Blockchain-Augmented FL IDS for Non-IID Edge-IoT Data Using Adaptive Trimmed Mean Aggregation

Kalibbala Jonathan Mukisa, Love Allen Chijioke Ahakonye, *Member, IEEE*, Dong-Seong Kim, *Senior Member, IEEE*, and Jae-Min Lee, *Member, IEEE*

***Abstract*—The rapid expansion of the Internet of Things (IoT) into edge networks, which are populated by resource-constrained devices, introduces significant security challenges. This article introduces a robust intrusion detection systems (IDS) for edge-IoT networks developed as a combination of blockchain technology and federated learning (FL) with a customized aggregation strategy, adaptive trimmed mean aggregation (ATMA). Our design leverages a permissioned blockchain to authenticate clients and immutably store the final global model, guaranteeing that only verified participants contribute to the training. ATMA strategy used in the FL departs from fixed-threshold schemes in other works by dynamically adjusting its trimming parameter according to the observed variance in client updates. This variance-aware trimming provides strong Byzantine resilience without sacrificing model accuracy, and its sorting-based implementation maintains an O(n log n) computational complexity. The proposed setup was evaluated under combined label-flipping and Gaussian-noise attacks at adversarial rates of 0%, 10%, 20%, 30%, 40%, 50%, and 60%, in both IID and non-IID data distributions. The results demonstrated that our blockchain-backed ATMA preserves high detection performance under severe attack scenarios and does so with minimal overhead, making it a scalable, secure solution for safeguarding Edge-IoT deployments.**

***Index Terms*—Blockchain, byzantine, edge, federated learning (FL), IID, Internet of Things (IoT), trimmed mean aggregation.**

Received 21 May 2025; accepted 9 August 2025. Date of publication 14 August 2025; date of current version 24 October 2025. This work was supported in part by the Innovative Human Resource Development for Local Intellectualization Program through the IITP Grant Funded by the Korea Government (MSIT) under Grant IITP-2025-RS-2020-II201612, 20%; in part by the Priority Research Centers Program through the NRF funded by the MEST under Grant 2018R1A6A1A03024003, 20%; in part by the MSIT, Korea, under the ITRC Support Program under Grant IITP-2025-RS-2024-00438430, 20%; in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-ICT Challenge and Advanced Network of HRD (ICAN) Grant funded by the Korea Government (Ministry of Science and ICT) under Grant IITP-2025-RS-2022-00156394, 20%; and in part by the Gyeongsangbuk-do RISE (Regional Innovation System & Education) Project (Specialized Industry Scale-up unit) (20%). *(Corresponding author: Jae-Min Lee.)*

Kalibbala Jonathan Mukisa, Dong-Seong Kim, and Jae-Min Lee are with the Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi 39177, South Korea (e-mail: kjonmukisa@kumoh.ac.kr; dskim@kumoh.ac.kr; ljmpaul@kumoh.ac.kr).

Love Allen Chijioke Ahakonye is with the ICT Convergence Research Center, Kumoh National Institute of Technology, Gumi 39177, South Korea (e-mail: loveahakonye@kumoh.ac.kr).

Digital Object Identifier 10.1109/JIOT.2025.3598832

I. Introduction

THE RAPID expansion of the Internet of Things (IoT) has led to the connection of various devices, transforming industries, such as healthcare, manufacturing, and transportation [1], [2]. These systems are increasingly deployed in edge networks, where devices operate in resource-constrained environments with limited processing power and bandwidth [3], [4]. Edge IoT devices generate vast amounts of data critical for automation and decision-making [5]. However, the distributed nature of Edge-IoT networks introduces significant security challenges, particularly in ensuring system integrity and protecting against malicious attacks [6], [7].

Intrusion detection systems (IDS) have long been vital defense mechanisms for detecting and mitigating cyberattacks [8]. Traditional IDSs often rely on centralized data collection and analysis, in which data from all devices is transmitted to a central server for processing. However, this approach is inefficient in Edge-IoT environments for several reasons [9], [10]. For example, the constant transmission of data from edge devices to a central server incurs a significant communication overhead, which is problematic in resource-constrained environments [11], [12]. Additionally, centralized IDSs raise privacy concerns by requiring raw data transmission to the server, which is particularly difficult when handling sensitive information, such as personal health records or critical infrastructure data [13], [14]. Furthermore, IoT devices in different locations collect data with distinct distributions, referred to as nonindependent and identically distributed (non-IID) data, making it difficult for traditional IDSs to effectively generalize [15], [16]. As the number of connected devices increases, centralized IDS architectures face scalability issues as data processing at a central location becomes a bottleneck [17], [18].

To address these challenges, federated learning (FL) has emerged as a decentralized solution that enables collaborative machine learning while preserving data privacy [19], [20]. In FL, edge devices (clients) train models locally using their data and share only the model updates (weights) with a central server, thus avoiding transmitting raw data [21], [22]. This significantly reduces communication costs and addresses privacy concerns [23], [24]. After receiving all client updates, the central server aggregates these updates to build a global model [25], [26]. However, despite its advantages, FL introduces new vulnerabilities, particularly those related

TABLE I
COMPARISON WITH EXISTING STUDIES HIGHLIGHTING THE SIGNIFICANCE OF THE PROPOSED TMA (USED: √, NOT USED: χ)

Ref	Year	Target System	Method	Dataset	Lightweight Model	Decentralized Learning	Blockchain	Trusted Aggregation	Non-IID Data Adaptability
[42]	2021	IoT	CNN	CICDDOS2019	√	√	χ	χ	√
[43]	2022	IoT	CNN-LSTM	CICDDOS2019	√	χ	χ	χ	√
[44]	2022	IIoT	Hybrid GAN	ToN-IoT	√	χ	χ	χ	χ
[45]	2023	Metaverse	RF	ToN-IoT	√	χ	χ	χ	χ
[46]	2023	IoT	WD-SPRT	-	√	χ	χ	χ	√
[47]	2023	IIoT	FedIDSNet	FedIoTset, EdgeIoTSet	√	√	χ	χ	√
[34]	2023	Metaverse	MLP	MNIST	χ	√	√	√	χ
[48]	2023	IoT	DNN	-	√	√	√	√	χ
[49]	2023	SDN	DNN	NSL-KDD	χ	√	√	√	χ
Ours	2024	Edge-IoT	MLP	CICIoV, Edge-IoT	√	√	√	√	√

to Byzantine attacks, where malicious clients submit poisoned model updates to corrupt the global model [27]. Byzantine attacks severely threaten FL because adversarial clients can manipulate their model updates to render the global model ineffective or harmful when aggregated [28], [29]. Adversarial attacks can send malicious model weights to the server, whether purposefully or not [30], [31]. The widely used federated averaging (FedAvg) algorithm is particularly vulnerable to such attacks because it aggregates all updates equally without considering potential malicious intent [32]. This highlights the need for robust aggregation methods that secure a training process against adversarial clients [33], [34].

