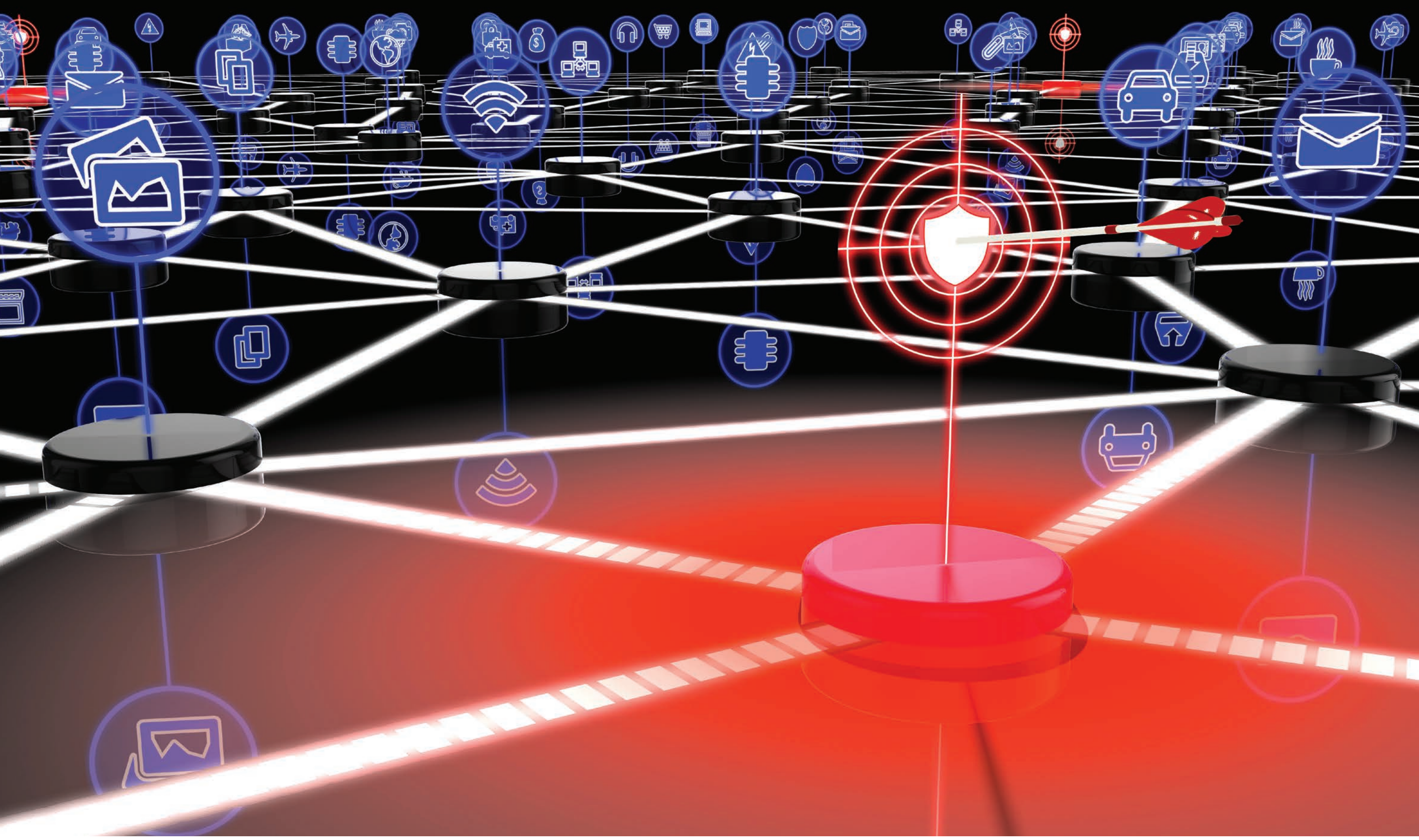

©SHUTTERSTOCK.COM/BEEBRIGHT

LOW-RATE DENIAL-OF-SERVICE ATTACK DETECTION IN OPEN RADIO ACCESS NETWORK

Integrating Federated Learning

Cheng-Feng Hung, Chui-Chen Kuo, and Shin-Ming Cheng

Security vulnerabilities have become increasingly critical with the growing connection of Internet of Things (IoT) devices and industrial control systems to 4G/5G private networks (PNs). Attacks in such environments are often complex and diverse, requiring advanced and adaptive detection mechanisms. However, conventional detection systems, which typically rely on the core network, incur high costs, introduce latency, and lack adaptability. The open radio access network (O-RAN) architecture provides a flexible and cost-effective framework by decoupling hardware and software components. This article presents a novel solution, the O-RAN advanced information security sharing system (O-RAN AIS3), which integrates federated learning (FL) to enhance detection capabilities. Detection

Digital Object Identifier 10.1109/MVT.2025.3623298
Date of publication 10 November 2025; date of current version 24 December 2025

> **THIS COLLABORATIVE LEARNING APPROACH ENABLES CROSS-DOMAIN KNOWLEDGE SHARING, WHICH IS ESSENTIAL FOR ADDRESSING THE DYNAMIC NATURE OF ATTACKS IN PNS AND OPTIMIZING THE DETECTION OF EMERGING THREATS IN REAL TIME.**

models are deployed in near real-time RAN intelligent controllers (Near-RT RICs) in diverse deployment scenarios and are periodically updated through parameter sharing. A Y1 consumer aggregates parameters from multiple Near-RT RICs, enabling knowledge synchronization across diverse network environments. This collaborative learning approach enables cross-domain knowledge sharing, which is essential for addressing the dynamic nature of attacks in PNs and optimizing the detection of emerging threats in real time. We evaluated the effectiveness of this approach in detecting unseen low-rate and high-rate denial-of-service (DoS) attacks, leveraging FL for enhanced cross-operator security. The dataset, collected in an emulated PN environment, included traffic generated from 11 distinct attack types, covering both low-rate and high-rate scenarios. Experiments conducted in two 5G O-RAN PN environments demonstrated improved detection accuracy in both cases. These results underscore the effectiveness of FL-based systems in addressing the increasing complexity of malicious behaviors in operational PNs.

