Federated Learning for Privacy-preserving Internet of Things (IoT) Security: A Decentralized Intrusion Detection Framework

Muhammad Shaharyar Ramzan [1], Arshad Ali [2], and Rafi us Shan [3,*]

[1] Department of Information Technology, Faculty of Computer Science and IT, Superior University, Lahore, Pakistan
[2] Faculty of Computer and Information Systems, Islamic University of Madinah, Almadinah Almunawarah, Saudi Arabia
[3] Faculty of Computer Information Science, Higher Colleges of Technology, Dhabi, United Arab Emirates.
Email: shaharyarramzan56@gmail.com (M.S.R.); a.ali@iu.edu.sa (A.A.); rshan@hct.ac.ae (R.u.S.)
*Corresponding author

***Abstract*—The proposed study suggests a to help cope with issues related to cybersecurity in Internet of Things and Industrial Internet of Things environments without compromising privacy. The proposed framework introduces several innovative features, such as federated learning with momentum-based optimization, adaptive differential privacy, trust verification via blockchain, and Byzantine-resilient aggregation, to enhance the security, scalability, and robustness of the system compared with traditional intrusion detection systems. It also integrates supervised classification with autoencoder-based anomaly detection to detect existing and emerging cyberattacks. The proposed system was assessed with respect to the extended Industrial Internet of Things Intrusion Dataset (X-IIoT) and Network-Based Botnet Attack detection for IoT (N-BaIoT) benchmark datasets, where the environments were simulated as federated ones. The accuracy of the Hybrid Robust Federated Intrusion Detection System increased to 97.15% on X-IIoT and 97.64% on N-BaIoT with only 41 communication rounds and was resilient against up to 20% of Byzantine clients. These results showcase its efficacy to secure, private and communication-efficient intrusion detection for next generation Internet of Things and Hybrid Robust Federated Intrusion Detection System networks.**

***Keywords*—Internet of Things (IoT), industrial internet of things, federated learning, intrusion detection system**

I. Introduction

One factor contributing to the rapid growth in the popularity of the Internet of Things (IoT) is the digitalization of modern industries and homes. As we move towards smart manufacturing systems in Industry 4.0, the Internet of Medical Things enables devices to continuously generate high-velocity telemetry, serving as a source of automation, real-time control, and intelligent decision-making [1]. Despite these advantages, the extensive interlinking of heterogeneous devices multiplies the cybersecurity risks in the IoT environment. Many Internets of Things devices, due to the limited computing resources available, which include limited memory, computing power, and battery, cannot support complex security measures, such as advanced cryptographic protocols or a real-time intrusion detection system, because they would require more computing resources. Consequently, such devices tend to offer an easy point of entry for cyberattacks, increasing the attack surface across vital systems such as healthcare facilities, industrial facilities, smart homes, and energy systems.

A. Evolution of the IoT Threat Landscape

The threat landscape of IoT environments has changed significantly. Previous attacks were mostly single denial-of-service incidents, but current cyberattacks have evolved into more-organized and complex attacks. In the present attack campaigns, it is often a multistage operation that includes botnet coordination, privilege escalation, lateral movement across interconnected systems, and ongoing intrusion into Industrial Control Systems [2]. Several attackers exploit lax authentication protocols, underlying firmware vulnerabilities, and unsecure communication channels to gain unauthorized access to IoT infrastructure.

The Internet of Things hardware platform and firmware versions also make it difficult to implement a uniform security mechanism because the heterogeneity of these products creates fragmented security environments that are difficult to control and secure. Such vulnerabilities are concerning because IoT networks are widely embedded in vital systems such as smart grids, healthcare systems, transportation systems, and automated factories. Hacking can result in data loss, work stoppages, and physical harm. Consequently, the development of adaptable, real-time, and resilient intrusion detection systems has become a vital concern in current cybersecurity systems [3].

Manuscript received March 27, 2026; revised May 10, 2026; accepted May 22, 2026; published August 25, 2026.

B. Limitations of Centralized Cloud-Based Detection

Conventional intrusion detection systems are usually implemented in a centralized design, in which telemetry data from distributed IoT devices are sent to cloud servers for analysis and threat identification. Cloud-based systems offer several benefits, including almost limitless computational capacity, scalable storage, and the ability to train highly complex deep learning models using large datasets. There are multiple constraints to centralized methods for finding in large-scale IoT systems. First, the constant broadcasting of raw network-wide traffic between edge devices and centralized cloud servers creates a high communication overhead and consumes a significant amount of bandwidth. Second, the architecture may lead to large detection latencies, which are undesirable in mission-critical systems that require real-time surveillance. Third, the infrastructure is centralized, creating points of failure that adversaries can easily exploit, thereby compromising the entire system. With the ever-increasing adoption of the Internet of Things and exponentially growing volumes of telemetry data, the constant flow of raw traffic data to cloud servers can be considered technically inefficient and financially unrealistic [4]. These challenges have inspired a shift toward decentralized, edge-centric security systems, where data processing and anomaly detection are performed closer to the data source. These architectures reduce communication overhead, decrease response time, and improve system resilience.

C. Federated Learning: A Decentralized Security Paradigm

Recently, Federated Learning has emerged as a promising paradigm for decentralized machine learning in distributed settings, such as Internet of Things networks. In contrast to conventional centralized learning methods, federated learning enables multiple devices to jointly learn a global model while preserving the privacy of their local training data. Participating clients do not share raw data; instead, model updates or gradient parameters are shared with an aggregation server in the center. The risks of information privacy leaks and communication overhead are greatly reduced in this type of collaborative learning model, which helps organizations comply with stringent data protection laws in sensitive sectors such as healthcare and industrial control systems. Moreover, federated learning enables international models to benefit from diverse data distributions generated by heterogeneous devices operating in different settings. Consequently, FL-based systems can enhance the generalization and indexing of deviant operations in distributed IoT networks [5].

D. Privacy Challenges in Federated Learning

Although federated learning is regarded as a privacy-preserving model, it does not guarantee complete protection against information leakage. As shown in the literature, common gradients or model updates can also reveal sensitive information owing to attacks such as membership inference and gradient inversion. In attacks of this type, the attackers aim to rebuild some of the original training data or determine whether certain data samples were used to train the models.

Multiple privacy-enhancing techniques have been proposed to mitigate these vulnerabilities, including homomorphic encryption, secure aggregation, and differential privacy. These methods aim to generate model updates at optimal points before transmission, either by introducing carefully controlled noise or by leveraging encrypted computations. Although these approaches are vital for safeguarding privacy, they may increase computational overhead and undermine model accuracy because of noise. This constitutes a privacy performance trade-off, which is a central challenge in building secure federated learning systems for IoT intrusion detection [6].

E. Adversarial Robustness and Model Poisoning

Another major challenge in federated learning environments is the presence of poor or malicious clients. In federated learning frameworks, it is assumed that participating clients are honest during local model training. Nevertheless, malicious clients can also, in theory, provide manipulated model updates to interfere with the global learning procedure or create covert back doors in the collected model. This type of attack is usually referred to as a Byzantine or model poisoning attack. Infrastructure currency is important for security-sensitive protection. To counter such threats, recent studies have investigated the incorporation of blockchain technology into federated learning models. Blockchain provides a decentralized, tamper resistant log that records updates made by models, with verifiable timestamps and cryptographic authentication. Federated learning systems built on blockchain and distributed consensus can improve trust and transparency and reduce the risk of malicious model updates.

F. Optimization Challenges in Heterogeneous Internet of Things Environments

Another notable problem with federated learning in IoT networks is that the data produced by heterogeneous devices are non-independent and non-identically distributed. Surveillance cameras generate data streams at high visual frequencies, whereas environmental sensors generate data streams at low frequencies. Such non-homogeneous data distributions can cause local model updates to be inconsistent and global optimization to be unstable during aggregation. Traditional aggregation algorithms, such as Federated Averaging, assume approximately equal and Independent and Identically Distributed data distributions across clients. However, federated averaging can exhibit slow convergence, oscillatory behavior during training, and decreased model accuracy in highly heterogeneous environments. To overcome these shortcomings, several optimization methods, including momentum-based aggregation, have been proposed to stabilize gradient updates and expedite convergence in distributed learning setups [7].