This study introduces a Blockchain-enhanced FL framework with adaptive trimmed mean aggregation (ATMA) to strengthen the IDS for Edge-IoT networks [35], [36]. The blockchain handles client authentication and verification, allowing only verified clients to train and store a tamper-proof record of the global model weights [37]. Unlike traditional approaches, ATMA dynamically adjusts its trimming parameter based on observed variance in client updates, providing superior defense against Byzantine attacks with O(n log n) computational efficiency [38], [39]. The edge server employs a comprehensive validation phase to evaluate the aggregated model before finalization, which ensures consistent performance in both IID and non-IID data environments [40], [41]. The proposed framework processes each neural network layer independently, allowing for fine-grained defense against layer-specific attacks-a capability absent in existing approaches like in [39]. The table I further compares this work with the other recent studies.

The main contributions of this work are as follows.

1) A Blockchain-enhanced FL framework with custom chain implementation for secure intrusion detection in Edge-IoT environments, ensuring authenticated client participation while maintaining tamper-proof model integrity with enhanced traceability.
2) Development of ATMA, which dynamically adjusts trimming thresholds based on variance patterns, providing superior Byzantine resilience compared to fixed-parameter approaches, while maintaining O(n log n) complexity versus the $O(n)$ or $O(n)$ requirements of alternatives.
3) Adoption of layer-wise processing capability that applies different trimming thresholds to each neural network layer, offering fine-grained protection against sophisticated attacks.
4) A comprehensive validation methodology that tests aggregated weights at the edge server before global model updates redistribution, ensuring performance consistency while preserving privacy through secure aggregation.
5) Extensive evaluation under both IID and non-IID data distributions with Byzantine attack rates up to 60%, demonstrating ATMA's superior resilience in maintaining high accuracy even with up to 60% Byzantine clients-outperforming traditional FedAvg.

II. BACKGROUND OF STUDY AND RELATED WORKS

A. Challenges in FL for Edge-IoT Networks

The emergence of FL with a collaborative model training without centralizing sensitive data aims to address IoT's significant security challenges due to the decentralized nature of the devices. However, securing the FL process is a non-trivial task, as it is susceptible to various adversarial attacks, especially Byzantine attacks, where some participating clients may submit malicious updates to corrupt the global model. Recent research has focused on enhancing the security and robustness of FL by integrating Blockchain technology and advanced aggregation strategies, such as the Krum [28], which improves robustness against Byzantine failures by effectively filtering out malicious updates; however, its complex and time-consuming nature makes it less ideal for real-time systems. Additional challenges associated with FL include the following.

1) *Data Heterogeneity (Non-IID Data):* In IoT networks, device-generated data is typically non-IID, meaning data distributions vary across devices, challenging standard FL algorithms that assume uniform data. As noted by Kim et al. [32], non-IID data can significantly affect learning processes, causing model bias, slower convergence, and reduced accuracy.
2) *Communication Overhead:* Efficient communication in FL is essential in resource-constrained IoT environments, where frequent model updates between clients and the server can be costly. This article presents a faster-converging model that also prioritizes client privacy, addressing what Que and Khan [50] identified as a key challenge in edge computing environments.
3) *Adversarial Attacks:* FL is vulnerable to Byzantine attacks, where malicious or malfunctioning clients submit compromised updates that corrupt the global model.

Traditional methods like FedAvg, which treat all updates equally, fail to filter out these harmful contributions.

Various approaches have been developed to enhance FL robustness. In [33], authors optimized model accuracy by implementing adaptive, hardware-friendly pruning ratios with iterative adjustments, improving system efficiency. However, determining optimal pruning ratios for each device involves high computational overhead, as discussed in [3]. Another study [24] introduced a trimming algorithm where benign clients minimize loss via gradient descent, while malicious ones inject interference through random matrices. The server then uses pairwise Euclidean distances to identify deviations, though its complexity limits practicality for real-time, resource-constrained Edge-IoT networks.

B. Blockchain Consensus Mechanisms for FL-Based IDS

A critical aspect of blockchain integration in FL-based IDS is selecting an appropriate consensus mechanism. Recent research by Saqib et al. [37] has demonstrated that consensus mechanism selection significantly impacts the performance and security of blockchain-enhanced FL systems for industrial IoT networks. Traditional mechanisms like Proof of Work (PoW) are resource-intensive and unsuitable for Edge-IoT environments with limited computational capabilities [37].

Proof of Authority (PoA) has emerged as a more suitable consensus mechanism for private permissioned blockchains in FL settings, particularly for intrusion detection applications [37]. The PoA relies on a limited number of trusted validators, typically administrators, to authenticate transactions. This approach, as demonstrated by Sun et al. [39], provides faster transaction validation while maintaining security, making it ideal for resource-constrained edge environments.

The lightweight model feature is essential for resource-constrained Edge-IoT environments, as noted by Zhu et al. [25], who emphasize that Edge or Fog resources are often less powerful compared to cloud. The proposed model performs better and requires significantly fewer parameters than CNN or hybrid GAN approaches. Decentralized learning capabilities address data privacy concerns without sacrificing model performance, as demonstrated by ProxyFL [26], which enables multi-institutional collaborations on decentralized data with improved protection for each collaborator's data privacy. Integrating blockchain with FL adds security with its immutable ledger [8].

Furthermore, trusted aggregation mechanisms are crucial for Byzantine resilience, as shown by Lavond et al. [30], whose TAG approach defends against backdoor attacks even when 40 percent of user submissions are malicious. Works such as BFLIDS [35] have utilized blockchain for enhancing FL-based IDSs in IoMT networks; however, they overlooked the critical balance between security and computational efficiency. Similarly, the hybrid CNN-BiLSTM approach by Zhou et al. [41] integrates blockchain with FL for IIoT security but lacks the adaptive aggregation mechanisms present in this work.

This study's approach addresses these limitations by implementing an ATMA alongside the PoA consensus, providing superior Byzantine resilience with $O(n\log n)$ complexity compared to existing solutions, making it suited for resource-constrained Edge-IoT devices. This combination enables the proposed approach to maintain high detection accuracy even under significant Byzantine attacks, outperforming recent implementations that struggle with the computational demands of blockchain integration while maintaining robust intrusion detection capabilities.

C. Byzantine Attacks and Non-IID Data

Byzantine attacks and non-IID data pose critical challenges in FL, especially within Edge-IoT networks [29]. In Byzantine attacks, malicious clients send corrupted updates to degrade model performance, while non-IID data, where each client's data distribution varies significantly, further complicates model training [31]. These issues lead to skewed global models that struggle with robustness and accuracy. Addressing these requires filtering out malicious updates and minimizing the model's sensitivity to adversarial inputs and distribution variance across clients.