Introduction

As an increasing number of IoT devices, user equipment (UE), and industrial control systems connect to 4G/5G PNs, the security challenges facing telecom networks have become increasingly severe [1], [2]. Relying exclusively on intrusion detection systems and core network components, such as the user plane function for traffic analysis and attack detection has several limitations. This approach often requires vendor-specific updates and adjustments, incurring high costs and operational delays. Furthermore, it constrains the network's capability to apply fine-grained security policies in response to real-time threats. To promote openness in telecom infrastructure and reduce vendor lock-in, the O-RAN Alliance was established in 2018 by a consortium of telecom operators, equipment vendors, and research institutions [3]. O-RAN standardizes interfaces to enable third-party development of interoperable white-box hardware and software [4]. O-RAN not only decouples traditional base stations into the O-RAN central unit (O-CU), O-RAN distributed unit (O-DU), and O-RAN radio unit (O-RU), but also introduces the RIC, which leverages artificial intelligence/machine learning (AI/ML)-based services to optimize operations [5]. By dynamically adjusting policies based on the real-time status of RAN components, O-RAN enhances network management flexibility and operational efficiency [6].

With the decoupling of base stations in the O-RAN architecture, O-CU and O-DU can transmit the collected packet data to the Near-RT RICs via the E2 interface and collaborate with the non-RT RIC for training and analysis while utilizing FL and deep reinforcement learning (DRL) frameworks to enhance malicious attack detection [5], [7], [8]. Houda et al. [9] and Sheikhi and Kostakos [10] integrated FL and deep learning to detect jamming and IP spoofing attacks, improving model convergence, and enhancing 5G intrusion detection. Rumesh et al. [11] used FL in a network digital twin (NDT) framework for anomaly detection, surpassing traditional classifiers in detecting user datagram protocol (UDP) distributed DoS (DDoS) and bandwidth hog attacks. The variability of attack patterns across domains and telecom operators complicates the development of generalized defense mechanisms, particularly for identifying low-rate DoS attacks [12]. To address this challenge, collaborative learning across domains is essential for sharing threat intelligence and enhancing detection accuracy [13]. FL offers a privacy-preserving solution by enabling each domain to retain its data locally while sharing only model parameters. This approach not only mitigates the privacy and security concerns associated with centralized learning (CL) but also reduces bandwidth consumption and aligns with the CIA security principles, making it well-suited for distributed and heterogeneous O-RAN environments.

In this article, we propose integrating the O-RAN architecture with FL to improve detection capabilities, hereafter referred to as the *O-RAN AIS3*. A local logistic regression (LR) model is deployed in the Near-RT RIC via xApp, facilitating real-time monitoring and anomaly detection. Near-RT RICs periodically update and transmit parameters to the Y1 consumer, utilizing the federated averaging (FedAvg) algorithm, which aggregates parameters from multiple Near-RT RICs to update the model weights. The updated parameters are then synchronized to the Near-RT RICs in each field. This approach enables the exchange of diverse attack patterns and detection capabilities across different environments, enhancing overall network security. To evaluate the effectiveness of FL in attack detection, we set up two 5G O-RAN experimental environments. The first environment is configured for high-rate DoS attacks, while the second is configured for low-rate attacks. We constructed a custom dataset by executing real-world attacks in a PN testbed. The dataset includes 10 types of high-rate DoS attacks: Internet control message protocol (ICMP) flood, TCP SYN, ACK, FIN, and RST floods, UDP, DNS, NTP, SNMP, and TFTP floods, as well as one low-rate attack based on the Slowloris technique. This dataset enables simulation and analysis of diverse attack traffic in PNs, providing a realistic basis for evaluating detection performance under both high-rate and stealthy low-rate conditions. We compare FL-based and CL models to assess the impact of parameter sharing on detection accuracy. Experimental results show that the first 5G O-RAN PN effectively detects previously unseen low-rate DoS attacks, while the second identifies previously unseen high-rate attacks, thereby validating the feasibility of FL-based cross-domain detection in O-RAN.

Background and Related Work

O-RAN Architecture

- *Near-RT RIC:* The Near-RT RIC communicates with the RAN via the E2 interface to manage E2 nodes, such as the O-CU and O-DU, as illustrated in Figure 1. The collected data are stored in the shared data layer, enabling xApps to access it for AI/ML-based RAN optimization [5], [7]. The Near-RT RIC also supports Y1 consumers, which subscribe to shared data layer data via the Y1 interface [5]. These consumers can aggregate data from multiple Near-RT RICs, providing broader network insights and enhancing the adaptability of AI/ML models [6]. Optimized results are fed back to the Near-RT RIC to improve RAN adaptability and resource orchestration [3].
- *Y1 consumer:* Y1 consumer applications within the O-RAN architecture access RAN analytical information from the Near-RT RIC via the Y1 interface [5], [13]. They use near real-time network analysis to optimize performance, enhance user experience, or formulate strategies. Y1 consumers actively subscribe to or query specific RAN data from the Near-RT RIC, such as key performance indicators (KPIs), event alarms, and predictive analytics. They then promptly acquire the required information through the Y1 interface. This enables them to rapidly respond to network changes, adjust resource allocation, and improve efficiency.
- *O-CU:* O-CU handles base station functions, such as radio resource control (RRC), service data adaptation protocol (SDAP), and packet data convergence protocol (PDCP) [6], including RRC signaling at layer 3 and user data at layer 2. SDAP maps quality of service flows to data radio bearers to meet requirements. PDCP manages user data transmission and reception, including integrity, encryption, compression, reordering, and sequencing, to improve efficiency and mobility. RRC configures wireless connections for communication. O-CU connects to the Near-RT RIC via E2, providing monitoring of UE traffic, quality of service, and handovers, and receiving control commands for resource management [3], [8].
- *O-DU:* The O-DU executes key layer 2 functions, including radio link control (RLC) for data transmission, medium access control (MAC) for resource allocation, and PHY-high for modulation and error control [6]. It interfaces with the Near-RT RIC via E2 [8], the O-CU via F1, and the O-RU via the open fronthaul, thereby supporting coordinated network operations.
- *E2 interface:* The E2 interface is specified by the O-RAN Alliance to connect the Near-RT RIC with E2

