Journal of Distributed Computing and Systems(JDCS)
Vol.9, Issue 1, 2026, Pages 83-88

TransFeL-NIDS: Federated Latent Zero-Day Signatures using a Memory-Based Transformer for Network Intrusion Detection System

Hamidreza Yazdanpanah[1], S. Mojtaba Matinkhah[1], AliAkbar Nikoukar[2]
[1]Department of Computer Engineering, Yazd University, Yazd, IRAN
[2]Department of Computer Science, Yasouj University, Yasouj, IRAN

Article History:
Received: 25 January 2026
Received in revised form: 14 March 2026
Accepted: 14 May 2026
Available online: 15 August 2026

Abstract

Zero-day attacks are a critical challenge for Network Intrusion Detection (NIDS) due to the unknown structure of the attack with no predefined signature. Machine learning (ML) and Deep Learning (DL) models have recently emerged as a promising solution for enhancing IDS capabilities by detecting anomalies. On the other hand, Federated Learning has emerged as a groundbreaking paradigm for distributed applications to enhance privacy and bandwidth efficiency. The standard Federated Learning (FL) approach generally shares model parameters, which cannot effectively transfer knowledge about unseen attack behaviors across clients. This paper introduces TransFeL-NIDS, a novel federated zero-day detection method that allows clients to share the latent zero-day signatures extracted from a memory-based Transformer. The latent signatures encode high-level behavioral semantics without disclosing raw packets or sensitive traffic data. Compared to the standard FL approach, this strategy improves cross-client generalization to detect unseen attacks while preserving privacy and bandwidth efficiency. Additionally, the latent signature updates help mitigate the negative effects of imbalanced and non-IID data distributions in federated learning.

Keywords: Federated Learning, Edge AI, Transformer, Latent Signature, Network Intrusion Detection

I. INTRODUCTION

The World Economic Forum's (WEF) Global Risks Report puts cybersecurity in the current and future top ten risks globally[1]. By the same token, Gartner analysts predict that 45% of global organizations will somehow be impacted by a cyber-attack over the following years[2]. By increasing the advanced security threats and considering the complexity of computer networks and systems, the need for early detection of intrusion in time has increased. [3]. A detection system that accurately monitors a network for malicious activities or policy violations that compromise its confidentiality, integrity, or availability. Intrusion detection systems (IDS) have evolved to respond to these needs.

While traditional signature-based IDSs are effective against known existing threats, they are unable to detect zero-day exploits in time[4]. Hence, anomaly detection in network traffic has been given more attention as a complementary method to overcome the weakness of signature-based IDSs. Anomaly detection finds unknown intrusions (zero-day exploits) using analyzer models that can learn and detect deviations from normal behavior. Recently, machine learning (ML) and deep learning (DL) models have been widely used for intrusion detection due to their ability to learn complex patterns from high-dimensional data.

Meanwhile, Federated Learning[5] (FL) has emerged as a innovative approach for distributed applications to enhance privacy, security, and bandwidth efficiency. A decentralized approach that achieves the purpose of collaborative learning from a large amount of heterogeneous data that belongs to different clients without sharing their data[6]. In the FL approach, local training using raw data on edge devices is employed to ensure data privacy and security. In this approach, in contrast to traditional centralized approaches where data is sent to a central server for training, only the model updates are shared with the central server, ensuring data privacy and having less communication overhead.

Nevertheless, in the standard FL paradigm, learning is limited to past distributions, the global model is overly dependent on weight and bias updates that become unstable under skewed or heterogeneous datasets, and the global adaptation occurs slowly. Also, the parameter sharing mechanism fails to transfer meaningful knowledge about emerging or unknown attack behaviors across clients. An issue that becomes particularly severe in the non-IID and imbalanced conditions typical of intrusion detection environments. To address these limitations, this paper proposes TransFeL-NIDS, a novel federated zero-day detection method that allows clients to share the latent behavioral signatures (representing anomalous or zero-day activity) alongside the parameters. These latent signatures

(extracted from the hidden states of a memory-based Transformer model) encode high-level semantic representations of anomalous or previously unseen traffic patterns without exposing raw data. The proposed method uses similarity in latent representation space to detect zero-day attacks across distributed clients. This method improves cross-client generalization by enriching global updates with high-level semantic representations of anomalies to identify unknown attacks while preserving privacy and bandwidth efficiency.