Building on prior works [1], [21], we introduce a blockchain-enhanced FL framework for Edge-IoT networks to address non-IID data and security challenges. A private, permissioned blockchain ensures that only authenticated clients contribute to model training, preserving client identity and data integrity. A central validation dataset handles non-IID data and improves model generalization. Additionally, ATMA dynamically adjusts its trimming parameter based on observed variance in client updates, providing superior defense against Byzantine attacks with $O(n\log n)$ computational efficiency. This combined approach strengthens the reliability and security of IDS in Edge-IoT environments. It secures the FL framework by ensuring client authentication, model integrity, and robust aggregation against adversarial attacks and non-IID data. The proposed system processes each neural network layer independently, allowing for fine-grained defense against layer-specific attacks, a capability absent in existing approaches. The Flask server-client architecture enhances scalability, allowing seamless client registration and participation in training.

The combination of Blockchain and ATMA offers a comprehensive solution for securing FL in edge IoT environments. Using blockchain for client authentication and model integrity and ATMA for robust aggregation with variance-based adaptation, the proposed framework can withstand adversarial attacks and handle non-IID data while maintaining computational efficiency suitable for resource-constrained environments.

III. Methodology

The FL system as proposed in the Fig. 1 with N clients which are the Edge-IoT devices 1, where each client i has a local dataset $\mathcal{D}_i = (\mathbf{x}j, y_j)j = 1^{m_i}$ consisting of m_i samples. A subset $\mathcal{B}$ of size B may be Byzantine or malicious, while the remaining clients $\mathcal{G} = 1, 2, \ldots, N \setminus \mathcal{B}$ of size $N-B$ are honest. The FL framework is aimed at enabling the different devices

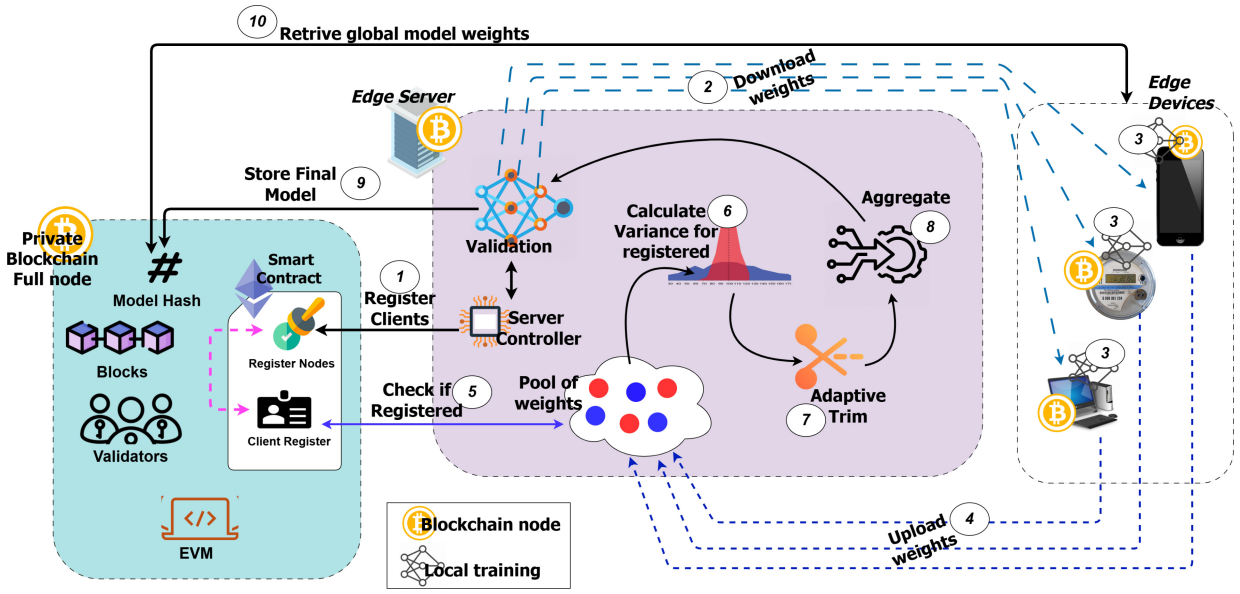


Fig. 1. Proposed system architecture showing edge-IoT devices connected through wired and wireless LAN, the edge server with ATMA aggregation, and blockchain integration for client registration and model storage.

on the network to collaboratively train a global model $\mathbf{w} \in \mathbb{R}^d$ that minimizes the following objective function in

$$F(w) = \frac{1}{N-B} \sum_{i \in \mathcal{G}} F_i(w) \tag{1}$$

where $F_i(\mathbf{w})$ represents the local objective function for client (Edge device) i with $\ell(\mathbf{w}; \mathbf{x}, y)$ being the loss function that measures the prediction error of model $\mathbf{w}$ on sample $(\mathbf{x}, y)$

$$F_i(w) = \frac{1}{m_i} \sum_{j=1}^{m_i} \ell\big(w; x_j, y_j\big). \tag{2}$$

The global model resulting from the aggregation done at the edge server is iteratively updated through multiple communication rounds. At each round t, the edge server distributes the latest global model weights obtained from the aggregation $\mathbf{w}^t$ to all clients. Each edge client i then performs local training on their local data to compute an updated model $\mathbf{w}_i^{t+1}$ by minimizing its local objective in

$$w_i^{t+1} = w^t \; - \; \eta \, \nabla F_i\big(w^t\big) \tag{3}$$

where η is the learning rate and $\nabla F_i(\mathbf{w}^t)$ is the gradient of the local objective. Without any measure against Byzantine, malicious clients can submit arbitrary updates $\tilde{\mathbf{w}}_i^{t+1}$ that deviate from the honest update rule.

ATMA strategy for aggregation on the server mitigates malicious update activity. Traditional robust aggregation methods like coordinate-wise median or trimmed mean use fixed parameters across all dimensions and rounds. ATMA dynamically adjusts its trimming parameter based on the observed variance in client updates, permitting more precise filtering of malicious contributions. When $\mathbf{w}_i^{t+1} \in \mathbb{R}^d$ denote the model update from client i at round $t+1$. For the neural network, these updates can be decomposed into L layers, where each layer l has parameters $\mathbf{w}i^{t+1}[l] \in \mathbb{R}^{d_l}$ such that $\sum l = 1^L d_l = d$. ATMA processes each layer independently, allowing for layer-specific defense mechanisms. For each layer l, ATMA performs the following steps, further captured in the edge server in the fig. 1.