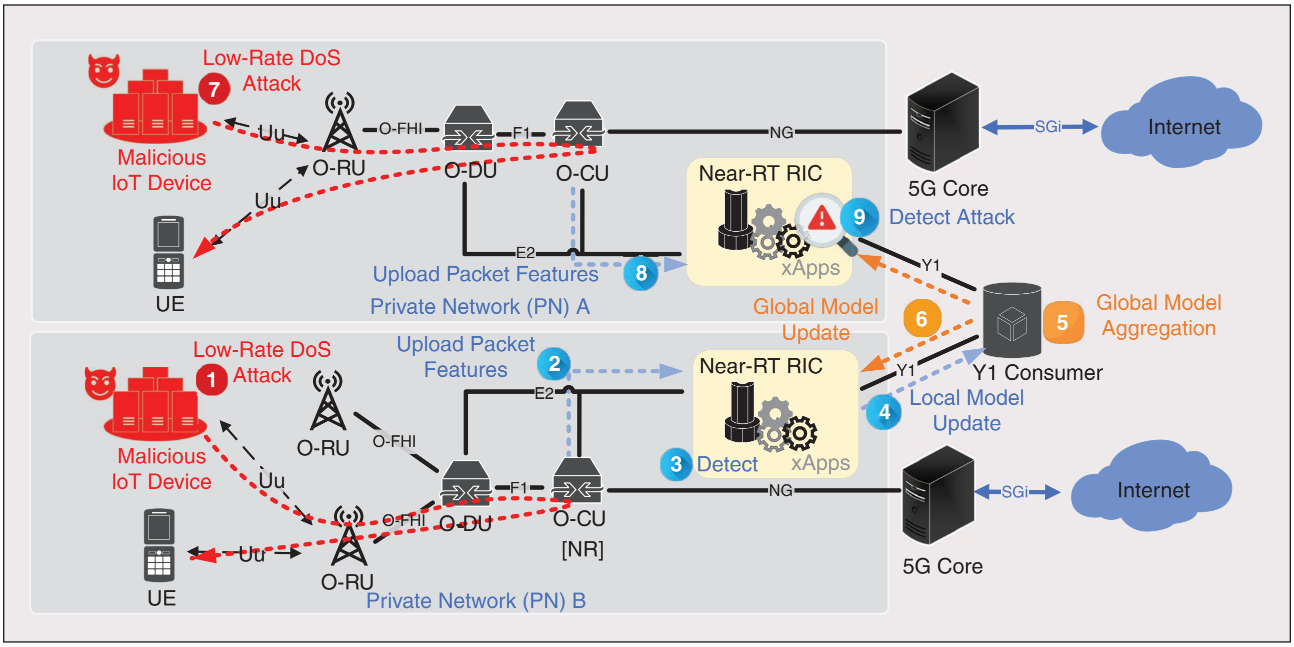

FIGURE 1 Overview of O-RAN architecture with O-RAN AIS3.

nodes, such as the O-CU and O-DU, as shown in Figure 1. It enables the Near-RT RIC to collect data from E2 nodes, supporting network optimization, resource allocation, and management [7], [8]. O-RAN WG11 recommends using Internet protocol security to protect E2 traffic [14].

- *Y1 interface:* The Y1 interface is a critical communication channel in the O-RAN architecture that connects the Near-RT RIC and Y1 consumer, as shown in Figure 1. The Y1 interface adopts TCP to ensure reliable data transmission, uses transport layer security (TLS) to guarantee communication security and integrity, employs HTTP/HTTPS at the application layer to facilitate communication and integration, and applies the lightweight JSON format for data representation [13]. Combining these protocols, the Y1 interface provides a secure and efficient communication channel between the Near-RT RIC and the Y1 consumer. Its functionalities enable Y1 consumers to subscribe to updates or query specific RAN analytical information, such as KPIs, event alarms, and predictive analytics. This supports network performance optimization and resource allocation adjustments.

O-RAN Security Threats

Compared with traditional 4G networks, O-RAN enhances flexibility and scalability through open interfaces and disaggregated components. This openness, however, enlarges the attack surface and strains legacy defenses. Adversaries can embed malicious commands in seemingly legitimate traffic to evade intrusion detection systems, and O-RAN's distributed control and multisite deployments expose networks to coordinated DDoS. Recent studies analyze malicious traffic and propose O-RAN–tailored defenses. Houda et al. [9] combine FL with deep reinforcement learning to detect and mitigate jamming, addressing data scarcity and convergence while optimizing resource usage; the focus remains on physical/MAC-layer interference. Sheikhi and Kostakos [10] employ unsupervised FL with a long short-term memory (LSTM) backbone to enhance intrusion detection for packet forwarding control protocol (PFCP) and IP spoofing, improving detection under privacy constraints but centering on control-plane anomalies. Rumesh et al. [11] introduce an NDT to pretrain models for Near-RT RIC deployment, where an FL-based detector outperforms traditional classifiers on UDP volumetric DDoS and bandwidth-hog traffic; the evaluation emphasizes high-rate behaviors within a single-operator scope. Collectively, these works confirm FL's promise for O-RAN security, yet most target high-rate attacks that are more tractable to detect and mitigate, leaving low-rate DoS comparatively underexplored in O-RAN.