G. Explainability and Transparency in IDS

An Intrusion Detection System (IDS) must offer both accurate predictions and results that can be understood, rather than mere guesses. In areas of vital necessity, such as healthcare and industrial automation, there is a real need to provide a description of what it did and why. Black-box machine learning models can raise operational and ethical issues when administrators are unable to understand the reasons for a detected anomaly. In the case where an Intrusion Detection System flags medical device network traffic as malicious, one of the system operators should have a clear picture of the root cause to ensure that corrective measures are implemented. Thus, the problem of explainable Artificial Intelligence methods has been proposed to increase the transparency of machine learning-based intrusion detection systems. Security analysts can use feature attribution, saliency maps, and model-agnostic explanation techniques to determine which network features or behavioral patterns lead to a particular detection decision. Combined with other federated learning systems, these interpretability mechanisms should operate without releasing sensitive local information, thereby preserving transparency and privacy of the data.

H. Proposed Framework and Contributions

To overcome these challenges, this study proposes a Hybrid Robust Federated Intrusion Detection System for secure and scalable Internet of Things and Industrial Internet of Things settings. The proposed framework combines federated learning, blockchain-based verification, and adaptive privacy protection mechanisms to enhance data security and system reliability. In particular, it enables a momentum-based aggregation method inspired by current federated optimization schemes to achieve more stable convergence given non-independent and identically distributed data. To make a more resilient learning process, the framework was enhanced based on a blockchain-based verification process and Byzantine-resilient aggregation process to tackle the effects of model poisoning and the malicious behavior of clients.

In addition, the framework utilizes a hybrid detection architecture that combines supervised classification and autoencoder-based anomaly detection to further enhance the detection of unseen or zero-day attacks. Comprehensive experimental tests on the X-IIoTID and N-BaIoT benchmark datasets indicate that the proposed Hybrid Robust Federated Intrusion Detection System framework can achieve close-to-centralized detection accuracy, provide considerable reductions in communication overheads, and offer high resilience against adversarial attacks in resource-constrained Internet of Things systems.

I. Research Novelty and Technical Contributions

Although a few federated intrusion detection frameworks for Internet of Things and Industrial Internet of Things networks have been proposed, most available studies have focused on isolated aspects such as privacy, optimization efficiency, adversarial robustness, and anomaly detection. The proposed Hybrid Robust Federated Intrusion Detection System does not exploit a completely new federated learning algorithm because it introduces a systematic integration framework that combines different complementary mechanisms, such as federated optimization with momentum, adaptive differential privacy, blockchain for verification, Byzantine-robust aggregation, and hybrid supervised-unsupervised intrusion detection in a unified manner, which is quite practical. The most distinctive aspect of this study is the coordinated deployment and assessment of these techniques in a federated intrusion detection system architecture tailored to the context of multi-type, non-independent, and identically distributed IoT systems. The framework also explores how these elements relate in the context of privacy preservation, convergence stability, communication efficiency, adversarial resilience, and zero-day attack detection in realistic federated scenarios.

II. Literature Review

The rapid growth of the Internet of Things and Industrial Internet of Things has considerably increased the scale of attacks on current cyber-physical systems. Millions of interlinked devices continue to generate massive amounts of network traffic and operational data, creating new opportunities for advanced cyber-attacks. A conventional centralized Intrusion Detection System uses raw traffic data collected at a central server to train detection models. Although centralized learning methods may achieve high detection accuracy, they have drawbacks such as high communication overhead, privacy concerns, and vulnerability to a single point of failure. These restrictions are especially important in large-scale distributed systems, such as industrial automation systems, smart healthcare infrastructure, and edge computing networks, where critical operational data cannot be shared at will.

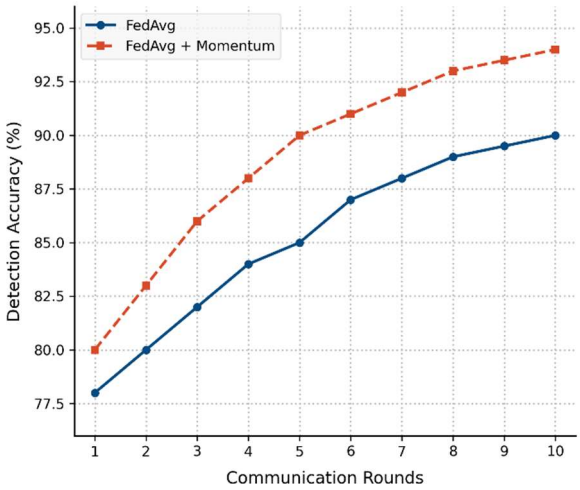

Fig. 1. Comparison of the performance of FedAvg and momentum aggregation.

Consequently, researchers have attempted to address these problems by increasing their interest in Federated Learning, a decentralized machine learning model that enables collaborative model training without exchanging

original data. Federated learning models are locally trained by individual clients at an endpoint in federated learning, who provide only model parameters or gradients to a central aggregation server and mix these parameters and gradients with the private dataset of additional clients. The method is much more effective at preserving data privacy and scalability and mitigating regulatory retreat, which is especially appropriate in distributed IoT ecosystems. Non-independent and identically distributed data distributions, communication inefficiency, gradient-based privacy intrusion, and adversarial or poisoning attacks are, however, the new challenges brought by federated learning. Therefore, recent studies have focused on improving federated Intrusion Detection systems through better optimization methods, privacy-preserving approaches, hybrid detector paradigms, and trust management systems, as shown in Fig. 1.

A. Federated Optimization for IoT Intrusion Detection

Federated learning is characterized by efficient optimization as the primary challenge owing to the heterogeneous distribution of client data and limited communication resources. The Privacy-Enhanced IoT Defence System (PEIoT-DS) framework introduces a refined system for federated intrusion detection, specifically for IoT. As IoT devices have limited computational capabilities and operate in non-federated averaging data scenarios, conventional centralized training methods are no longer feasible. The proposed framework consequently uses federated learning to enable distributed learning on IoT devices to train a global intrusion detection model while preserving local data privacy show in Fig. 2.

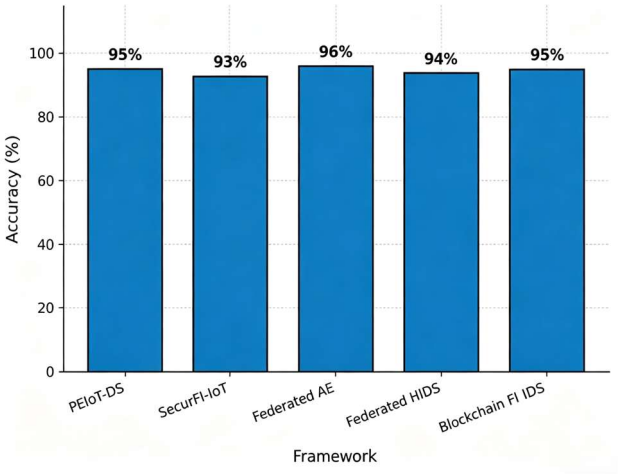

Fig. 2. Comparative performance analysis of federated Intrusion Detection System (IDS) frameworks.

One of the significant contributions of this study is the improvement of the significantly popular algorithm, Federated Averaging. Although federated averaging aggregates local model parameters by averaging, non-uniform distributions of client data may lead to poor performance. To address this problem, the authors present a momentum-based aggregation mechanism that stabilizes global model updating. Adding momentum to the aggregation phase can eliminate sudden parameter changes, and even the gradient updates become more stable between the training rounds.

B. Privacy-preserving Federated Learning in Industrial Internet of Things

The key consideration for industrial IoT systems is privacy protection, as sensitive operational information and company secrets related to their processes and structures should not be disclosed. In this regard, the Secure Federated Learning for IoT framework, a privacy-rooted federated intrusion detection system specifically designed for an industrial setting, was proposed [8]. This structure enables distributed industrial nodes to collectively train intrusion detection models without revealing raw operations.