1) *Pooling Client Weights:*
Gather updates $\mathbf{w}_1^{t+1}[l], \mathbf{w}_2^{t+1}[l], \ldots, \mathbf{w}_N^{t+1}[l]$ from all clients.
2) *Variance Calculation:*

$$\sigma^2[l,j] = \frac{1}{N} \sum_{i=1}^{N} \big(w_i^{t+1}[l,j] - \bar{w}^{t+1}[l,j]\big)^2 \tag{4}$$

where $\bar{\mathbf{w}}^{t+1}[l,j] = (1/N) \sum_{i=1}^{N} \mathbf{w}_i^{t+1}[l,j]$ is the mean of the jth parameter in layer l.
3) *Variance Normalization:*

$$\hat{\sigma}^2[l,j] = \frac{\sigma^2[l,j] - \min_j \sigma^2[l,j]}{\max_j \sigma^2[l,j] - \min_j \sigma^2[l,j]} + \epsilon \tag{5}$$

where ϵ is a small constant (e.g. 10^{-8}).
4) *Adaptive Alpha Calculation:*

$$\alpha[l] = \alpha_{\text{base}} + \big(\alpha_{\max} - \alpha_{\text{base}}\big) \frac{1}{d_l} \sum_{j=1}^{d_l} \hat{\sigma}^2[l,j] \tag{6}$$

where α_{base} and $\alpha_{\max}$ are the minimum and maximum trimming proportions, respectively.
5) *Trimming and Aggregation:*
For each parameter j in layer l, Sort the values $\mathbf{w}1^{t+1}[l,j], \mathbf{w}2^{t+1}[l,j], \ldots, \mathbf{w}N^{t+1}[l,j]$ in ascending order to obtain the ordered statistics $x(1)[l,j], x(2)[l,j], \ldots, x(N)[l,j]$. After which, calculate the number of clients to trim: $k[l] = \lceil \alpha[l] \cdot N \rceil$. And finally compute the trimmed mean in

$$\text{ATMA}[l,j] = \frac{1}{N - 2\,k[l]} \sum_{i=k[l]+1}^{N-k[l]} x_{(i)}\big[l,j\big]. \tag{7}$$

Fig. 2. Block showing smart contract deployment to the blockchain.

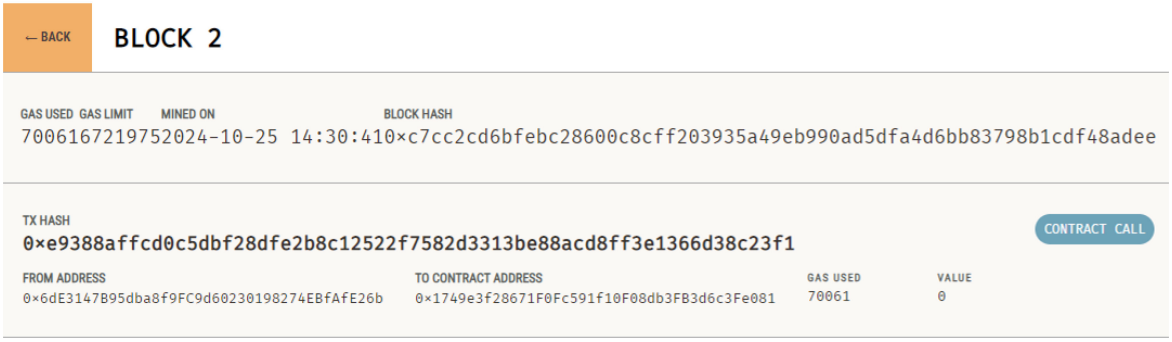


Fig. 3. Block details after server registers first client.

Fig. 1 shows the proposed framework of the proposed system. The blockchain layer serves as the system's core, providing security, immutability, and decentralized verification of the FL process. It includes a full node, and it is near the edge server. This facilitates faster validation and transaction processing, crucial for time-sensitive applications in edge IoT environments. The edge server and client devices function as light blockchain nodes, allowing them to interact with the blockchain without maintaining the entire ledger.

The ethereum virtual machine (EVM) on the blockchain executes smart contracts in Algorithm 1 that govern client registration and model verification. The smart contract manages the client registry, a whitelist of authorized participants in the FL process. When clients attempt to submit model updates, the system verifies their identity against this registry, ensuring that only pre-registered edge devices can contribute to the model training. The private blockchain validators use PoA consensus. It comprises validators that confirm transactions and maintain the integrity of the blockchain, creating new blocks that permanently record critical information about the FL process. The Algorithm 2 is a sample on interacting with the blockchain to register an edge device. Figs. 2 and 3 illustrate smart contract deployment and initial client registration on Ganache, with each client entry costing 70061 gas and recorded on-chain to ensure robust participation tracking.

Algorithm 1 Smart Contract: Register Client and Add Model Hash (Simplified)

Require: Server address: *server*
registeredClients ← {} ▷ Client registry
finalGlobalModelHash ← *null*
procedure REGISTERCLIENT(client_address)
  if client_address **not in** registeredClients **then**
    registeredClients[client_address] ← true
    emit ClientRegistered(client_address)
  else
    throw "Client already registered"
  end if
end procedure
procedure SAVEFINALMODEL(model_hash)
  if model_hash **then**
    finalGlobalModelHash ← model_hash
    emit FinalModelSaved(model_hash)
  else
    throw "Invalid model hash"
  end if
end procedure
function GETFINALMODELHASH
  return finalGlobalModelHash **or throw** "No model saved"
end function

A. Dataset Description

This study evaluates the proposed intrusion detection framework on three datasets: 1) CICIoV2024; 2) Edge-IIoTset; and 3) ForgeIIOT Pro [40]. CICIoV2024 contains CAN-BUS vehicle data with DoS and spoofing attacks. Edge-IIoTset covers 14 IoT/IIoT attack types and supports centralized and FL. ForgeIIoT Pro focuses on IIoT with 17 attack classes and 1M samples featuring realistic industrial scenarios. Two sampling strategies created IID and non-IID data: 1) IID data was randomly shuffled and 2) evenly distributed across clients. In contrast, non-IID data was partitioned by attack class to simulate client heterogeneity. This setup tested the

Algorithm 2 Interaction With Smart

w3: Web3 instance, contract_address: address, contract_abi: ABI, private_key: string Transaction receipt with hash

```
procedure REGISTERCLIENT(w3, contract_address, contract_abi, private_key)
    contract ← w3.eth.contract(address=contract_address, abi=contract_abi)
    account ← w3.eth.accounts[0]
    client_address ← w3.eth.accounts[1]
    nonce ← w3.eth.get_transaction_count(account)
    txn ← contract.functions.registerClient(client_address)
        .build_transaction({
        'chainId': 1337,
        'gas': 2000000,
        'gasPrice': to_wei(20, 'gwei'),
        'nonce': nonce,
        'from': account})
    signed_txn ← w3.eth.account.sign_transaction(txn, private_key=private_key)
    txn_hash ← w3.eth.send_raw_transaction(signed_txn.raw_transaction)
    txn_receipt ← w3.eth.wait_for_transaction_receipt(txn_hash)
    return txn_receipt.transactionHash.hex()
end procedure
```

model's robustness and adaptability under realistic, skewed data distributions.

B. Experimental Setup

The proposed framework was evaluated with a custom Python and Flask setup. The model training and aggregation processes were implemented using TensorFlow 2.x, running on a high-performance system with 128 GB of RAM and dual NVIDIA A100 GPUs. This configuration allowed efficient model training and weight aggregation across multiple communication rounds.

Ganache was used as a private Ethereum-compatible blockchain simulator to secure and trace model updates within the FL system. Blockchain interactions, including update tracking and verification, were handled via web3.py. Solidity smart contracts deployed on Ganache logged and verified client model updates, ensuring contribution integrity. After initializing the global model weights, the server distributes them to all 25 clients for local training. It collects updated weights serialized as lightweight JSON files containing unique client IDs for tracking. The server deserializes and aggregates these weights to update the global model, which is then validated on a server-side dataset to ensure improved accuracy and robustness, enhancing generalization across heterogeneous client data.

