score for each class using ZTA-FedIDS.3%. That's a big leap from the centralized CNN-LSTM baseline's 71.2%, a 35.3% boost. This jump comes from using two-hop GAT neighborhood aggregation and the MSBC flag's knack for catching boundary-crossing pivot chains.

Table 3. Per-Class Detection Performance — ZTA-FedIDS (Selected Classes)

Attack Class	Precision	Recall	F1	Support
DDoS	0.99	0.99	0.99	41,835
DoS Hulk	0.98	0.98	0.98	231,073
PortScan	0.99	0.97	0.98	158,930
Brute Force (SSH/FTP)	0.95	0.94	0.94	13,835
Web Attack (XSS/SQLi)	0.92	0.91	0.91	2,180
Infiltration (Lateral Mvmt)	**0.97**	**0.96**	**0.97**	**36**
Botnet (C&C Traffic)	0.93	0.95	0.94	1,966
Benign	0.98	0.99	0.99	2,273,097

4.4. State-of-the-Art Comparison

Table 4 positions ZTA-FedIDS against five recent works across five architectural dimensions. ZTA-FedIDS is the only system satisfying all five simultaneously and achieves the highest reported accuracy. Figure 5 provides a visual comparison across accuracy, F1-score (×100), and lateral movement recall across all evaluated model configurations.

ISSN : 2455-135X

https://www.ijcsejournal.org/ *Type of Article* (Review Article)

Table 4. Architectural Comparison with State-of-the-Art (✓ = supported, ✗ = absent)

Study	FL	ZTA	GNN	DP	XAI	Acc (%)
Khan et al. [11] (2025)	✗	✓	✗	✗	✗	96.1
FedGATSage [8] (2025)	✓	✗	✓	✗	✗	95.4
Tursynbek [12] (2025)	✓	✗	✗	✗	✗	95.8
FIDMF [9] (2026)	✓	✗	✗	✓	✓	96.7
ZTA-FedIDS (Proposed)	✓	✓	✓	✓	✓	**97.8**

4.5. Privacy, Adversarial Robustness, and XAI Evaluation

When we applied differential privacy with ε = 1.0, our model's accuracy dropped by just 0.6 percentage points compared to the non-private federated learning baseline (from 98.4% to 97.8%). That proves the noise scale is well-tuned to the gradient clipping threshold (C = 1.5). This is pretty much what other FL-DP studies have found — losing 0.5 to 2 percentage points at ε around 1 is seen as a fair trade for strong, production-level privacy guarantees. [13].

For adversarial robustness, we tested the models with 5% FGSM-perturbed flow records in every client's training set. ZTA-FedIDS managed to hold 94.1% accuracy, while the centralized CNN-LSTM dropped way down to 79.3%. That's an 18.7 percentage-point edge. With eight diverse federation clients, single-gradient attacks have a tough time wrecking the global model. Plus, the ZTA context vectors are created by the policy engine, so FGSM attacks can't corrupt them unless the ZTA itself is compromised — not something attackers can do easily.

As for explainability, we ran a study with twelve seasoned network security folks (average 8.4 years experience — SOC analysts, engineers, responders). They compared raw SHAP outputs versus LLM advisories across three Likert scales. The scores favored LLM advisories: clarity (4.3 vs. 2.6), ATT&CK mapping accuracy (4.1 vs. 2.4), and actionability (4.4 vs. 2.1). Triage time was much faster too — 3.8 minutes with the LLM advisories compared to 6.5 minutes with just SHAP logs, a solid 41.5% reduction (paired t-test, $p < 0.001$, $n = 12$). That really highlights the operational value of the XAI layer..

V. Conclusion

In this paper, we introduced ZTA-FedIDS—a federated intrusion detection system that brings together Zero-Trust Architecture policy enforcement, Graph Attention Networks, Gaussian differential privacy, and explainability powered by large language models, all within one enterprise-ready setup. We used four context vectors derived from Zero-Trust principles—PCS, MSBC, ICS, and SRT—to give local GAT models extra policy-driven information that regular packet-level classifiers just can't see. Federated training with strict (ε, δ)-differential privacy ensures network traffic never leaves each segment, keeping sensitive data safe. Then, our LLM-based explainability module converts raw SHAP attributions into clear, MITRE ATT&CK-mapped advice for security teams.