One of the main innovations of Secure Federated Learning for IoT is the integration of homomorphic comparison and differential privacy into the federated learning process. Homomorphic encryption allows model parameters to be summed while they are encrypted, effectively preventing unauthorized access to the training data. Meanwhile, differential privacy adds noise to model updates to mitigate the risk that adversaries can reconstruct sensitive information from shared gradients, as shown in Table I.

TABLE I. COMPARISON OF FEDERATED IDS FRAMEWORKS

Framework	Dataset	Key Technique	Accuracy (%)
PEIoT-DS	IoT Dataset	Momentum Aggregation	95
SecurFL-IoT	X-IIoTID	Differential Privacy + HE	N/A
Federated AutoEncoder	N-BaIoT	Unsupervised Detection	N/A
Federated HIDS	System Calls	Host-Based Monitoring	94
Blockchain FL IDS	IIoT Dataset	Blockchain Verification	95

The extended Industrial Internet of Things Intrusion Dataset (X-IIoT) dataset was used to evaluate the framework, and it reflects realistic cyber-attack scenarios in an industrial setting. The experiments have shown that the privacy-preserving mechanisms introduce additional computational complexity and a minor drop in detection accuracy; however, the overall performance remains competitive. The framework incurs lower communication overhead and higher privacy, suggesting that it should be applied in large-scale distributed industrial contexts where data privacy and object detection are crucial.

C. Unsupervised Federated Learning Approaches

Another significant research path in federated intrusion detection is unsupervised anomaly detection models, which are particularly applicable for detecting cyber threats that have not been observed previously. A comparative analysis of the performance of supervised and unsupervised models in a federated learning setting is presented using the N-BaIoT dataset, which comprises real network traffic from various IoT devices infected with botnet malware.

Supervised models use labelled datasets to distinguish between malicious and benign traffic patterns. Although such models can be highly accurate in known attack categories, they often become inactive in instances involving zero-day attacks that were not encountered during training. In contrast, untrained models, such as autoencoders, can learn typical behavioral patterns and generate anomalies via reconstruction errors, thereby naturally yielding new attacks.

The results show that the federated Autoencoder model has better generalization performance on heterogeneous Internet of Things devices. Because the autoencoder model is trained on normal device behavior, it can recognize new threats that it has not been trained on. In addition, the federated training system guarantees the privacy of raw device data at the local level, but it is sufficient to enable collaborative anomaly detection across distributed clients.

D. Host-based Federated Intrusion Detection

Most studies on intrusion detection systems focus on analyzing network traffic; however, another level of security is provided by host-based intrusion detection systems, which act as an overlay to network traffic monitoring and observe activities at the device level. The authors suggested a lightweight Hybrid Robust Federated Intrusion Detection System that uses system call traces to detect abnormal host behavior [9]. System calls are interactions between software processes and the operating system kernel, and atypical system call behavior can signal malicious behaviors, such as privilege escalation, malware, or unauthorized access by an attacker. The adopted architecture is a decentralized federated learning architecture, in which each device performs local system call sequence analysis and only shares model updates with a central aggregator. This architecture maintains the confidentiality of host-level data while enabling cooperative model refinement across distributed nodes.

An experimental comparison shows that the lightweight Hybrid Robust Federated Intrusion Detection System achieves high detection accuracy while maintaining a relatively low detection latency, making it a good fit for resource-constrained edge devices. Host-level attack detection complements network-based intrusion detection systems, especially in distributed Internet of Things environments, where internal device responses can reveal threats that network monitoring systems cannot detect.

E. Centralized vs Federated Model Comparison

Several studies have compared the performance of federated learning with conventional centralized machine learning methods for intrusion detection. The University of New South Wales-Network-Based 2015 (UNSW-NB15) dataset was used for a comparative analysis, as it comprises various attack types, including denial-of-service, reconnaissance, exploits, and worms [10], Centralized learning. In this case, all training data are concentrated on a single server, so the model can access the entire data distribution. This generally results in the best performance at the expense of high privacy vulnerability and low scalability. Conversely, federated learning distributes training across multiple clients while preserving local privacy. The findings show that Random Forest achieves the highest classification accuracy in both centralized and federated configurations. The centralized model has a slight performance advantage over the federated model, which is relatively small (typically less than 1%). These results indicate that federated learning can achieve near-centralized detection accuracy and offers significant benefits in terms of privacy, scalability, and regulatory compliance.

F. Hybrid and Blockchain-Enhanced Federated Intrusion Detection System Frameworks

To increase the resilience and reliability of federated intrusion detection systems, recent studies have investigated hybrid architectures and blockchain-based verification schemes. The authors introduced a hybrid federated intrusion detection architecture that combines multiple learning components to enable near-real-time threat detection in distributed IoT networks. The detection accuracy of the system exceeded 96%, indicating that collaborative federated training is useful in latency-constrained settings. On the same note, the presents a blockchain-based federated intrusion detection system optimized to reflect Industrial Internet of Things settings. In federated learning, malicious participants can exploit the system to conduct model poisoning attacks, leading to biased updates during aggregation. Technology is used to overcome this risk, enabling algorithm-proof verification of model changes and guaranteeing the authenticity of customers involved [11]. Transactions for model updates and client contributions are recorded in the blockchain layer, which provides greater trust in the federated ecosystem. The experimental outcomes reveal that the detection capability exceeds 95% and enhances resilience against adversarial manipulation. The massive Internet of Things, which poses severe data privacy challenges, is also an area where federated intrusion detection system solutions have been investigated. proposed a federated-learning-based intrusion detection system in which hospitals and medical devices jointly train intrusion detection models without sharing sensitive patient information. The system's detection rates exceeded 94%, indicating the feasibility of federated intrusion detection in highly controlled healthcare settings, as shown in Table II.

G. Comparative Performance Analysis and Research Gap

In a related comparison of the articles reviewed, it can be noted that federated-learning-based Intrusion Detection System models are generally reported to achieve consistent detection rates of 89.98% across different IoT and IIoT datasets. These findings reveal that even when raw data are managed by edge devices, decentralized collaborative learning remains an efficient method for detecting cyber threats, as shown in Table III.

TABLE II. TAXONOMY AND COMPARATIVE ANALYSIS OF FEDERATED IDS FRAMEWORKS

Framework/Study	Privacy Preservation	Byzantine Robustness	Blockchain Integration	Zero-Day Detection Capability	Non-IID Optimization Support	Main Limitation
FedAvg-based IDS [11]	Moderate	Weak	No	Limited	Weak	Slow convergence under heterogeneous non-IID environments
SecuFL-IoT [6]	Strong	Weak	No	Moderate	Moderate	Accuracy degradation due to differential privacy noise
Federated Autoencoder IDS [7]	Moderate	Weak	No	Strong	Moderate	Limited resistance against adversarial poisoning attacks
Blockchain FL-IDS [2]	Moderate	Moderate	Yes	Moderate	Weak	Additional blockchain verification overhead
Host-based FL-IDS [10]	Moderate	Moderate	No	Moderate	Moderate	Limited visibility of network-wide attack behavior
Proposed HRF-IDS	Strong	Strong	Yes	Strong	Strong	Slight increase in computational and verification overhead

TABLE III. RESEARCH GAP ANALYSIS

Issue	Existing Solutions	Limitations
Privacy Protection	Differential Privacy/Encryption	Accuracy Degradation
Optimization	FedAvg Variants	Non-IID Instability
Trust Management	Blockchain Verification	High Latency
Zero-Day Detection	Autoencoder Models	Lack of Hybrid Detection
Adversarial Security	Byzantine Aggregation	Limited Integration with Privacy

Some of the most effective methods include unmonitored federated models and host-based detection models. Federated Autoencoder models should be more effective at generalizing and detecting zero-day attacks, whereas host-based intrusion detection systems have lower detection latency owing to the direct analysis of device-level behavior. Nonetheless, the risk of a minor loss of accuracy noise injection and encrypted aggregation constraints could be observed in frameworks with strong privacy measures, such as encryption or, more specifically, differential privacy.