The updated global model weights, accuracy, and loss validation metrics are serialized into JSON and returned to clients for subsequent federated training rounds. This framework integrates FL, JSON serialization, and blockchain verification to ensure secure, scalable, and privacy-preserving model updates. Training employs categorical cross-entropy loss, the Adam optimizer with adaptive learning rate, ten local epochs, five communication rounds, and nine simulated heterogeneous Edge-IoT clients. The neural network is a sequential multiclass classifier with a Gaussian noise input layer for robustness, three dense ReLU layers (60, 30, 15 neurons) for feature extraction, and a softmax output layer for classification.

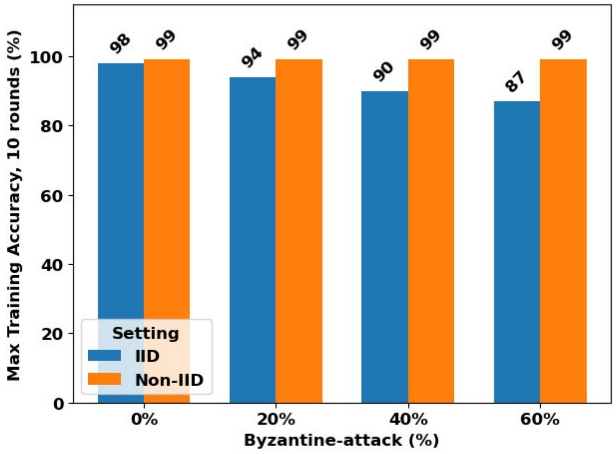


Fig. 4. Performance comparison under varying Byzantine attack rates for Fog dataset across IID and non-IID data distributions

Byzantine attack simulations were conducted varying the proportion of compromised clients among 25 participants (0% to 60%) using random sampling without replacement. The Byzantine attacks comprised of label flipping and feature perturbation via Gaussian noise injection with parameters $\mu = 0,\ \sigma = 0.5$. This dual-vector attack simultaneously disrupts decision boundaries and feature representations, intensifying adversarial challenges beyond single-vector methods. The robustness of ATMA was assessed under both IID and non-IID data distributions.

IV. Results and Discussion

Table II shows that the ATMA approach achieves accuracy of 91.8% for IID and 88.4% for non-IID, while maintaining high Byzantine tolerance similar to Bulyan but with significantly better computational complexity ($O(n\log n)$ versus $O(n^3)$). This is particularly noteworthy given the challenging nature of the Edge-IIoTset dataset, which contains 14 different attack types across 1.9 million records, as mentioned in the search results.

The baseline FedAvg method shows the highest accuracy under IID conditions (92.3%) but performs poorly with non-IID data (78.5%) and offers low Byzantine tolerance, highlighting its vulnerability to adversarial attacks. Krum and Median provide moderate Byzantine tolerance with reasonable accuracy but cannot match the Byzantine resilience of Bulyan and our ATMA approach. Our method's strong performance on non-IID data (88.4%) is particularly significant given the highly imbalanced nature of the Edge-IIoTset dataset, where attack classes like MITM and Fingerprinting represent just 0.02% and 0.04% of the data, respectively. This demonstrates ATMA's ability to handle the practical challenges of real-world Edge-IoT environments where data heterogeneity is common and its not complex.

Figs. 4 and 5 illustrate our ATMA framework's performance under varying Byzantine attack rates (0%–60%) across IID and

TABLE II
PERFORMANCE COMPARISON OF ROBUST AGGREGATION METHODS UNDER IID AND NON-IID SETTINGS

Aggregation Method	Accuracy (IID)	Accuracy (Non-IID)	Byzantine Tolerance	Computational Complexity
FedAvg (baseline)	92.3%	78.5%	Low	$O(n)$
Krum	89.7%	82.1%	Medium	$O(n^2)$
Median	90.1%	84.3%	Medium	$O(n \log n)$
Bulyan	91.2%	85.7%	High	$O(n^3)$
Adaptive TMA (Ours)	91.8%	88.4%	High	$O(n \log n)$

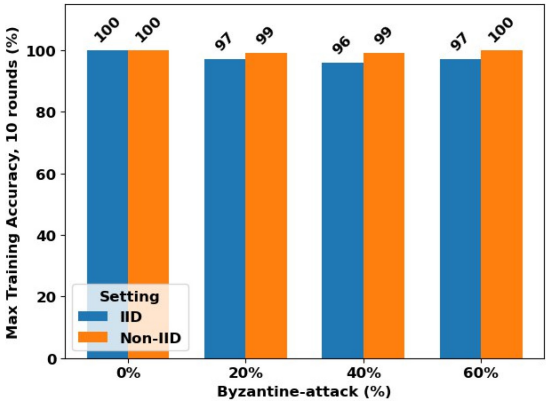


Fig. 5. Performance comparison under varying Byzantine attack rates for CICIoV dataset across IID and non-IID data distributions.

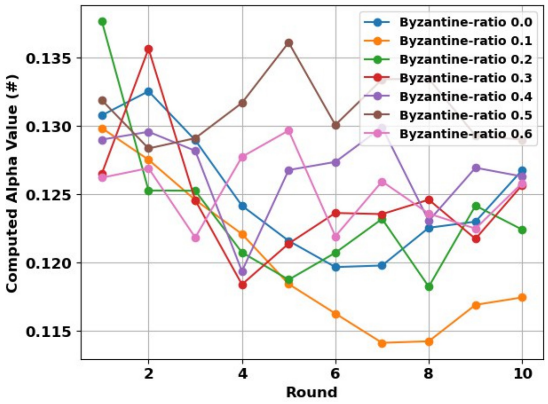


Fig. 6. Computed alpha values per round for different Byzantine ratios, showing ATMA's adaptive response to attack intensity.

non-IID data distributions for the Fog and CCIOV datasets. In the Fog dataset (Fig. 5), IID accuracy decreases progressively from 98.40% (0% attacks) to 87.35% (60% attacks), while non-IID accuracy remains remarkably stable (99.40%–99.99%) across all attack rates. Similarly, the CCIOV dataset (Fig. 6) demonstrates robust performance, with IID accuracy declining minimally from 100% to 97.01% as attack rates increase, while non-IID accuracy maintains 100% throughout all attack scenarios. This counterintuitive resilience in non-IID environments can be attributed to ATMA's adaptive trimming mechanism, which effectively leverages the natural variance in heterogeneous data to distinguish between legitimate distribution differences and malicious updates.