II. LITERATURE REVIEW

For years, traditional signature-based IDS relied on a predefined database of attack signatures to identify malicious activity or policy violations. By increasing the advanced security threats and the complexity of networks and computer systems, signature-based techniques were unsuccessful in identifying unknown zero-day attacks. Hence, ML algorithms became prominent for detecting anomalies in network traffic by providing a better detection rate. In this regard, various supervised and unsupervised models such as k-Nearest Neighbors (KNN), Support Vector Machines (SVM), K-means, decision tree, and hybrid and ensemble models were proposed that make intrusion detection systems more effective[7]. In recent years, deep learning (DL) has achieved outstanding performance. DL-based IDSs have gained vast popularity due to the ability of deep feature learning to generate excellent results in detecting anomalies and malicious attacks[8]. Alongside, researchers try to extend the FL concept to the realm of intrusion detection systems, presenting a novel approach that leverages its decentralized architecture to address the challenges of intrusion detection methods in terms of privacy, security, and communication overhead. In this context, Li et al.[9] proposed DeepFed, a novel federated deep learning scheme that uses a convolutional neural network and a gated recurrent unit to detect cyber threats against industrial cyber-physical systems. The proposed method emphasizes privacy-preserving collaborative learning and improved detection versus localized training. In another work, Huang et al. [10] introduce a personalized FL architecture that separates execution and evaluation networks to allow local personalization while benefiting from global knowledge; targets heterogeneity and concept drift in cyber-physical systems domains, and proposes similarity-based aggregation mechanisms. In 2023, Edrissi et al. [11] proposed Fed-ANIDS, a method for distributed network intrusion detection that leverages the power of federated learning approach and autoencoder techniques. They compute an intrusion score based on the reconstruction error of normal traffic. Their findings show that autoencoder-based models outperform other generative adversarial network based models and achieve high detection accuracy coupled with fewer false alarms. Bhavsar et al. [12] demonstrate an FL deployment for vehicular/transportation IoT where edge devices train local models and share updates to a central server. They have claimed that their FL-IDS system outperforms traditional centralized learning with machine learning and deep learning approaches regarding accuracy and loss. In 2025, Nguyen et al. [13] present a two-stage federated NIDS combining local DNN feature extractors with global aggregation to boost packet-level attack detection performance. They try to enhance the detection accuracy of known attacks, robustness and resilience to new attack patterns, and privacy preservation, using packet-level granular data.

To achieve FL goals and solve the problem of privacy, security, and bandwidth efficiency, most current methods consider standard FL by sharing parameters (weights and biases) and hyperparameters that control the learning process itself. These processes are time-consuming and less effective in detecting zero-day exploits in time. Also, the imbalanced and non-IID distribution of data in a heterogeneous environment can be prone to the risk of generating a biased model. These gaps motivate the development of a new federated knowledge-sharing mechanism. This paper introduces TransFeL-NIDS, a novel federated zero-day detection method that allows clients to share the latent zero-day signatures alongside other parameters without violating privacy or having communication overhead.

Table 1. The Advantages and Limitations of Related Research

paper	Advantages	Limitations
DeepFed [9]	Tailored CNN+GRU; secure aggregation (Paillier); privacy-preserving FL demonstration.	Shares weights only; limited zero-day propagation; potential communication cost.
EEFED [10]	Personalized FL; handles heterogeneity & drift; improved local performance.	Extra computation /communication; complex aggregation
Fed-ANIDS [11]	Autoencoder-based anomaly scoring; low communication cost; effective anomaly detection across clients.	Reconstruction-based scores can vary by client; limited mechanism for propagating zero-day semantics.
FL-IDS [12]	Edge/vehicular IoT focus; practical deployment on constrained devices; privacy via local training.	Uses standard update aggregation; limited support for rare/zero-day transfer across nodes.
FedNIDS[13]	Packet-level two-stage design; improved detection through federated DNNs.	Parameter-sharing approach; sensitive to non-IID distributions and rare classes.

Journal of Distributed Computing and Systems(JDCS)
Vol.9, Issue 1, 2026, Pages 83-88

III. METHODOLOGY

In this section, we introduce the proposed framework and elaborate on our federated setup and the deep learning model that is more suitable for the proposed method. In TransFel-NIDS, multiple clients participate to share their model knowledge and gain experience to detect local unknown attacks. A single client can produce a latent signature for a zero-day attack that is absent in many other clients.

A. Federated Setup

This paper proposes a new federated paradigm. In our FL approach, local training using raw data on edge servers is employed to ensure data privacy and security. The overall architecture of our FL approach is shown in Fig. 1.