Low-rate DoS depletes server resources not through volume but by sustaining long-lived, low-throughput connections or by slowly transmitting protocol elements. Representative types include slow headers (Slowloris-like partial or throttled headers), slow bodies (R.U.D.Y./slow POST with trickled payloads), slow reads (receiver-window throttling that forces server-side buffering), and low-frequency bursts that trigger timeouts while remaining below volumetric thresholds [15]. These behaviors share consistent footprints: long flow duration, sparse and high-variance interarrival times, low packet and byte rates, small payloads, and pronounced active/idle cycles. They resemble benign long sessions and therefore weaken the effectiveness of signature- and rate-based rules. These factors make LR-DoS detection in O-RAN PNs particularly challenging, as operators face both subtle traffic footprints and strict constraints on observability and data sharing.

Comparison With Similar Approaches

As summarized in Table 1, prior O-RAN FL defenses differ from our design across several operational dimensions. Houda et al. [9] aggregate at the Non-RT RIC and

TABLE 1 Comparison with related works.

Work	*Domain*	*Collaboration Scope*	*Aggregator Placement*	*Attack Coverage (LR-DoS?)*	*Observability*	*Stability under non-IID*	*Algorithm*
Houda et al. [9]	O-RAN	Intraoperator	Non-RT RIC	Jamming (no)	Radio/KPIs (no payload)	—	FL + DRL
Sheikhi and Kostakos [10]	5G	Intraoperator	Cloud/external	PFCP/IP spoofing (no)	Control-plane sequences	—	Unsupervised FL + LSTM
Rumesh et al. [11]	O-RAN	Intraoperator	Hierarchical: Near-RT RIC → Non-RT RIC	UDP DDoS, bandwidth hog (no)	KPIs (NDT)	—	FL + LSTM
This article	O-RAN	Cross-operator	Y1 consumer	ICMP/TCP/UDP floods; Slowloris (yes)	Flows at O-CU/Near-RT RIC	Standard-sharing	FL + LR

non-IID: nonindependent and identically distributed; Standard-sharing: presharing per feature mean/standard before training; Y1 consumer: a cross-operator service endpoint for model aggregation.

mitigate jamming from radio KPIs. Sheikhi and Kostakos [10] detect PFCP/IP spoofing with control plane sequences at a cloud server. Rumesh et al. [11] address volumetric UDP and bandwidth-hog anomalies using KPIs within an NDT with hierarchical RIC aggregation. These efforts remain confined to intraoperator collaboration and emphasize high-rate attacks that are easier to flag by throughput metrics, while low-rate DoS is not explicitly considered.

By aggregating at a Y1 consumer, we enable cross-operator federation without reliance on any single RIC. Our system detects both low-rate DoS attacks (e.g., Slowloris) and high-rate DoS attacks using flow-level features available at the O-CU and Near-RT RIC, and it stabilizes training under nonindependent and identically distributed (non-IID) traffic through cross-operator feature standardization. In practice, these design choices break operator data silos, close the LR-DoS gap left in prior work, and respect privacy and bandwidth constraints while supporting real-time detection in heterogeneous O-RAN deployments. These design choices illustrate our key contributions to collaborative intrusion detection in O-RAN.

O-RAN AIS3

This article proposes the O-RAN AIS3 to enhance security intelligence sharing across PNs and telecom operators, as illustrated in Figure 1. Many IoT devices lack rigorous security validation, rendering them susceptible to embedded malware or backdoors during deployment. Once compromised, these devices can be exploited to launch internal attacks, potentially resulting in system failures. Malicious traffic passes through the O-RU, O-DU, and O-CU before reaching the target UE. As specified by O-RAN WG11, wireless environments are vulnerable to radio jamming (threat ID: T-RADIO-01) [14], which highlights the need to address both end-device intrusions and PHY/MAC-layer threats.

In this article, we assume a trusted FL setting, where both the Y1 consumer and the participating Near-RT RICs operate honestly and follow the prescribed protocol. O-RAN AIS3 captures packets as they enter the O-CU and converts them into network flow data. The system then sends these flows to the Near-RT RIC via the E2 interface for localized detection and analysis of malicious traffic. As shown in Figure 1, once the Y1 consumer subscribes to the Near-RT RICs in PN A and PN B, each RIC performs preliminary local detection and model training. Each RIC sends its detection results and model parameters (weights w_i and bias b) to the Y1 consumer through the Y1 interface for global model aggregation. After updating the global model, the Y1 consumer redistributes the optimized parameters to the subscribed RICs, enabling PN A to identify low-rate DoS attacks and PN B to detect high-rate DoS attacks. This collaborative process supports real-time threat detection and mitigation within each local network.

FL Model Parameter Exchange

According to the Y1 interface application protocol specification [13], we design a subscription and transmission mechanism that complies with the standard. The parameter exchange procedure is illustrated in Figure 2. Initially, the Y1 consumer sends an HTTP POST request to the Near-RT RICs to subscribe to RAN analytics information notifications. Upon successful subscription and integration with O-RAN AIS3, the Near-RT RICs respond with an HTTP 201-created status, completing the subscription process.

To address the non-IID nature of data across PNs, we introduce a preprocessing step in which each Near-RT RIC shares its local data mean and standard deviation