Although significant progress has been made in federated intrusion detection research, several crucial issues remain. First, many current solutions focus on privacy preservation, optimization efficiency, or adversarial robustness; however, few incorporate all these issues into a single framework. Second, federated systems are susceptible to malicious client behavior and model poisoning, highlighting the need for robust trust management systems. Finally, most current methods rely solely on supervised or unsupervised detection models, which could limit their ability to identify known and novel attack patterns. Such constraints motivate the development of a hybrid, robust, and privacy-protecting federated intrusion detection system that combines sophisticated aggregation techniques, dynamic privacy procedures, blockchain-secured verification, and hybrid anomaly detection to enhance security and stability in distributed Internet of Things settings.

The Hybrid Robust Federated Intrusion Detection System builds on previous work by offering a coherent integration of several detectors, each tailored for use in realistic federated intrusion detection configurations. Based on Federated Averaging aggregation, an optimization method under non-independent and identically distributed data distributions, the Hybrid Robust Federated Intrusion Detection System is different from PEIoT-DS for mainly introducing an explicit stabilization mechanism in the aggregation that also includes a momentum term, allowing for global convergence under highly heterogeneous IoT data distributions. Unlike Secure Federated Learning for Internet-of-Things (SecuFL-IoT) based privacy-preserving approaches to intrusion detection systems primarily use either differential privacy or homomorphic encryption. The hybrid robust federated Intrusion Detection System uses an adaptive privacy budget that trades detection accuracy for privacy preservation. Moreover, blockchain-based intrusion detection systems offer trust verification; however, most of them have the disadvantage of introducing high latency and are not well integrated with learning dynamics. This problem is overcome using a Hybrid Robust Federated Intrusion Detection System lightweight reputation-based blockchain layer built specifically for federated update validation. Finally, compared to the majority of existing solutions, which are supervised or unsupervised, the Hybrid Robust Federated Intrusion Detection System uses a hybrid architecture that merges deep supervised classification with autoencoder-based anomaly detection, and thus achieves a more robust detection of zero-day attacks.

III. MATERIALS AND METHODS

This section introduces a proposed communication-efficient, adversarial robust and privacy-preserving Hybrid Robust Federated Intrusion Detection System for intrusion detection in various kinds of Internet of Things and Industrial Internet of Things. The architecture fuses traditional features of federated

learning, such as federated optimization, adaptive Differential Privacy, trust management with blockchain, Byzantine-resilient aggregation, and hybrid detection methods.

A. System Overview

Fig. 3 shows that the Hybrid Robust Federated Intrusion Detection System framework operates in a distributed federated environment consisting of a central aggregation server and a set of IoT clients. Each client represents a resource-constrained device that generates local network traffic data.

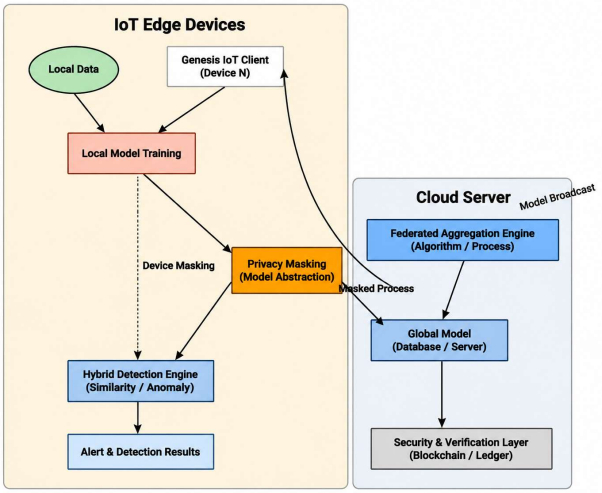

Fig. 3. Proposed architecture of the hybrid robust federated intrusion detection system.

Let the global dataset be:

$$D = \{(x_i, y_i)\}_{i=1}^{N} \tag{1}$$

where $x_i \in \mathbb{R}^d$ denotes network traffic features and $y_i \in \{0,1\}$ represents normal or attack labels.

The dataset is partitioned across K clients under a non-IID distribution, as expressed in Eq. (2):

$$D = \bigcup_{k=1}^{K} D_k \, , D_i \neq D_j \text{ for } i \neq j \tag{2}$$

This setting reflects realistic Internet of Things environments where devices generate heterogeneous traffic patterns.

B. Local Model Training Objective

Each client minimizes the following empirical loss (Eq. (3)) on its local data:

$$\mathcal{L}_k(w) = -\frac{1}{|D_k|} \sum_{(x,y)\in D_k} \begin{bmatrix} y \log p_w(x) + (1-y) \\ \log(1 - p_w(x)) \end{bmatrix} \tag{3}$$

where:

- w is model parameter vector
- $p_w(x)$ is predicted probability
- D_k is local dataset of client k

The local model parameters are updated using stochastic gradient descent, as given in Eq. (4):

$$w_k^{t+1} = w_k^t - \eta \nabla \mathcal{L}_k(w_k^t) \tag{4}$$

where η is learning rate.

C. Adaptive Differential Privacy Mechanism

To prevent gradient leakage attacks (e.g., inversion and membership inference), each client applies adaptive Gaussian differential privacy before sending updates.

To protect against gradient leakage, each client adds Gaussian noise to its gradient, yielding the privatized update shown in Eq. (5):

$$\tilde{g}_k^t = g_k^t + \mathcal{N}(0, \sigma^2) \tag{5}$$

where noise scale is defined as:

$$\sigma = \frac{\Delta f}{\epsilon}$$

- ϵ: Privacy Budget.
- Δf: sensitivity of gradients

Lower ϵ ensures stronger privacy but increases noise, introducing a controlled privacy-utility trade-off.

D. Blockchain-Assisted Trust and Verification Layer

To mitigate Byzantine and poisoning attacks, a lightweight permissioned blockchain is integrated.

Each client submits a hashed version of its update for verification, as defined in Eq. (6):

$$H_k^t = \text{Hash}(w_k^t) \tag{6}$$

Smart contracts validate:

- client identity.
- update integrity.
- historical contribution quality.

A dynamic reputation score is updated as:

$$R_k^{t+1} = \alpha R_k^t + (1-\alpha) Q_k^t$$

where:

- R_k^t: reputation score.
- Q_k^t: quality score of updates.
- $\alpha \in (0,1)$: decay factor.

Clients with: $R_k^t < \tau_r$ are excluded from aggregation.

E. Byzantine-resilient Trimmed Mean Aggregation

To ensure robustness against adversarial updates, Hybrid Receptive Field Intrusion-Detection System (HRF-IDS) applies trimmed mean aggregation.

The sorted client updates are denoted as follows (Eq. (7)):

$$\tilde{g}_{(1)}^t \leq \tilde{g}_{(2)}^t \leq \ldots \leq \tilde{g}_{(K)}^t \tag{7}$$

The aggregated gradient is:

$$g^t = \frac{1}{K-2m}\sum_{i=m+1}^{K-m} \tilde{g}_{(i)}^t$$

where:

- m: Number of suspected Byzantine clients.

Extreme values are eliminated prior to aggregation. This mechanism guarantees system robustness in the presence of malicious Byzantine clients.

F. Momentum-Based Global Aggregation (Convergence Enhancement)

To stabilize convergence, we introduce a momentum term in the global aggregation, as shown in Eq. (8):

$$v^{t+1} = \beta v^t + (1-\beta) g^t \tag{8}$$

Global update:

$$w^{t+1} = w^t - \eta v^{t+1}$$

where:

- β: momentum coefficient
- v^t: accumulated global direction

This stabilizes training under heterogeneous IoT distributions and reduces convergence rounds

G. Hybrid Intrusion Detection Model

The Hybrid Robust Federated Intrusion Detection System integrates a validation-based percentile strategy to identify the threshold τ for anomaly detection. The sorting of reconstruction errors derived from the normalization samples is performed, and the threshold τ is set to the 95th percentile value. By applying this method, normal traffic, even traffic fluctuations, is maintained, and all other abnormal traffic can be discovered, even by unknown and zero-day attacks. Training is a time-consuming process that updates the threshold reasonably often to reflect changes in the IoT traffic distribution.