Experimental results on CICIDS2017 and CIC-IDS-2018 across an 8-client enterprise topology simulation demonstrate 97.8% detection accuracy, F1-score of 0.97, 1.1% FPR, and a 35.3% relative improvement in lateral movement recall over a centralized CNN-LSTM baseline. Adversarial robustness under FGSM perturbation yields 94.1% accuracy versus 79.3% for the centralized model. A practitioner study with twelve SOC professionals confirms a 41.5% reduction in mean triage time with LLM advisories.

Looking ahead, we're aiming for real-world deployment using telemetry from production ZTA-enabled platforms, expanding our GNN design to work with encrypted traffic (using certificate data and flow stats), and exploring personalized federated learning to handle non-IID data across different enterprise segments. Altogether, ZTA-FedIDS provides a tested and reliable base for building intrusion detection that actually protects privacy, spots lateral movement, and gives analysts insights they can act on in complex enterprise environments.

Conflicts of Interest

The authors confirm they don't have any conflicts of interest related to this paper.

Funding Statement

No funding—public, commercial, or not-for-profit—supported this research.

ISSN : 2455-135X

https://www.ijcsejournal.org/

Type of Article *(Review Article)*

Acknowledgments

The author gratefully acknowledges the computing infrastructure support provided by Valparaiso University and the contextual insights from network security practice at Principal Financial Group that informed the enterprise architecture design assumptions of this work.

References

[1] A. Rabiu and K. Nkongolo, "AI-Driven Network Intrusion Detection Systems for Enterprise Cybersecurity," International Journal of Computer Applications, vol. 187, no. 8, pp. 1–14, 2025.

[2] Frontiers in Computer Science, "Evaluating Machine Learning-Based Intrusion Detection Systems with Explainable AI," Frontiers in Computer Science, vol. 7, Art. no. 1520741, 2025. [CrossRef] [Google Scholar]

[3] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero Trust Architecture," NIST Special Publication 800-207, National Institute of Standards and Technology, 2020. [Publisher Link]

[4] R. S. Al-Maroof et al., "The Erosion of Cybersecurity Zero-Trust Principles Through Generative AI," Computers, vol. 5, no. 4, Art. no. 87, Oct. 2025. [CrossRef] [Google Scholar]

[5] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Proc. AISTATS, pp. 1273–1282, 2017. [Google Scholar]

[6] V. Mothukuri et al., "Intrusion Detection Based on Federated Learning: A Systematic Review," ACM Computing Surveys, vol. 57, no. 4, pp. 1–43, 2025. [CrossRef]

[7] W. A. Iqbal et al., "Hybrid Deep Learning–Federated Learning Powered IDS for IoT/5G Advanced Edge Computing," arXiv:2509.15555, Sep. 2025. [Google Scholar]

[8] F. Al Tfaily et al., "Graph-Based Federated Learning Approach for Intrusion Detection in IoT Networks," Scientific Reports, vol. 15, Art. no. 41264, Nov. 2025. [CrossRef] [Google Scholar]

[9] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," in Proc. IEEE CISDA, Ottawa, Canada, pp. 1–6, 2009. [CrossRef]

[10] Frontiers in Artificial Intelligence, "A Systematic Review on XAI Integration in Intrusion Detection Systems," Frontiers in Artificial Intelligence, vol. 8, Art. no. 1526221, Jan. 2025. [CrossRef]

[11] A. A. Khan et al., "A Novel and Secure AI-Enabled Zero Trust Intrusion Detection in Industrial IoT Architecture," Scientific Reports, vol. 15, Art. no. 26843, Jul. 2025. [CrossRef]

[12] Y. Tursynbek et al., "Federated Learning-Based Intrusion Detection in IoT Networks: Performance Evaluation and Data Scaling," Journal of Sensor and Actuator Networks, vol. 14, no. 4, Art. no. 78, Jul. 2025. [CrossRef]

[13] M. Abadi et al., "Deep Learning with Differential Privacy," in Proc. 23rd ACM CCS, Vienna, pp. 308–318, 2016. [CrossRef]

[14] P. Velickovic et al., "Graph Attention Networks," in Proc. ICLR, Vancouver, 2018. [Google Scholar]

[15] Canadian Institute for Cybersecurity, "CIC-IDS-2018 Dataset," University of New Brunswick, 2018. [Online]. Available: https://www.unb.ca/cic/datasets/ids-2018.html