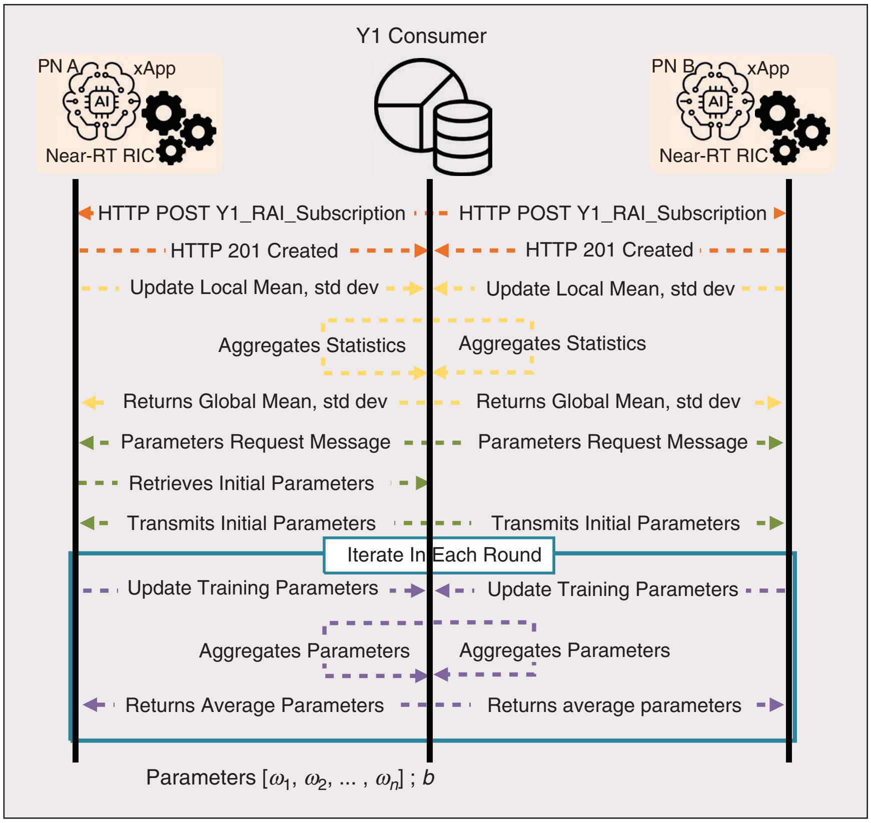
Y1 Consumer
PN A
xApp
AI
Near-RT RIC
PN B
xApp
AI
Near-RT RIC
HTTP POST Y1_RAI_Subscription
HTTP POST Y1_RAI_Subscription
HTTP 201 Created
HTTP 201 Created
Update Local Mean, std dev
Update Local Mean, std dev
Aggregates Statistics
Aggregates Statistics
Returns Global Mean, std dev
Returns Global Mean, std dev
Parameters Request Message
Parameters Request Message
Retrieves Initial Parameters
Transmits Initial Parameters
Transmits Initial Parameters
Iterate In Each Round
Update Training Parameters
Update Training Parameters
Aggregates Parameters
Aggregates Parameters
Returns Average Parameters
Returns average parameters
Parameters [ω_1, ω_2, ... , ω_n] ; b

FIGURE 2 FL parameter transmission in O-RAN AIS3.

with the Y1 consumer. After aggregating these statistics, the Y1 consumer returns the global average mean and standard deviation, enabling consistent data standardization across all Near-RT RICs. Standardization is particularly crucial for models sensitive to feature scales. It reduces knowledge disparities and accounts for variations in attack traffic. This approach enables O-RAN AIS3 to reduce false negatives (FNs) by providing a more comprehensive view of network flows, thereby enhancing detection performance across all participating PNs.

Model initialization starts when the Y1 consumer requests initial parameters from PN A at training-round zero. The Y1 consumer then distributes these parameters to all participating Near-RT RICs. In each subsequent round, every Near-RT RIC transmits its updated local model parameters back to the Y1 consumer, which averages them to update the global model. The system repeats the procedures illustrated in the blue block in Figure 2 until it completes the specified number of training rounds.

Algorithm of Model

Equation (1) defines the LR algorithm implemented at the Near-RT RICs. After each training round, the model computes the feature weights w_i and bias b for classification. These parameters are used to calculate z from the input features x_i, which is then passed through a sigmoid function to yield a probability between 0 and 1. Inputs with a probability greater than or equal to 0.5 are classified as *attack*, while others are labeled *benign*, indicating the presence or absence of malicious network flows within the PN.

The classification rule for the sigmoid function algorithm is given by

$$\sigma(z) = \frac{1}{1+e^{-z}} \text{ and } z = \sum_{i=1}^{n} w_i x_i + b \qquad (1)$$

where z is calculated as the weighted sum of features, $\sigma(z)$ sigmoid function maps the input z to a probability value between 0 and 1, w_i represents the weight associated with each feature, x_i are the input data of the feature, b is the bias term, and n is the total number of features.

In O-RAN AIS3, the LR model comprises two primary parameters: coefficients and intercept, corresponding to the weights w_i and bias b in (1). During training, the coefficients form a 1×79 matrix, where 1 represents the single output unit for binary classification, and 79 corresponds to the number of input features. The intercept represents the bias term b. When all feature weights are zero, the classification result depends solely on the value of b, which directly influences the predicted probability.

After each training round, the participating Near-RT RICs send their local model parameters to the Y1 consumer, which applies the FedAvg algorithm while considering the number of contributors. The Y1 consumer then updates the global model and redistributes the new parameters to all RICs, ensuring consistent and balanced detection performance across PNs. This process strengthens the overall effectiveness of collaborative learning.

Evaluation

Our evaluation consists of four main steps:

1) *Dataset construction and non-IID partitioning:* PN A is exposed only to high-rate attacks, while PN B is limited to low-rate attacks.
2) *Local model training:* Each Near-RT RIC trains an LR model using standardized LR features.
3) *Cross-operator aggregation:* Parameters are aggregated at the Y1 consumer to produce a federated global model.
4) *Flow-level testing:* The shared test set is evaluated using accuracy, precision, recall, and F1 score, as defined earlier, with direct comparison against a CL baseline under identical settings.

This design allows us to evaluate whether PN A, after federation, can detect low-rate DoS attacks it has never encountered during training, and whether PN B can likewise detect high-rate attacks, thereby demonstrating the benefit of cross-operator knowledge sharing in O-RAN AIS3.

Experimental Setup

We deployed three isolated, high-performance infrastructures to build the experimental environment. Two served as PN A and PN B, each containing an O-DU, O-CU, Near-RT RIC, and a 5G core. The third operated as the Y1 consumer. All setups used identical hardware configurations: an Intel i7-12700KF 12-core CPU, 32 GB of RAM, an NVIDIA RTX 3060 Ti GPU, a 1.5 TB solid-state drive, and Ubuntu 22.04.5 LTS. For the software stack, we utilized FlexRIC (developed by EURECOM) to implement the Near-RT RIC. At the same time, the 5G core and RAN components were constructed using the OpenAirInterface (OAI) open source platform. We employed Flower 1.14.0, running on Python 3.12.8, to establish the FL architecture. Additional relevant package versions include scikit-learn 1.6.1, NumPy 2.2.1, Pandas 2.2.3, and Matplotlib 3.10.0.