1) Supervised classifier (softmax output)

The supervised classifier outputs a probability distribution over classes, given by Eq. (9):

$$P\,(\,y \mid x\,) = \text{Softmax}\left(f_w(x)\right) \tag{9}$$

2) Unsupervised autoencoder for zero-day detection

For zero-day detection, the autoencoder reconstructs the input as described in Eq. (10):

$$\hat{x} = Decoder\big(Encoder(x)\big) \tag{10}$$

Reconstruction error:

$$E(x) = \| x - \hat{x} \|^2$$

Anomaly decision rule:

$$E(x) > \tau$$

Threshold definition:

$$\tau = \text{Percentile}_{95}\,(E(x_{validation}))$$

H. Communication Efficiency Model

The total communication cost is estimated using Eq. (11):

$$C_{total} = R \times S_{update} \tag{11}$$

where:

- R: communication rounds.
- S_{update}: model update size.

Momentum-based convergence reduces R, improving efficiency.

I. Complexity and Robustness Analysis

The proposed Hybrid Robust Federated Intrusion Detection System framework was mainly analyzed using experimental simulations and performance analysis. The present study does not provide a full proof of convergence and adversarial robustness guarantees of the framework; however, it uses concepts for momentum-based optimization, adaptive differential privacy, blockchain-assisted verification, and Byzantine-resilient aggregation. Instead, the purpose of the theoretical discussion is to facilitate an intuitive understanding of the behavior of the framework in the context of a heterogeneous federated Internet of Things setting.

The complexity of the framework can be summarized as follows.

Local client-side model training complexity: $O\,(E \times B \times d)$. Where:

- E denotes local training epochs.
- B represents mini-batch operations.
- d is the number of trainable parameters.

Aggregation complexity at the server: $O(Kd)$. Where K represents the number of participating clients.

Trimmed mean sorting complexity: $O(K\log K)$.

Blockchain verification overhead stems from transaction validation and smart-contract execution.

The Byzantine-robustness mechanism adopted in the Hybrid Robust Federated Intrusion Detection System resembles widely-adopted schemes in trimmed-mean aggregation literature. These schemes assume that the fraction of malicious clients in each aggregation round does not exceed a tolerable threshold. The resiliency of Byzantine systems is measured experimentally by simulating participation ratios of adversarial Byzantine agents up to 20%. As such, robustness claims assume the future existence of adversarial participation rather than assuming their validity with a formal guarantee. Similarly, for non-intrusion detection system data distributions, the convergence stability of the momentum-based aggregation mechanism was observed in the experiments. However, the convergence proof of the mechanism under a federated optimization problem with heterogeneity was not considered in this study and is left as an important direction for future research. Moreover, the adaptive

differential privacy mechanism applies Gaussian noise injection to reduce the risk of gradient leakage. Our formulation relies on the privacy budget parameter ϵ; however, there is no formal Rényi Differential Privacy accountant, and no cumulative privacy-loss tracking over multiple communication rounds is implemented. This means that the analysis of privacy should be understood as an approximation of privacy rather than a formal privacy claim.

It should be noted that the proposed framework was mainly tested using simulation-based experiments in this study, without a formal theoretical analysis. This means that claims of convergence behavior, Byzantine resistance, and privacy protection must be understood as observed over the tested experimental conditions and are not mathematically provable properties. An important direction for future research is formal theoretical validation, including convergence proofs, certified adversarial robustness analysis, and robust privacy accounting, such as that provided by Rényi Differential Privacy.

IV. Result and Discussion

To ensure reproducibility and practicality, the proposed Hybrid Robust Federated Intrusion Detection System was simulated in a federated learning environment using a multi-client emulation framework. The simulation process was created to realistically model distributed Internet of Things and Internet of Industrial Things environments, where each client runs independently with local data partitions stored within the client. The implementation was performed in Python using the commonly used deep learning frameworks TensorFlow and PyTorch. Experiments were run on a common daily use laptop with an Intel Core i5 or equivalent processor and 8–16 GB of Random-Access Memory (RAM), with built-in graphics in a typical computing environment for many researchers and practitioners in resource-limited environments. Although the system had some hardware limitations, it was made as efficient as possible by employing a range of different lightweight models with the help of batch processing methods for training and evaluation. These federated clients were emulated as autonomous standalone training processes and locally distributed non-Independent and Identically Distributed (non-IID) data partitions extracted from the X-IIoTID and N-BaIoT datasets. To mimic communication with a central server, the only thing exchanged between clients and the central server was the model parameters and not the raw data, following the standard communication procedures in federated learning. The above-mentioned communication overhead is reported as model update size estimates sent for each communication round, not a measurement in a real communication environment. This type of estimation, which is based on simulation, is frequently utilized in federated learning studies to compare the communication efficiency between various approaches while remaining fair. The proposed architectural structure provides the feasibility of general personal computing tools and allows for the evaluation of the convergence behavior of various models for privacy preservation analysis under the conditions of heterogeneous and adversarial IoT, without losing realism.

A. Classification Performance on X-IIoTID Dataset

A comparison of the various models based on their classification results is shown in Fig. 4 using the X-IIoTID dataset. The highest accuracy provided by the learning method was 97.82% for centralized learning. The proposed Hybrid Robust Federated Intrusion Detection System achieves an accuracy of up to 97.15 %, which is only 0.67% lower than that of centralized learning, demonstrating that the federated structure can achieve a performance similar to that of centralized learning without violating data privacy.

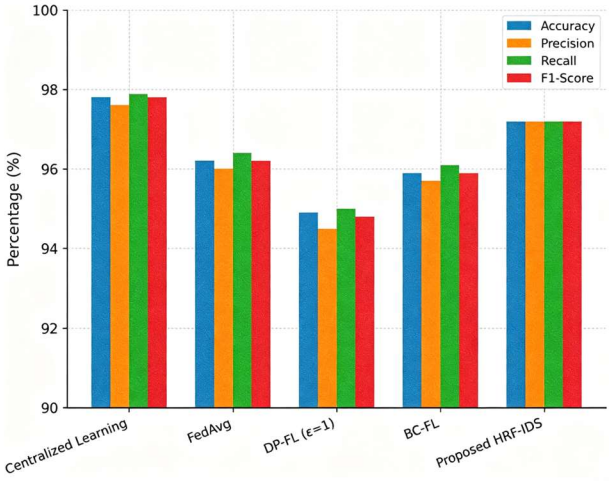

Fig. 4. Comparative performance analysis on the X-IIoTID dataset.

The integrated model, Hybrid Robust Federated Intrusion Detection System, shows better performance than the conventional Federated Averaging 96.21% accuracy by 0.94%, with the difference largely owing to momentum-based aggregation, which stabilizes training under non-independent and identically distributed data distributions. The accuracy of the Differential Privacy Federated Learning model $e = 1$ is lower than 94.87% because of noise injection during optimization. However, the adaptive privacy mechanism of the Hybrid Robust Federated Intrusion Detection System minimizes this performance loss. Overall, the findings affirm that the Hybrid Robust Federated Intrusion Detection System is useful for balancing accuracy, privacy, and robustness in Industrial Internet of Things systems.

B. Performance on N-BaIoT Dataset (Zero-Day Evaluation)

The N-BaIoT dataset was used to assess the zero-day detection performance of the proposed hybrid Robust Federated Intrusion Detection System architecture, which combines supervised learning with autoencoder-based anomaly detection. The hybrid model outperformed the benchmark models, achieving a recall of 97.88% and an F1-score of 97.09%. Notably, the high recall demonstrates that the model is highly effective at detecting previously

unseen attacks, as exemplified by its zero-day detection capability (see Fig. 5).