The layer-wise variance analysis adapts trimming thresholds dynamically, as further illustrated in Figs. 6 and 7. Fig. 6 presents the computed alpha values per round for different

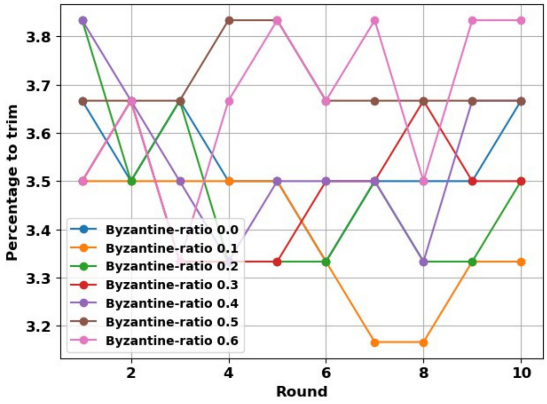


Fig. 7. Percentage of clients trimmed each round under different Byzantine attack rates, demonstrating ATMA's dynamic adjustment to adversarial conditions.

Byzantine ratios, showing that ATMA increases the trimming parameter in response to higher attack rates, especially as variance in client updates rises.

Fig. 7 displays the corresponding percentage of clients trimmed each round, confirming that higher Byzantine presence leads to more aggressive exclusion of outlier updates. The dynamic trends in alpha values and trimming percentages demonstrate that ATMA continuously adapts its aggregation strategy, maintaining high robustness even as adversarial conditions fluctuate. Together, these results show that properly designed adaptive aggregation methods like ATMA can transform data heterogeneity from a challenge into a defensive advantage for Byzantine-resilient FL.

Fig. 8 demonstrates our ATMA framework's performance under severe Byzantine conditions (60% malicious clients) with IID data distribution in Fog dataset. Despite this challenging scenario, the model achieves remarkable stability, with training accuracy steadily increasing from 78.2% to 87.3% across ten rounds while validation accuracy rises from 79.1% to 88.4%. The corresponding loss values show consistent improvement, with training loss decreasing from 48.7% to 29.3% and validation loss from 43.5% to 25.1%. Despite most clients behaving maliciously, this robust performance highlights ATMA's exceptional Byzantine resilience through its adaptive trimming mechanism.

Fig. 9 presents ATMA's performance under the same 60% Byzantine attack rate but with non-IID data distribution (CICIoV dataset). Remarkably, the model maintains near-perfect accuracy throughout training, with both training and validation accuracy consistently above 99.8%.

Loss values remain exceptionally low (below 0.005) as in the fig. 9 across all rounds. This counterintuitive result, superior performance in non-IID versus IID conditions under

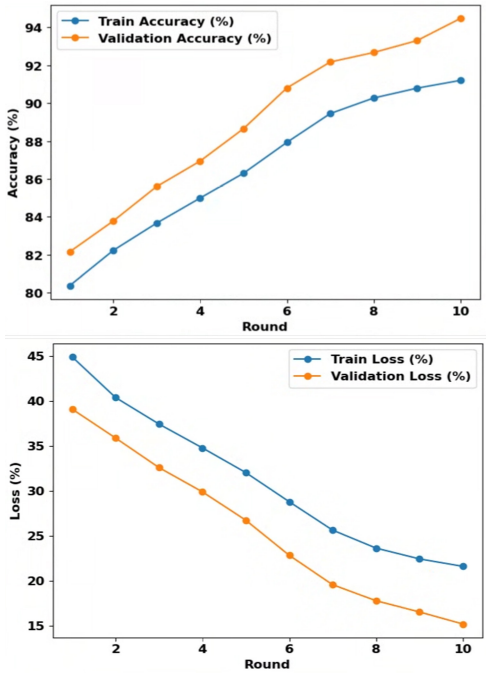


Fig. 8. ATMA performance under 60% Byzantine attack rate with IID data distribution, showing training and validation accuracy/loss across ten rounds for the Fog dataset.

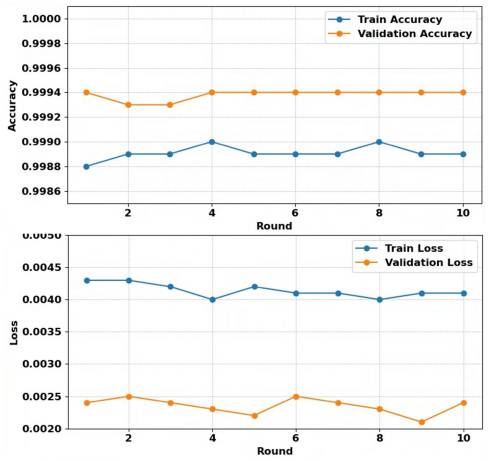


Fig. 9. ATMA performance under 60% Byzantine attack rate with IID data distribution, showing training and validation accuracy/loss across ten rounds for the CICIoV dataset.

identical attack rates, suggests that ATMA's adaptive variance-based trimming effectively leverages the natural heterogeneity in non-IID data to better distinguish between legitimate client diversity and malicious interference, demonstrating a unique advantage in realistic FL environments where data heterogeneity is the norm.

V. Conclusion

The Blockchain-enhanced FL IDS with ATMA is a practical approach for handling non-IID data in Edge-IoT networks. Results demonstrate its effectiveness for handling IID and non-IID data in Edge-IoT networks. The experimental results across three datasets (CICIoV2024, Edge-IIoTset, and ForgeIIOT Pro) highlight the framework's remarkable resilience, particularly in non-IID scenarios where it maintained near-perfect accuracy even under 60% Byzantine attack rates. Adopting blockchain for secure client authentication and immutable model verification with ATMA's dynamic trimming mechanism creates a robust defense against sophisticated adversarial threats while preserving computational efficiency. The layer-wise variance analysis in ATMA transforms data heterogeneity from a challenge into an advantage, enabling the system to distinguish between legitimate distribution differences and malicious updates. One of the limitations of the current implication, is the need for a centralized validation dataset, which may not be feasible in all edge environments. Additionally, while efficient compared to alternatives, the $O(n\log n)$ complexity could still present challenges in extremely resource-constrained settings.

Future work should explore fully decentralized validation mechanisms, investigate ATMA's performance with deeper neural architectures, and extend the framework to handle concept drift in streaming data environments. Additionally, exploring privacy-preserving techniques like differential privacy in conjunction with ATMA could further enhance the security guarantees of the system while maintaining its Byzantine resilience.

References

[1] L. A. C. Ahakonye, C. I. Nwakanma, and D.-S. Kim, "Tides of blockchain in IoT cybersecurity," *Sensors*, vol. 24, no. 10, p. 3111, 2024.

[2] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning," *IEEE Access*, vol. 10, pp. 40281–40306, 2022.

[3] S. Wang et al., "Adaptive federated learning in resource constrained edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1205–1221, Jun. 2019.

[4] Y. Ye, S. Li, F. Liu, Y. Tang, and W. Hu, "EdgeFed: Optimized federated learning based on edge computing," *IEEE Access*, vol. 8, pp. 209191–209198, 2020.

[5] A. Eghmazi, M. Ataei, R. J. Landry, and G. Chevrette, "Enhancing IoT data security: Using the blockchain to boost data integrity and privacy," *IoT*, vol. 5, no. 1, pp. 20–34, 2024.