Explanation of Dataset

We used the OAI platform to emulate both low-rate and high-rate DoS attacks, along with benign traffic, in an O-RAN environment. Using the hping3 tool, we generated 10 types of high-rate attacks: ICMP floods; TCP floods (SYN, ACK, FIN, RST); and UDP-based floods, including DNS, NTP, SNMP, and TFTP. For TCP floods, we randomized source ports while targeting common service ports (e.g., 80, 443). For UDP floods, we randomized only the source ports. To create low-rate traffic, we implemented Slowloris in Python, opening 700 long-lived HTTP connections to a web server on port 80. Benign traffic was collected by browsing popular social media, e-commerce, and video-streaming sites. All traffic was captured at the O-CU for later processing.

We removed GTP headers and converted packets into flows using CICFlowMeter v4.0 (built with jnetpcap 1.4.1 r1425). We adopted a 70:30 train–test split and maintained a benign-attack ratio of 3:2. PN A's training set included 105,000 benign and 70,000 high-rate attack flows (175,000 total). In contrast, PN B's training set included 105,000 benign and 70,000 Slowloris flows (175,000 total). The shared test set consisted of 45,000 benign, 15,000 high-rate, and 15,000 low-rate attack flows (75,000 total).

This non-IID design exposed PN A only to high-rate attacks and PN B only to low-rate attacks, allowing us to evaluate whether each PN could detect previously unseen attacks after federation within the O-RAN AIS3 framework. The dataset was collected in a controlled testbed rather than a production network, but the attack tools (hping3, Slowloris) and the resulting flow-level footprints (long duration, sparse interarrival times, low packet rates) closely match those observed in real DoS incidents. These traces provide a reasonable proxy, and future work will validate the framework using operator-grade traffic to strengthen real-world applicability.

Implementation

We implemented O-RAN AIS3 using LR, a supervised learning algorithm well-suited for binary classification of network traffic. LR offered low parameter complexity and minimal overhead, enabling efficient flow classification within the FL framework, with each decision completed in under 10 ms. We evaluated model size, communication efficiency, and convergence behavior, and confirmed that LR delivered competitive performance while significantly reducing communication costs. Before training, we applied Z-score standardization to balance feature influence, avoid numerical instability in the sigmoid function, and accelerate convergence. Experimental results showed that standardization substantially improved model stability and accuracy. To ensure meaningful parameter aggregation, we applied consistent standardization across all local models. For this purpose, each participating Near-RT RIC shared its local means and standard deviations before FL training, as illustrated in Figure 2, which further enhanced the performance, stability, and convergence of LR under heterogeneous data distributions.

We configured the LR model with $L2$ regularization (Ridge) to mitigate overfitting and stabilize model weights. We selected $L2$ over $L1$ regularization because it retains all features without reducing coefficients to zero. We set the maximum number of iterations (*max_iter*) to 50, which proved sufficient based on empirical evaluation. Starting from the second training round, we enabled the *warm_start* parameter to preserve model weights across rounds, aligning with the iterative nature of FL. We chose the *saga* solver for its scalability, support for both $L1$ and $L2$ regularization, and efficiency in handling large or sparse datasets. We set the regularization strength (C) to 1, providing a balanced tradeoff between generalization and accuracy through parameter tuning. This configuration enabled the LR model to perform robustly, offering resilience to imbalanced traffic while maintaining computational efficiency.

We evaluated flow-level classification in a binary setting, distinguishing between benign and attack flows, using four standard metrics: accuracy, precision, recall, and F1 score. A true positive is an attack flow correctly flagged as an attack; a false positive (FP) is a benign flow incorrectly flagged as an attack; a true negative is a benign flow correctly identified; and a FN is an attack flow missed by the detector. Accuracy measures the proportion of correct decisions among all flows. Precision is the fraction of flagged attack flows that are truly attacks, while recall is the fraction of actual attack flows that are correctly flagged. The F1 score is the harmonic mean of precision and recall. Because undetected attacks (FN) are critical in security, we emphasize recall, while precision reflects false-alarm control.

To evaluate the effectiveness of O-RAN AIS3 within the FL architecture, we conducted a comparative experiment using the same LR model against a CL baseline. We emulated a realistic 5G telecom environment using the OAI platform and generated both Slowloris and 10 high-rate attack types to represent real-world PN threats. The CL results served as a benchmark to assess the performance of O-RAN AIS3 under FL. This experimental setup reflected the non-IID distribution of attacks across telecom networks, enabling a realistic evaluation of the model's adaptability and performance under various conditions.

Accuracy Comparison Between O-RAN AIS3 and CL

The experimental results demonstrate that O-RAN AIS3 significantly improves detection performance under non-IID data distributions compared to the CL baseline. As shown in Table 2, the CL-LR models

TABLE 2 Experiment result of CL and O-RAN AIS3.

Model	Accuracy	Precision	Recall	F1 Score
CL-LR in PN A	81.73%	85.84%	81.73%	80.27%
CL-LR in PN B	80%	85%	80%	78.09%
FL-LR 5th round	97.94%	98%	97.94%	97.93%
FL-LR 25th round	97.97%	98.03%	97.97%	97.96%
FL-LR 50th round	97.97%	98.04%	97.97%	97.97%