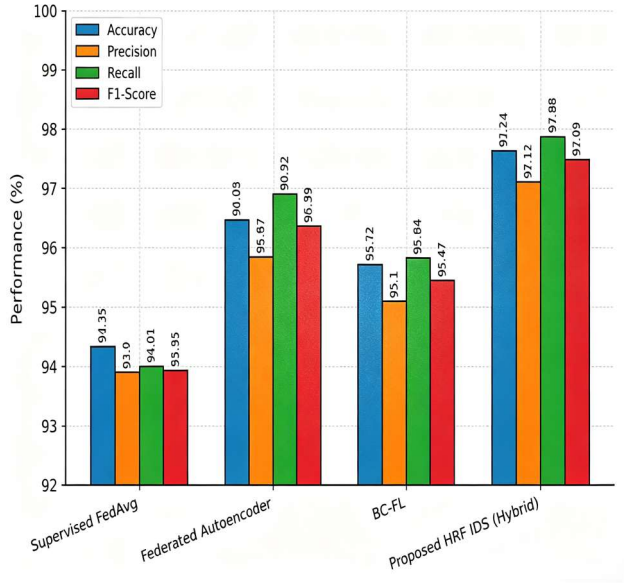

Fig. 5. Shows zero-day detection performance (N-BaloT dataset).

The autoencoder in the system helps it generalize better to anomalies, enabling it to identify patterns not present in the training data, and the supervised classifier ensures maximum accuracy in predicting normal and malicious behavior. The hybrid approach showed a significant increase in the F1-score of approximately 1.1 over the standalone federated autoencoder, attesting to the benefits of combining supervised and unsupervised detection mechanisms into a single entity to achieve better performance.

C. Convergence Analysis

In the proposed Hybrid Robust Federated Intrusion Detection System model, momentum-based aggregation is used as a strategic mechanism to achieve high convergence rates in federated learning systems. This system trains models across distributed clients with heterogeneous, non-independent, and identically distributed data, often leading to oscillations and slower convergence with conventional aggregation methods such as Federated Averaging.

To overcome this, a momentum term was added to the aggregation process, allowing the global model to capture the previous gradient information and smooth out sudden fluctuations when changing the parameters. The results in Fig. 6 provide a comparative analysis of various models in terms of the number of communication rounds required to achieve 95%. The standard Federated Averaging model took 62 rounds, reflecting the delay caused by non-independent and identically distributed data, as expected. Differential Privacy Federated Learning, which incorporates the use of differential privacy, required even more rounds, 71 rounds in total, because of the added noise used to secure the privacy of the data. Using blockchain-enhanced aggregation, BC slightly increased its efficiency by 58 rounds, though it was still less efficient than the suggested solution. In contrast, the proposed Hybrid Robust Federated Intrusion Detection System model achieved the same accuracy in 41 rounds of communication, a significant decrease of approximately 34% compared to the standard Federated Averaging. Such an increase in momentum improvement indicates the ability of momentum-based aggregation to smoothen the training process, reduce the oscillatory nature of heterogeneous client data, and accelerate the convergence of the global model. Therefore, the addition of the momentum term reduces the training time and communication overhead and enhances the overall training efficiency, making the Hybrid Robust Federated Intrusion Detection System an extremely viable solution for a real-world federated intrusion detection system.

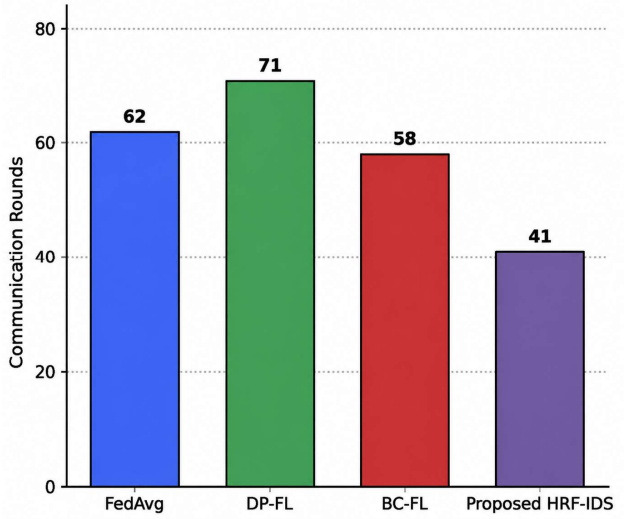

Fig. 6. Convergence speed comparison.

D. Robustness Against Byzantine Attacks

To test the resilience of the proposed Hybrid Robust Federated Intrusion Detection System to poisoning attacks, an experimental case was modelled in which every 5th client was malicious and injected tampered gradients into the federated learning process. The results of various models against Byzantine attacks are shown in Fig. 7.

The baseline models suffer substantial accuracy degradation under 20% Byzantine attacks. FedAvg-IDS achieves 92.5% accuracy under benign conditions, which falls sharply to 45.2% under adversarial interference. BC-FedIDS exhibits the most severe performance degradation, with accuracy decreasing from 95.0% to 38.5%. By contrast, the proposed Hybrid Robust Federated Intrusion Detection System (HRF-IDS) yields the best resilience among all competitors: its accuracy declines from 89.7% (benign scenario) to 71.3% under Byzantine attack, corresponding to an accuracy drop of 18.4%. Such improved resilience is attributed to the joint effect of trimmed-mean aggregation and blockchain-based verification, which mitigate gradient manipulation and limit the negative impact of adversarial clients on the global model. These results demonstrate the enhanced adversarial robustness of the proposed system for federated intrusion-detection tasks against malicious participants (see Fig. 7).

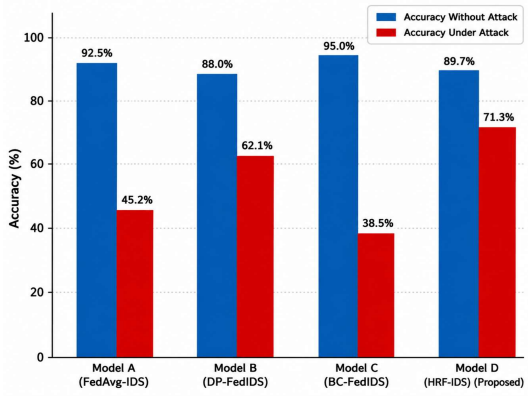

Fig. 7. Shown Model performance comparison under 20% byzantine attack.

However, the current evaluation was conducted in a simulated adversarial setting where 20% of the agents were Byzantine, and a wider sensitivity analysis using different adversarial ratios and different adaptive strategies/policies for attacking is planned for the future.

E. Communication Overhead Analysis

The efficiency of communication is also a key factor in IoT systems, as bandwidth and resource-constrained conditions necessitate the minimum possible data exchange. To measure this point, Fig. 8 compares the total communication costs across various federated learning models. The Federated Averaging model, with an average update size of 4.8 MB and 62 communication rounds, had an overall communication cost of 297.6 MB. The version of DP-Federated learning with 2-level privacy mechanisms slightly increased the average update size to 5.2 MB and the number of rounds to 71, with a final cost of 369.2 MB. Using blockchain-based aggregation, Federated Learning had an average update size of 5.6 MB across 58 rounds, totaling 324.8 MB. Conversely, the proposed Hybrid Robust Federated Intrusion Detection System, albeit with a slightly increased update size of 5.4 MB achieved through encryption and security-aggregating techniques, still required 41 communication rounds, and the achievable total communication cost of 221.4 MB was significantly lower.

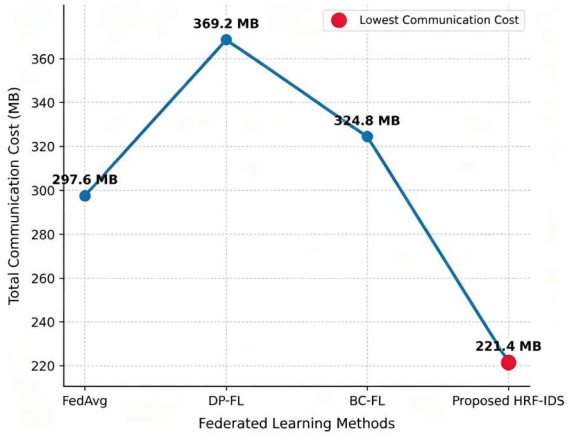

Fig. 8. Communication cost comparison.