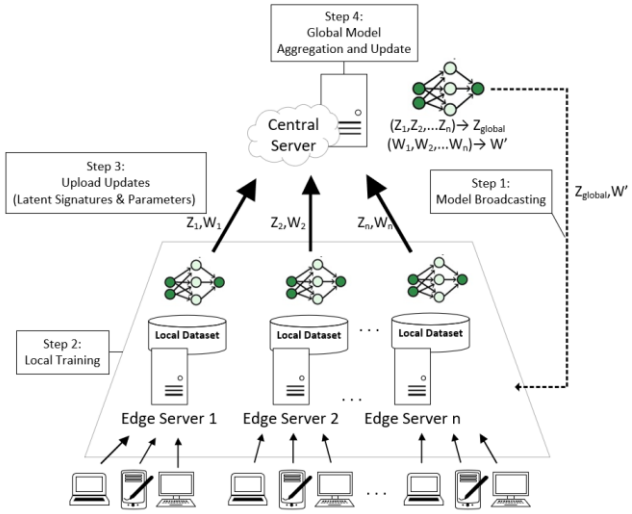

Figure 1. The Architecture of Our Proposed Method (TransFL-IDS)

In the proposed FL approach, the server broadcasts a copy of the latest version of the global model's parameters to the chosen clients (Step 1). Participants then use the shared model's parameters to start the training based on the local dataset (Step 2). After local training, clients send their local model parameters and the latent signatures (representing anomalous or zero-day activity) to the server (Step 3). Sending compact embeddings alongside the parameters has a low communication cost and is far cheaper than sharing raw data in non-FL setups. Finally, the central server aggregates updates from different clients using the aggregation algorithms to create a new global model that integrates knowledge from all participants (Step 4). It should be noted that latent signatures cannot be aggregated using standard FL algorithms (such as FedAvg, FedProx, and so on), because latent signatures require representation-level aggregation. Therefore, the central server uses representation-level aggregation techniques such as clustering, density-based aggregation, or knowledge distillation to produce global signature prototypes. The latent signatures aggregation outperforms model-weight aggregation when addressing non-IID and unknown attack propagation. At the end, the central server sends new updates (new latent signatures and new parameters) to clients, and the training process is repeated iteratively until a fully trained model is achieved, capable of performing well on data and detecting unknown threats from all client nodes. Here is a clear Python-style pseudocode for the server and client side.

Algorithm1. TranFel-IDS Server-side algorithm

```
Initialize global model W^0
Initialize global signature set P^0 = ∅

for each round t = 1 … T do
   Select a set of clients C_t
   Send (W^t, P^(t−1)) to all clients in C_t

   Receive from each client k ∈ C_t:
      w_k  ← weight update
      Z_k  ← latent signatures

   # Aggregate parameters
   W^(t+1) = (1 / |C_t|) * Σ_k w_k

   # Aggregate latent signatures
   Z_global = {Z_1, Z_2, …, Z_n} from all clients
   P^t = ClusterAndSelect(Z_global)    # e.g., cluster centroids

   Broadcast (W^(t+1), P^t) to clients
end for
```

Algorithm2. TranFel-IDS Client-side algorithm

```
# Receive global model
W_global = receive(W)

# Load global weights
model.load(W_global)

# ----- Local training -----
for epoch in range(E):
   for x, y in local_data:
      z = model.encoder(x)          # latent vector
      L = loss(model, z, y)         # classification loss
      L.backward()
      optimizer.step()
      optimizer.zero_grad()

# ----- Prepare weight update -----
W_local = model.get_weights()
w1, w2, ..., wn = W_local - W_global     # weight update vector

# ----- Extract latent signatures -----
Z_list = []
for x in local_data:
   if is_rare_or_anomalous(x):
      Z_list.append(model.encoder(x))  # Z1, Z2, ..., Zn

# Keep only top-K signatures
Z1, Z2, ..., Zn = select_K(Z_list)

# ----- Send results to server -----
send(w1, w2, ..., wn,  Z1, Z2, ..., Zn)
```

B. Transformer-XL as a Deep Learning Model

To effectively capture long-range temporal dependencies in network traffic, the proposed method leverages Transformer-XL[14] in the FL setup as the core sequence-modeling component. Each client deploys a local Transformer-XL model that processes sequences of network traffic tokens. Standard transformer model[15] operate on fixed-length segments and lack an explicit mechanism for retaining information beyond the segment boundary, which limits their ability to model evolving patterns commonly observed in intrusion detection scenarios. Transformer-XL overcomes this limitation through its segment-level recurrence mechanism, enabling the model to reuse hidden states from previous segments. As a result, it maintains contextual information over substantially longer time spans compared to standard architectures. In the context of intrusion detection, such extended memory is critical for identifying subtle, gradually emerging anomalies and handling concept drift in dynamic network environments. By leveraging the recurrent memory, Transformer-XL processes raw or feature-extracted traffic data as a continuous temporal stream, allowing it to learn both short-term fluctuations (e.g., bursty traffic) and long-term behavioral trends (e.g., slow-moving attacks). Furthermore, its relative positional encoding scheme ensures consistent performance across variable-length sequences, improving robustness to irregular or non-uniform traffic patterns.