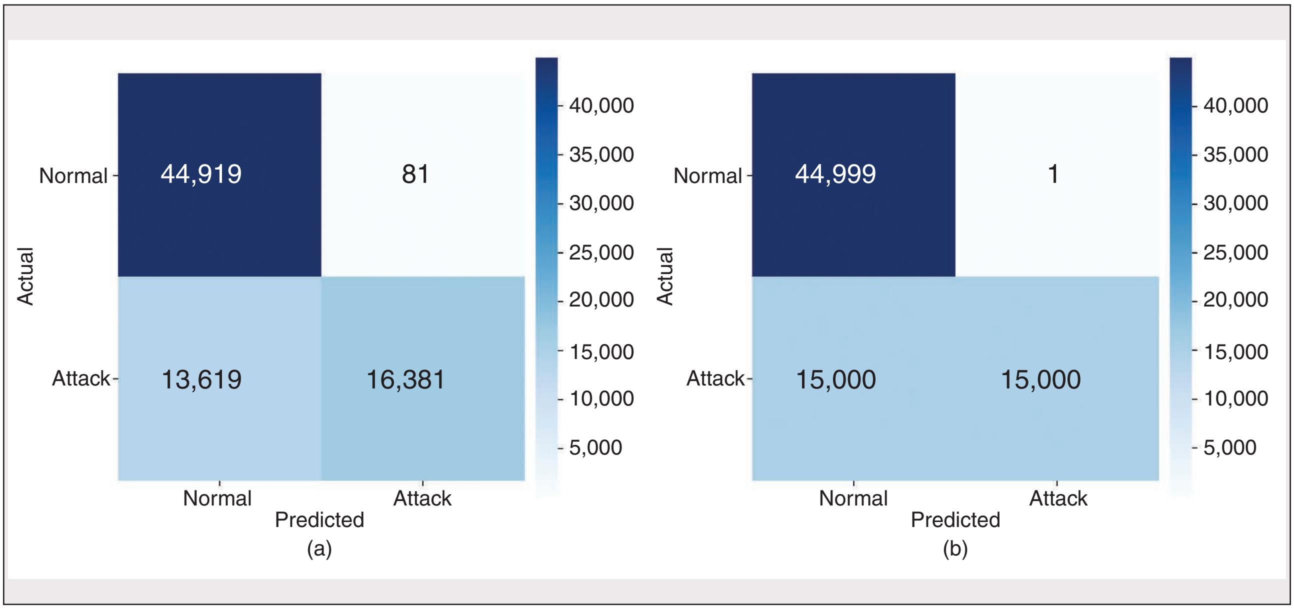

FIGURE 3 Confusion matrix of CL: (a) CL-LR at PN A and (b) CL-LR at PN B.

trained independently at PN A and PN B achieved limited performance due to their restricted exposure to diverse attack patterns. Specifically, PN A achieved an accuracy of more than 80.27% and PN B exceeded 78.09% across all metrics, including accuracy, precision, recall, and F1 score. These findings highlight the limitations of relying solely on local data, which constrains the model's ability to generalize to unseen threats.

In contrast, the FL-LR model in O-RAN AIS3 consistently outperformed the CL baseline, surpassing 97.97% across all metrics after 50 training rounds. Even at rounds 5 and 25, the model achieved near-optimal performance, demonstrating rapid convergence enabled by parameter aggregation and standardized feature scaling. Before training, all participating Near-RT RICs shared their local feature-wise statistics to apply consistent Z-score standardization, which ensured stable model aggregation across heterogeneous data distributions.

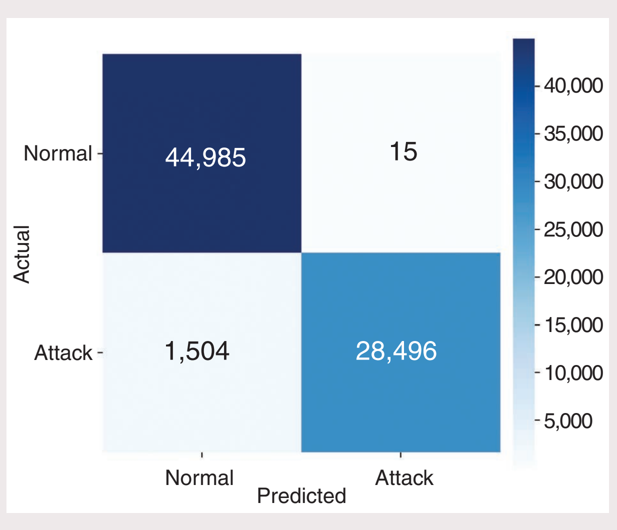

FIGURE 4 Confusion matrix of O-RAN AIS3. Fiftieth round at PN A and B.

Figure 3 shows that the CL-LR models suffered from high FN rates, with 13,619 and 15,000 misclassified attack flows in PN A and PN B, respectively. PN A also recorded 81 FPs, indicating that the model overfitted to its local data. In contrast, the FL-LR model, as illustrated in Figure 4, significantly reduced FNs to 1,504, lowering misclassifications by 12,115 in PN A and 13,496 in PN B. It also maintained a low FP count of 15. Although PN B experienced a slight increase in FPs, the overall error remained low and well-balanced, reflecting the improved generalization achieved by the global model.

These results show that O-RAN AIS3 combines FL, Y1 consumer aggregation, and cross-operator feature standardization to handle non-IID traffic in distributed PNs. It shares models across operators without exposing raw flows, reduces bandwidth use, and protects privacy. The design aligns with the O-RAN vision for intelligent, interoperable, and secure RANs and shows promise for deployment at multioperator scale. We also recognize limits: The dataset comes from an OAI testbed rather than a live network, the detector uses a lightweight LR core, and the threat scope covers only DoS and low-rate DoS. Future work will validate on operator-grade traces, broaden the attack set, and compare against deeper models.

Conclusion

This article proposes O-RAN AIS3, integrating FL into O-RAN PNs to enhance the detection of malicious traffic. Experimental results demonstrate that, compared to CL-LR models trained solely on local data, the FL-LR model significantly improves accuracy, precision, recall, and F1 score, raising performance in both PN A and PN B from about 80% to 97.97%. The FN counts also dropped sharply, decreasing from 13,619 to 1,504 in PN A and from 15,000 to 1,504 in PN B. O-RAN AIS3 effectively leverages FL to enable collaborative learning among Near-RT RICs across domains, allowing the sharing of attack patterns and consistent standardization without exposing raw data. This design strengthens detection capabilities while preserving data privacy, offering a scalable and practical solution for real-time threat identification in multioperator environments. Future work will investigate broader attack pattern coverage and explore privacy-preserving techniques, such as homomorphic encryption and differential privacy to enhance model security.