This is a general decline of nearly one-fourth of the Federated Averaging. The findings suggest that security provisions may slightly increase individual updates. Nevertheless, the overall effect of momentum-based aggregation and shorter communication rounds is a significant improvement in communication efficiency. These findings suggest that the proposed framework may be applicable to resource-constrained IoT environments; however, validation on real hardware platforms remains necessary.

F. Privacy-performance Trade-off Analysis

Fig. 9 shows the performance of the proposed Hybrid Robust Federated Intrusion Detection System with different privacy budgets (e). The privacy budget e regulates the extent of differential privacy used when updating the models, trading off data protection and detection accuracy.

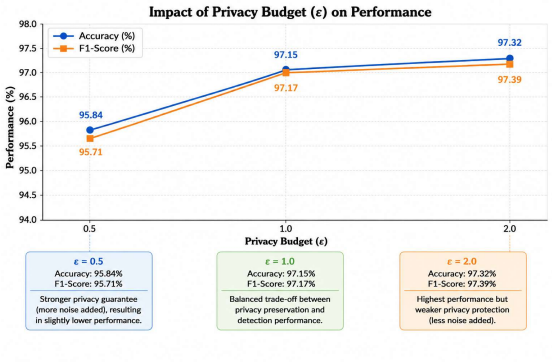

Fig. 9. Impact of privacy budget on performance.

The model achieved an accuracy of 95.84% and an F1-score of 95.71% when e was set to 0.5. A smaller privacy budget imposes a tighter privacy guarantee and introduces larger gradient noise, which slightly degrades model performance. Further increasing e to 1 improves accuracy to 97.15% and the F1-score to 97.17%, yielding a favorable balance between privacy preservation and detection capability. Increasing e to 2 further elevates accuracy to 97.32% and the F1-score to 97.39%, yet this comes at the expense of reduced privacy protection. These results suggest that the privacy budget $e = 1$ provides the best trade-off: it maintains satisfactory intrusion-detection performance while preserving strong privacy for client-side data within the federated-learning framework.

G. Ablation Study

In non-independent and identically distributed federated environments, to thoroughly understand the contribution of each component in the proposed Hybrid Robust Federated Intrusion Detection System, an ablation study was performed on the X-IIoTID and N-BaIoT datasets. This analysis aimed to evaluate the effects of momentum-based aggregation, adaptive differential privacy, blockchain-assisted verification, trimmed mean Byzantine defense, and hybrid supervised-unsupervised detection architecture on the overall system performance. The performance of the proposed framework is compared in Table IV.

TABLE IV. ABLATION ANALYSIS OF HRF-IDS COMPONENTS

Model Configuration	Privacy Protection	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Communication Rounds	Byzantine Robustness	Zero-Day Detection Capability
Standard FedAvg	No	96.21	95.88	96.19	96.04	62	Low	Moderate
FedAvg + Momentum Aggregation	No	96.88	96.54	96.90	96.71	47	Moderate	Moderate
FedAvg + Differential Privacy	Yes	95.74	95.41	95.80	95.60	71	Moderate	Moderate
FedAvg + Blockchain Verification	Partial	96.42	96.18	96.45	96.31	58	High	Moderate
FedAvg + Trimmed Mean Defense	No	96.67	96.33	96.75	96.54	55	High	Moderate
FedAvg + Hybrid Detector	No	97.01	96.82	97.10	96.95	49	High	High
Proposed HRF-IDS	Yes	97.15	97.12	97.22	97.17	41	Very High	Very High

The experimental results show the distinctiveness of the contribution of each module to the effectiveness of the proposed framework. Based on the non-independent and identically distributed client distribution, the number of rounds was reduced from 62 to 47 when aggregation methods based on Momenta were used, which shows significant convergence efficiency. This shows the effectiveness of the momentum mechanism for stabilizing federated optimization, which reduces the oscillatory behavior when aggregating information from the rest of the system. The adaptive differential privacy effect scales the classification accuracy, as Gaussian noise is added to make local model updates harder to determine in gradient leakage attacks; however, its effect on privacy is improved. Overall, while this trade-off exists, the degradation in performance is not significantly large, showing that the proposed adaptive privacy mechanism is an optimal compromise between the detection capability and privacy. Blockchain-assisted verification is used to improve the trust management process and adversarial resilience by validating client updates before aggregation. The experimental results suggest that blockchain verification has the potential to increase the resistance of participants to attacks and enhance the robustness of the blockchain against model poisoning attacks. Likewise, the aggregation mechanism that removes the top and bottom 25% of the gradients to obtain the aggregated average gradient is efficient in resisting the Byzantine clients and stabilizing the aggregated value under adversarial environment.

The most significant improvement was observed in the hybrid supervised-unsupervised detection architecture for attacks that occurred on an unknown date. The artificial neural network framework based on the supervised classification technique and the autoencoder-based anomaly detection technique were able to increase the detection rate of previously unseen attack patterns while preserving the accuracy of detection of known attack patterns. The overall results of the entire Hybrid Robust Federated Intrusion Detection System framework were the highest for the overall accuracy with 97.15% and the lowest number of communication rounds 41, while retaining high adversarial robustness. These results validate that all of the suggested components are needed to simultaneously guarantee the privacy of protection, stability of convergence, efficient communication, and a power intrusion detection system in distributed Internet of Things and Industrial Internet of Things federated systems.

H. Statistical Significance Analysis

All experiments were repeated five times throughout this study, with the same federated configuration, such as client distribution, participation ratio of adversarial clients, and hyperparameters, to guarantee reliability, repeatability, and statistical consistency in the outcomes. The reported performance parameters correspond to the average value of all experimental runs, and the standard deviation values were calculated to check the convergence consistency and experimental stability.

The proposed Hybrid Robust Federated Intrusion Detection System achieved promising performance, with average classification accuracy values listed as follows:

X-IIoTID dataset: $97.15 \pm 0.31\%$

N-BaIoT dataset: $97.64 \pm 0.28\%$

Likewise, the F1-score remained consistent through repeated runs, verifying the detection consistency within the non-independent and Identically Distributed federated heterogeneous environment with Byzantine participants.

To further enhance the statistical rigor, 95% Confidence Intervals were calculated for key assessment metrics based on the experimental data observed from repeated runs. The approximate 95% confidence intervals for the classification accuracy of the hybrid robust federated intrusion detection system WERE:

X-IIoTID dataset: $97.15\% \pm 0.27\%$ (95% CI)

N-BaIoT dataset: 97.64% ± 0.24% (95% CI)

Furthermore, we performed a comparative evaluation of repeated experimental observations between the Hybrid Robust Federated Intrusion Detection System and some baseline approaches, such as federated averaging, differential privacy federated learning, and BC-Federated learning, to measure the consistency of the trends of techniques across different runs. The results consistently demonstrated empirical performance improvements in convergence efficiency, adversarial robustness, and classification accuracy across repeated experimental runs. However, the number of repeated experimental executions was restricted, and the statistical analysis is a supportive empirical validation and cannot be interpreted as a formal statistical validation. There are new aspects of the statistical analysis that are explicitly included in the reported statistical analysis, such as the values of standard deviations and confidence interval estimates.

This assessment will be expanded in future research with larger repeated run samples, formal statistical testing methods, and graphical control of variability in the form of error bars over comparative performance measures.

Although the experimental results show that it exhibits stable convergence under simulated Byzantine settings and has good empirical robustness, it does not prove any formal theoretical convergence bound or any certified adversarial guarantee. Robustness observations were thus conducted in limited federated simulation environments with empirical experiences. Future work will concentrate on considering formal convergence analysis, guaranteed Byzantine robustness, and cutting-edge strategies for privacy accounting, including Rényi Differential Privacy analysis, for better theory validation.