During the training process, hidden states (memory vectors) are generated and anomalous sequences encoded as latent zero-day signatures. In other words, these signatures are extracted from the hidden states of a Transformer-XL model trained locally. The edge servers share latent behavioral signatures (including a small set of anomaly latent vectors and optional metadata such as anomaly score) beside parameter updates. Actually, instead of raw traffic, the system transmits latent vectors that encode the behavioral characteristics of anomalies. The central server aggregates these vectors using a representation-level aggregation technique such as clustering, density-based aggregation, or knowledge distillation.

C. Cross-Client Zero-Day Detection

To address the challenge of detecting zero-day attacks distributed across multiple clients, the proposed framework leverages similarity in the latent representation space extracted by the Transformer-XL encoder. When a client observes new traffic, it generates its own latent representation Z_β that captures high-level behavioral patterns of network traffic, including unknown or previously unseen anomalies. By sharing and aggregating these latent signatures at the central server, the system constructs a global prototype dictionary representing diverse attack behaviors observed across the federation. When a client encounters a new, rare pattern, it compares the corresponding latent representation against this global dictionary. If a close match is found, the client flags the event as potentially malicious. This cross-client similarity-based detection enables early identification of zero-day threats by exploiting shared knowledge encoded in the latent space, without requiring access to raw network data. Moreover, it enhances generalization to novel attacks by unifying distributed observations into a collective anomaly detection mechanism.

$$\text{Anomaly\_score} = \text{sim}(Z_\beta, Z\_\text{global})$$

IV. COMPARISON

To evaluate the effectiveness of the proposed method, we compare the standard FL approach-where clients only transmit model parameters-with our proposed paradigm, in which clients send both model parameters and latent signatures extracted from the hidden layer of the model.

Here is a clear comparison table between our proposed method and standard FL.

Table 2. Our proposed method vs Standard FL.

Feature / Criterion	Standard FL (parameters)	Our FL Approach (parameters & latent signatures)
Handles Non-IID Clients	weak - based on weights, causing model drift and degraded performance.	Strong - latent signatures give the server extra information about client distributions and semantics, improving aggregation.
Handles Data Imbalance (common in IDS datasets)	Weak - often ignores minority classes.	Strong - latent signatures highlight minority-class embeddings, improving global model sensitivity to rare attacks.
Representing Local Knowledge	Limited - weights/biases alone cannot express nuanced client-specific behaviors.	Strong - latent signatures act as compressed knowledge vectors of local attack semantics.
Detecting Emerging (Zero-Day) Patterns	Weak - model learns only from past distributions.	Strong - latent signatures expose novel patterns before full training, aiding zero-day detection.
Communication Cost	Very low	Low (parameters + signature vector)
Client Privacy	High - no raw data shared.	High - latent signatures are compressed non-reconstructable representations; privacy is preserved.
Global Model Accuracy	Weak in non-IID.	Higher and more stable across all settings.
Convergence Speed	Slow - inconsistent client gradients.	Faster - latent signatures guide aggregation toward common decision boundaries.
Scalability	Good	Good - latent signature size is small (e.g., 128–1024 dims).

V. Conclusion and Future Work

In this paper, we introduced a novel federated learning approach that allows clients to share the latent behavioral signatures (representing anomalous or zero-day activity) alongside the parameters. Unlike standard FL methods that only use model parameters, our method allows the server to capture important structural information about rare, minority, and emerging attack behaviors without accessing any raw traffic data. By combining both weight updates and latent signatures, the global model becomes much more resilient to non-IID and imbalanced network traffic, which are two challenges that significantly impair the performance of existing FL-based NIDS. The latent signatures effectively convey zero-day characteristics and high-level behavioral patterns while keeping privacy intact and non-invertible. Our design also creates a unified representation space across clients, leading to faster convergence, improved anomaly sensitivity, and better generalization, especially for unknown threats. Overall, our method provides a practical and scalable solution for collaborative intrusion detection in distributed settings and presents a promising direction for future research in privacy-preserving detection of zero-day threats. In future work, we plan to implement the proposed framework, conduct a comprehensive comparison with the baseline federated NIDS model, and perform an extensive evaluation to validate its effectiveness.