Acknowledgment

This work was supported in part by the National Science and Technology Council (NSTC), Taiwan, under Grants NSTC 113-2923-E-011-005-MY2 and 113-2634-F-011-002-MBK, and by project “PREACT: Privacy-respecting Explainable Assessment and Collection of Threats” (2024–2025), no. TM05000014, funded by the Technological Agency of the Czech Republic.

Author Information

Cheng-Feng Hung (hungchengfeng@nict.go.jp) is a fixed-term researcher with the Center for Research on AI Security and Technology Evolution, Cybersecurity Research Institute, National Institute of Information and Communications Technology, Tokyo 184-8795, Japan. His research interests include open radio access network and artificial intelligence security. Hung received his Ph.D. in computer science and information engineering from the National Taiwan University of Science and Technology, Taipei in 2025. He is a Member of IEEE.

Chui-Chen Kuo (M11215019 @mail.ntust.edu.tw) is currently studying for an M.S. degree in computer science and information engineering at the National Taiwan University of Science and Technology, Taipei 10617, Taiwan. Her research interests include open radio access network security and mobile networks. Kuo received her B.S. degree in computer science and information engineering from the National Taiwan University of Science and Technology, Taiwan, in 2023.

Shin-Ming Cheng (smcheng@mail.ntust.edu.tw) is a professor in the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei 10617, Taiwan. His research interests include mobile network security, Internet of Things system security, malware analysis, and artificial intelligence robustness. Cheng received his Ph.D. degree in computer science and information engineering from National Taiwan University, Taipei, Taiwan, in 2000. He is a Member of IEEE.

References

[1] R. Khan, P. Kumar, D. N. K. Jayakody, and M. Liyanage, “A survey on security and privacy of 5G technologies: Potential solutions, recent advancements, and future directions,” *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 196–248, 1st Quart. 2020, doi: 10.1109/COMST.2019.2933899.

[2] S. Soltani, A. Amanloo, M. Shojafar, and R. Tafazolli, “Intelligent control in 6G open RAN: Security risk or opportunity?” *IEEE Open J. Commun. Soc.*, vol. 6, pp. 840–880, 2025, doi: 10.1109/OJCOMS.2025.3526215.

[3] M. Polese, L. Bonati, S. D’Oro, S. Basagni, and T. Melodia, “Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges,” *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 1376–1411, 2nd Quart. 2023, doi: 10.1109/COMST.2023.3239220.

[4] M. Liyanage, A. Braeken, S. Shahabuddin, and P. Ranaweera, “Open RAN security: Challenges and opportunities,” *J. Netw. Comput. Appl.*, vol. 214, May 2023, Art. no. 103621, doi: 10.1016/j.jnca.2023.103621.

[5] O-RAN ALLIANCE Working Group 1, “O-RAN architecture description 12,” O-RAN ALLIANCE, Alfter, Germany, Tech. Specification, Jun. 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[6] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, “Toward next generation open radio access networks: What O-RAN can and cannot do!,” *IEEE Netw.*, vol. 36, no. 6, pp. 206–213, Nov./Dec. 2022, doi: 10.1109/MNET.108.2100659.

[7] C.-F. Hung, Y.-R. Chen, C.-H. Tseng, and S.-M. Cheng, “Security threats to xApps access control and E2 interface in O-RAN,” *IEEE Open J. Commun. Soc.*, vol. 5, pp. 1197–1203, Feb. 2024, doi: 10.1109/OJCOMS.2024.3364840.

[8] O-RAN ALLIANCE Near-Real-Time RIC and E2 Interface Work Group 3, “O-RAN E2 service model E2SM, RAN control 4.0,” O-RAN ALLIANCE, Alfter, Germany, Tech. Specification, Oct. 2023. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[9] Z. A. E. Houda, H. Moudoud, and B. Brik, “Federated deep reinforcement learning for efficient jamming attack mitigation in O-RAN,” *IEEE Trans. Veh. Technol.*, vol. 73, no. 7, pp. 9334–9343, Jul. 2024, doi: 10.1109/TVT.2024.3359998.

[10] S. Sheikhi and P. Kostakos, “Advancing security in 5G core networks through unsupervised federated time series modeling,” in *Proc. IEEE Int. Conf. Cyber Secur. Resilience (CSR)*, London, U.K., 2024, pp. 353–356, doi: 10.1109/CSR61664.2024.10679491.

[11] Y. Rumesh, D. Attanayaka, P. Porambage, J. Pinola, J. Groen, and K. Chowdhury, “Federated learning for anomaly detection in Open RAN: Security architecture within a digital twin,” in *Proc. Joint Eur. Conf. Netw. Commun. 6G Summit (EuCNC/6G Summit)*, Antwerp, Belgium, 2024, pp. 877–882, doi: 10.1109/EuCNC/6GSummit60053.2024.10597083.

[12] D. Mimran et al., “Security of open radio access networks,” *Comput. Secur.*, vol. 122, 2022, Art. no. 102890, doi: 10.1016/j.cose.2022.102890.

[13] O-RAN ALLIANCE Working Group 3, “O-RAN Y1 interface: Application protocol 1.01,” O-RAN ALLIANCE, Alfter, Germany, Tech. Specification, Oct. 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[14] O-RAN ALLIANCE Working Group 11, “O-RAN security threat modeling and risk assessment 5.0,” O-RAN ALLIANCE, Alfter, Germany, Tech. Rep., Feb. 2025. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[15] W. Zhijun, L. Wenjing, L. Liang, and Y. Meng, “Low-rate DoS attacks, detection, defense, and challenges: A survey,” *IEEE Access*, vol. 8, pp. 43,920–43,943, Feb. 2020, doi: 10.1109/ACCESS.2020.2976609.

VT