[6] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, "Machine learning-based network vulnerability analysis of industrial Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6822–6834, Aug. 2019.

[7] S. Selvarajan et al., "An artificial intelligence lightweight blockchain security model for security and privacy in IIoT systems," *J. Cloud Comput.*, vol. 12, no. 1, p. 38, 2023.

[8] A. Saqib, L. Qianmu, and Y. Abdullah, "Blockchain and federated learning-based intrusion detection approaches for edge-enabled industrial IoT networks: A survey," *Ad Hoc Netw.*, vol. 152, Jan. 2024, Art. no. 103320.

[9] L. A. C. Ahakonye, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Trees bootstrap aggregation for detection and characterization of IoT-SCADA network traffic," *IEEE Trans. Ind. Informat.*, vol. 20, no. 4, pp. 5217–5228, Apr. 2024.

[10] L. A. C. Ahakonye, G. C. Amaizu, C. I. Nwakanma, J. M. Lee, and D.-S. Kim, "Classification and characterization of encoded traffic in SCADA network using hybrid deep learning scheme," *J. Commun. Netw.*, vol. 26, no. 1, pp. 65–79, Feb. 2024.

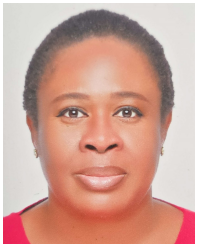
[11] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning," *IEEE Netw.*, vol. 33, no. 5, pp. 156–165, Sep./Oct. 2019.

[12] H. G. Abreha, M. Hayajneh, and M. A. Serhani, "Federated learning in edge computing: A systematic survey," *Sensors*, vol. 22, no. 2, p. 450, 2022.

[13] V. K. Prasad et al., "Federated learning for the Internet-of-Medical-Things: A survey," *Mathematics*, vol. 11, no. 1, p. 151, 2022.
[14] D. C. Nguyen et al., "Federated learning for smart healthcare: A survey," *ACM Comput. Surv.*, vol. 55, no. 3, pp. 1–37, 2022.
[15] H. Kaur, V. Rani, M. Kumar, M. Sachdeva, A. Mittal, and K. Kumar, "Federated learning: A comprehensive review of recent advances and applications," *Multimedia Tools Appl.*, vol. 83, pp. 54165–54188, May 2024.
[16] J. Wen, Z. Zhang, Y. Lan, Z. Cui, J. Cai, and W. Zhang, "A survey on federated learning: Challenges and applications," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 2, pp. 513–535, 2023.
[17] H. Fan, C. Huang, and Y. Liu, "Federated learning-based privacy-preserving data aggregation scheme for IIoT," *IEEE Access*, vol. 11, pp. 6700–6707, 2023.
[18] Z. A. E. Houda, B. Brik, A. Ksentini, L. Khoukhi, and M. Guizani, "When federated learning meets game theory: A cooperative framework to secure IIoT applications on edge computing," *IEEE Trans. Ind. Informat.*, vol. 18, no. 11, pp. 7988–7997, Nov. 2022.
[19] J. KonecnÍ, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," 2016, *arXiv:1610.02527*.
[20] G. A. Baumgart, J. Shin, A. Payani, M. Lee, and R. R. Kompella, "Not all federated learning algorithms are created equal: A performance evaluation study," 2024, *arXiv:2403.17287*.
[21] D. C. Nguyen et al., "Federated learning meets blockchain in edge computing: Opportunities and challenges," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12806–12825, Aug. 2021.
[22] A. Makkar, T. W. Kim, A. K. Singh, J. Kang, and J. H. Park, "SecureIIoT environment: Federated learning empowered approach for securing IIoT from data breach," *IEEE Trans. Ind. Informat.*, vol. 18, no. 9, pp. 6406–6414, Sep. 2022.
[23] X. Wang et al., "Toward accurate anomaly detection in industrial Internet of Things using hierarchical federated learning," *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7110–7119, May 2022.
[24] W. Fan, K. Yang, Y. Wang, C. Chen, and J. Li, "Data-free adaptive structured pruning for federated learning," *J. Supercomput.*, vol. 80, pp. 18600–18626, May 2024.
[25] W. Zhu, M. Goudarzi, and R. Buyya, "FLight: A lightweight federated learning framework in edge and fog computing," *Softw., Pract. Exp.*, vol. 54, no. 5, pp. 813–841, 2024.
[26] S. Kalra, J. Wen, J. C. Cresswell, M. Volkovs, and H. R. Tizhoosh, "Decentralized federated learning through proxy model sharing," *Nat. Commun.*, vol. 14, no. 1, p. 2899, 2023.
[27] D. Yin, Y. Chen, R. Kannan, and P. Bartlett, "Byzantine-robust distributed learning: Towards optimal statistical rates," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 5650–5659.
[28] F. Colosimo and F. De Rango, "Median-Krum: A joint distance-statistical based Byzantine-robust algorithm in federated learning," in *Proc. Int. ACM Symp. Mobile Manag. Wireless Access*, 2023, pp. 61–68.
[29] Z. Wang and P. Zhao, "Byzantine detection for federated learning under highly non-IID Data and majority corruptions," *Wireless Netw.*, vol. 31, pp. 865–877, Jun. 2024.
[30] J. Lavond, M. Cheng, and Y. Li, "Trusted aggregation (TAG): Backdoor defense in federated learning," *Trans. Mach. Learn. Res.*, pp. 1–21, Nov. 2024. [Online]. Available: https://openreview.net/forum?id=r9eNUDe2im
[31] Y. Li, X. Wang, F. Yu, L. Sun, W. Zhang, and X. Wang, "FedCAP: Robust federated learning via customized aggregation and personalization," 2024, *arXiv:2410.13083*.
[32] J. Kim, G. Kim, and B. Han, "Multi-level branched regularization for federated learning," in *Proc. 39th Int. Conf. Mach. Learn.*, 2022, pp. 11048–11064.
[33] K. Nishimoto, Y.-H. Chiang, H. Lin, and Y. Ji, "FedATM: Adaptive trimmed mean-based federated learning against model poisoning attacks," in *Proc. IEEE 97th Veh. Technol. Conf. (VTC)*, 2023, pp. 1–5.
[34] V. T. Truong, D. N. Hoang, and L. B. Le, "BFLMeta: Blockchain-empowered metaverse with Byzantine-robust federated learning," in *Proc. IEEE Glob. Commun. Conf.*, 2023, pp. 5537–5542.
[35] K. Begum, M. A. I. Mozumder, M.-I. Joo, and H.-C. Kim, "BFLIDS: Blockchain-driven federated learning for intrusion detection In IoMT networks," *Sensors*, vol. 24, no. 14, p. 4591, 2024.
[36] A. Almaghthawi, E. A. Ghaleb, N. A. Akbar, L. Asiri, M. Alrehaili, and A. Altalidi, "Federated-learning intrusion detection system based blockchain technology," *Int. J. Online Biomed. Eng.*, vol. 20, no. 11, p. 16, 2024.
[37] M. A. Manolache, S. Manolache, and N. Tapus, "Decision making using the blockchain proof of authority consensus," *Procedia Comput. Sci.*, vol. 199, pp. 580–588, Jan. 2022.
[38] R. Bensaid, N. Labraoui, A. A. A. Ari, H. Saidi, J. H. M. Emati, and L. Maglaras, "SA-FLIDS: Secure and authenticated federated learning-based intrusion detection system for IoMT-enabled smart healthcare," *PeerJ Comput. Sci.*, vol. 10, Dec. 2024, Art. no. e2414.
[39] N. Sun, W. Wang, Y. Tong, and K. Liu, "Blockchain based federated learning for intrusion detection for Internet of Things," *Front. Comput. Sci.*, vol. 18, no. 5, 2024, Art. no. 185328.
[40] P. Kumar, S. Mullick, R. Das, A. Nandi, and I. Banerjee, "IoTForge pro: A security testbed for generating intrusion dataset for industrial IoT," *IEEE Internet Things J.*, vol. 12, no. 7, pp. 8453–8460, Apr. 2025.
[41] Q. Zhou, X. Mao, and Y. Chen, "DDoS attack detection method combining federated learning and hybrid deep learning in software-defined networking," *Comput. J.*, Apr. 2025, Art. no. bxaf049, doi: 10.1093/comjnl/bxaf049.
[42] Q. Tian, C. Guang, C. Wenchao, and W. Si, "A lightweight residual networks framework for DDoS attack classification based on federated learning," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, 2021, pp. 1–6.
[43] V. T. Truong, L. Le, and D. Niyato, "Blockchain meets metaverse and digital asset management: A comprehensive survey," *IEEE Access*, vol. 11, pp. 26258–26288, 2023.
[44] A. Zainudin, L. A. C. Ahakonye, R. Akter, D.-S. Kim, and J.-M. Lee, "An efficient hybrid-DNN For DDoS detection and classification in software-defined IIoT networks," *IEEE Internet Things J.*, vol. 10, no. 10, pp. 8491–8504, May 2023.
[45] B. Bütün, A. T.-J. Akem, M. Gucciardo, and M. Fiore, "Fast detection of cyberattacks on the metaverse through user-plane inference," in *Proc. IEEE Int. Conf. Metaverse Comput., Netw. Appl. (MetaCom)*, 2023, pp. 350–354.
[46] S.-Y. Kuo, F.-H. Tseng, and Y.-H. Chou, "Metaverse intrusion detection of wormhole attacks based on a novel statistical mechanism," *Future Gener. Comput. Syst.*, vol. 143, pp. 179–190, Jun. 2023.
[47] A. Zainudin, R. Akter, D.-S. Kim, and J.-M. Lee, "Federated learning inspired low-complexity intrusion detection and classification technique for SDN-based industrial CPS," *IEEE Trans. Netw. Service Manag.*, vol. 20, no. 3, pp. 2442–2459, Sep. 2023.
[48] Z. Liu, K. Zheng, L. Hou, H. Yang, and K. Yang, "A novel blockchain-assisted aggregation scheme for federated learning In IoT networks," *IEEE Internet Things J.*, vol. 10, no. 19, pp. 17544–17556, Oct. 2023.
[49] Z. A. E. Houda, A. S. Hafid, and L. Khoukhi, "MiTFed: A privacy preserving collaborative network attack mitigation framework based on federated learning using SDN and blockchain," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 4, pp. 1985–2001, Jul./Aug. 2023.
[50] C. Que and F. N. Khan, "A scalable federated learning-based approach for accurate traffic prediction in edge computing-enabled metro optical networks," *Comput. Ind. Eng.*, vol. 203, May 2025, Art. no. 111004.