REFERENCES

[1] "World Economic Forum's (WEF) The Global Risks Report,"2023.[Online].Available: https://www3.weforum.org/docs/WEF_Global_Risks_Report_2023.pdf

[2] "Top trends in cybersecurity," 2022. [Online]. Available: https://www.gartner.com/en/articles/7-top-trends-in-cybersecurity-for-2022

[3] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," Cybersecurity, vol. 2, no. 1, pp. 1–22, 2019.

[4] M. Alkasassbeh, E. H. Omoush, M. Almseidin, and A. Aldweesh, "A Self-Adaptive Intrusion Detection System for Zero-Day Attacks Using Deep Q-Networks," IEEE Access, 2025.

[5] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, "Federated learning of deep networks using model averaging," arXiv Prepr. arXiv1602.05629, vol. 2, no. 2, 2016.

[6] Y. Liu et al., "Fedbcd: A communication-efficient collaborative learning framework for distributed features," IEEE Trans. Signal Process., vol. 70, pp. 4277–4290, 2022.

[7] D. N. P. Suthishni and K. S. S. Kumar, "A review on machine learning based security approaches in intrusion detection systems," in 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, 2022, pp. 341–348.

[8] J. Lansky et al., "Deep learning-based intrusion detection systems: a systematic review," IEEE Access, vol. 9, pp. 101574–101599, 2021.

[9] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, "DeepFed: Federated deep learning for intrusion detection in industrial cyber–physical systems," IEEE Trans. Ind. Informatics, vol. 17, no. 8, pp. 5615–5624, 2020.

[10] X. Huang, J. Liu, Y. Lai, B. Mao, and H. Lyu, "EEFED: Personalized federated learning of execution and evaluation dual network for CPS intrusion detection," IEEE Trans. Inf. Forensics Secur., vol. 18, pp. 41–56, 2022.

[11] M. J. Idrissi et al., "Fed-anids: Federated learning for anomaly-based network intrusion detection systems," Expert Syst. Appl., vol. 234, p. 121000, 2023.

[12] M. H. Bhavsar, Y. B. Bekele, K. Roy, J. C. Kelly, and D. Limbrick, "Fl-ids: Federated learning-based intrusion detection system using edge devices for transportation IoT," IEEE Access, vol. 12, pp. 52215–52226, 2024.

[13] Q. H. Nguyen, S. Hore, A. Shah, T. Le, and N. D. Bastian, "FedNIDS: A federated learning framework for Packet-based network intrusion detection system," Digit. Threat. Res. Pract., vol. 6, no. 1, pp. 1–23, 2025.

[14] Z. Dai, "Transformer-xl: Attentive language models beyond a fixed-length context," arXiv Prepr. arXiv1901.02860, 2019.

[15] A. Vaswani, "Attention is all you need," Adv. Neural Inf. Process. Syst., 2017.

How to cite: H. Yazdanpanah, S. M. Matinkhah, AA Nikoukar, **TransFeL-NIDS: Federated Latent Zero-Day Signatures using a Memory-Based Transformer for Network Intrusion Detection System,** Journal of Distributed Computing and Systems (JDCS), Vol 9, No 1, Pages 83-88, 2026.

Hamidreza Yazdanpanah is a Ph.D. Student in Computer Networks at Yazd University, with a focus on AI based NIDS. His research interests include Network Security, Intrusion Detection Systems, and Artificial Intelligence.

Email: yazdanpanah.hr@stu.yazd.ac.ir

S. Mojtaba Matinkhah received the Ph.D. degree in Computer Networks from Tehran Polytechnic, Tehran, Iran. He is currently a Faculty Member of the Computer Engineering Department, Yazd University, Iran. His research interests include Computer Networks, Privacy&Security, Internet of Things, and Cloud Edge Fog Computing.

Email: matinkhah@yazd.ac.ir

AliAkbar Nikoukar received the Ph.D. degree in Computer Networks from Yuan Ze University, Taoyuan, Taiwan. He is currently a Faculty Member of the Computer Science Department, Yasouj University, Iran. His research interests include Optical Networking, 6G, Zero Touch Networks, and AIoT.

Email: nikoukar@yu.ac.ir