Kalibbala Jonathan Mukisa received the B.Sc. degree in computer engineering from Kyungdong University, Goseong, South Korea, in 2023. He is currently pursuing the master's degree in IT-convergence engineering with the Networked Systems Laboratory, Kumoh National Institute of Technology, Gumi, South Korea.

He currently collaborates with Otic-Uganda. His interdisciplinary research covers blockchain-based offline transactions, NFT-enabled inventory management, simulated islanded microgrids, federated demand forecasting with Byzantine-resilient aggregation, TRIZ-inspired compression for large language models, and multilingual AI educational platforms, focused on securing decentralized systems, optimizing energy/data workflows, and enabling scalable personalized learning.



Love Allen Chijioke Ahakonye (Member, IEEE) received the B.Sc. degree in mathematics/computer science from the University of Port Harcourt, Port Harcourt, Nigeria, in 2001, the M.Sc. degree in information technology from the Federal University of Technology, Akure, Nigeria, in 2016, and the Ph.D. degree in IT-convergence engineering from the Networked System Laboratory, IT-Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea, in 2024.

She has over a decade of experience with the Nigerian oil and gas sector as a Network and System Administrator from 2002 to 2016. In 2017, she briefly worked as a Logistics Superintendent with the Nigerian Petroleum Development Company until 2019. She is also a Postdoctoral Researcher with the ICT Convergence Research Center, Kumoh National Institute of Technology. She is a Computer Professionals Registration Council of Nigeria. Her research interest is artificial intelligence application to the Internet of Things (IoT) and industrial IoT, supervisory control and data acquisition and industrial control system for intrusion/anomaly detection, cyber-security, fault detection, XAI, and manufacturing execution systems.

Dong-Seong Kim (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2003.

From 1994 to 2003, he worked as a full-time Researcher with the ERC-ACI, Seoul National University. From March 2003 to February 2005, he served as a Postdoctoral Researcher with the Wireless Network Laboratory, School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA. From 2007 to 2009, he was a Visiting Professor with the Department of Computer Science, University of California, Davis, CA, USA. He served as the Dean of IACF from 2019 to 2022. He is currently a Professor with the Department of IT Convergence Engineering, School of Electronic Engineering, Kumoh National Institute of Technology, Gumi, South Korea. He is also the Director of the KIT Convergence Research Institute and the ICT Convergence Research Center (ITRC and NRF advanced research center program), supported by the Korean Government at Kumoh National Institute of Technology, and the Director of NSLab Company, Ltd. His primary research interests include real-time IoT and smart platforms, industrial wireless control networks, networked embedded systems, Fieldbus, metaverse, and blockchain. Prof. kim is a Senior Member of ACM.

Jae-Min Lee (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2005.

From 2005 to 2014, he was a Senior Engineer with Samsung Electronics, Suwon, South Korea. From 2015 to 2016, he was a Principal Engineer with Samsung Electronics. Since 2017, he has been an Associate Professor with the School of Electronic Engineering and the Department of IT Convergence Engineering with Kumoh National Institute of Technology, Gyeongbuk, South Korea. Since 2024, he has been the Director of the Smart Defense Logistics Innovation Convergence Research Center (ITRC Program). He is the Executive Director of the Korean Institute of Communications and Information Sciences. His current main research interests are smart IoT convergence application, industrial wireless control network, UAV, metaverse, and blockchain.