The present paper is mainly statistical in character, with the empirical test being conducted by means of repeated simulations rather than by any formal testing procedures such as paired *t*-tests or any non-parametric statistical tests. Statements of average improvement, as reported, should thus be taken as a general observation of repeated experiments, rather than as the formal level of statistical significance. In the future, further experiments can be conducted on a larger scale, and formal statistical testing with *p*-value analysis can be performed.

I. ROC-AUC Performance Evaluation

To check the detailed performance of the proposed Hybrid Robust Federated Intrusion Detection System, Receiver Operating Characteristic analysis and Area Under the Curve metrics were performed on both the datasets, X-IIoTID and N-BaIoT. Several researchers have established that Receiver Operating Characteristic (ROC) analysis is one of the best metric methods that can be used to assess the performance of machine learning models when distinguishing between malicious and benign traffic at varying decision thresholds. Unlike accuracy, the intensity, sensitivity, and specificity of a model, as well as the overall discriminating ability, can be understood by Receiver Operating Characteristic, Area Under the Curve analysis. The experimental results show that the proposed Hybrid Robust Federated Intrusion Detection System obtained an Area Under the Curve of 0.986 and 0.991 on the X-IIoTID and N-BaIoT datasets, respectively. The very high Area Under the Curve values support the good ability of the framework to correctly classify normal traffic from attack traffic in various IoT scenarios. The findings also show that incorporating momentum-based federated aggregation, adaptive differential privacy, blockchain-based verification, and hybrid anomaly detection substantially affects the stability and reliability of the classification performance. Furthermore, the Receiver Operating Characteristic graph was used to assess the proposed framework, which generated a consistently low false-positive rate, indicating that the proposed framework provides high attack detection sensitivity while reducing false alarms. This quality is very relevant in real IoT scenarios, as many false alarms can lead to excessively high operational costs and degradation of the system's reliability. Additionally, the hybrid supervised-unsupervised detection architecture achieved high performance for unknown/zero-day attacks, indicating the robustness and generalization ability of the designed HRF-IDS framework in a heterogeneous federated learning environment.

J. Experimental Limitations and Future Deployment Considerations

The proposed Hybrid Robust Federated Intrusion Detection System framework was shown to perform well in simulated federated IoT environments; however, some practical challenges need to be addressed. Compared to real deployment on edge/embedded devices, which is quite difficult because they are typically remote, require manual configuration, and typically cannot handle complex software, this initial study used simulated devices based on the software and offers a framework for federated simulation. Therefore, performance measurements (e.g., inference latency, occupancy of memory resources, CPU resource usage, runtime overhead, battery consumption, and energy efficiency specific to the hardware) were not measured in this study.

This study presents an analysis of the communication overhead that assumes the volume of communications transmitted through the model assumes each communication round in a federation, but it does not measure the exact latency of the communications through the physical network. Likewise, the overhead in blockchain verification was estimated theoretically by increasing the complexity of the updates that need to be processed rather than by real implementation on a distributed blockchain infrastructure.

Moreover, the main focus of the adversarial robustness evaluation was on Byzantine participation ratios of up to 20% of malicious clients with simulated participation. The proposed aggregation mechanism, based on trimmed means, and the proposed verification mechanisms based on blockchain, empirically demonstrated robust performance under these conditions. Further investigation along the lines of varying attack intensities, attack adaptive strategies, Byzantine clients, and larger proportions of malicious clients should be considered in future research.

Moreover, the current implementation was performed under the assumption of a controlled simulation environment that runs on a workstation. However, other realistic constraints, such as intermittent network connectivity, customer dropouts, asynchronous communication, varying hardware capabilities, and limited resources, can also influence federated convergence and detection.

Therefore, future work will involve investigating the deployment and validation on real edge-IoT hardware platforms, Physical System dimensioning and timings, power and energy consumption, Distributed Federated deployment Evaluation, and Negative attack simulation with a variety of attack scenarios, and replacement as the primary feature to consider for the IoT in constrained situations, which is easy to achieve and orchestrated in real time with blockchain technology.

In future work, we will evaluate the framework on real edge-Internet of Things hardware platforms to verify its deployment ability in practical operating environments.

The present study mainly assessed the framework utilizing software-based simulated federated clients and hypothetical communication behaviors. Consequently, performance metrics in terms of network latency, runtime overhead, memory utilization, CPU utilization, battery consumption rate, energy efficiency, and blockchain infrastructure overhead were not experimentally measured. Therefore, these results should be regarded as simulation evidence of potential applicability rather than as confirmation of deployment readiness.

V. Conclusion

This study introduced a Hybrid Robust Federated Intrusion Detection System that provides privacy protection and scalability in distributed and Industrial Internet of Things environments. This study addressed the most significant drawbacks of centralized intrusion detection systems, such as communication overhead, privacy exposure, high detection latency, and single point of failure. The proposed integrated framework demonstrates the feasibility of integrating existing federated learning techniques into a single architecture for secure and decentralized Intrusion Detection Systems. The results show that the proposed integrated framework can potentially use federated methods for intrusion detection systems. The high accuracy of the detection and significantly lower number of communication costs were experimentally verified when implemented on the benchmark X-IIoTID and N-BaIoT datasets, respectively. The combination of supervised learning and the use of an autoencoder with anomaly detection increased the accuracy of known and zero-day attacks while boosting the adaptability of the system to new cyberthreats. Moreover, the integration of strong aggregation capabilities and blockchain verification mechanisms bolstered the system's security and resistance to adversarial influences and model poisoning attacks, thereby enhancing overall trust and resilience in collaborative learning environments. The results of this study show that federated learning can be an effective support to build the next generation of intrusion detection systems, specifically for safe smart environments, such as smart healthcare, industrial automation, smart cities, and other Internet of Things critical infrastructures, where the preservation of privacy and the ability to detect threats in real time are crucial concerns. Despite their success, there are some drawbacks, such as high computational costs and the need to manage resources efficiently on highly constrained edge devices. Future research could explore lightweight federated optimization schemes, adaptive client selection strategies, energy-efficient security mechanisms, and the incorporation of Explainable Artificial Intelligence models to further enhance transparency and decision interpretability. The proposed framework is poised to play a significant role in shaping the future of IoT cybersecurity in modern distributed networks, including secure, scalable, and intelligent approaches.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Muhammad Shaharyar Ramzan conceived the study, designed the methodology, conducted the experiments, analyzed the results, and prepared the original manuscript draft; Arshad Ali supervised the research, contributed to the study design, reviewed and edited the manuscript, and provided technical guidance throughout the study; Rafi us Shan contributed to data analysis, validation, manuscript review, and critical revisions. All authors read and approved the final version of the manuscript.

References

[1] V. Holubenko, D. Gaspar, R. Leal, and P. Silva, "Autonomous intrusion detection for IoT: A decentralized and privacy preserving approach," *Int. J. Inf. Secur*., vol. 24, no. 1, 7, 2025.

[2] A. Ali, M. Husain, and P. Hans, "Federated learning-enhanced blockchain framework for privacy-preserving intrusion detection in industrial IoT," arXiv preprint, arXiv:2505.15376, 2025.

[3] Y. Bai, W. Jiang, J. Mu, S. Liu, W. Gu, and S. Wang, "Enhancing IoT Security via federated learning: A comprehensive approach to intrusion detection," *IET Inf. Secur*., vol. 2025, no. 1, 8432654, 2025.

[4] A. Khraisat, A. Alazab, M. Alazab, A. Obeidat, S. Singh, and T. Jan, "Federated learning for intrusion detection in IoT environments: A privacy-preserving strategy," *Discover Internet of Things*, vol. 5, no. 1, 72, Jun. 2025.

[5] M. S. Ramzan *et al*., "An innovative machine learning based end-to-end data security framework in emerging cloud computing databases and integrated paradigms: Analysis on taxonomy, challenges, and opportunities," *Spectrum of engineering Sciences*, vol. 3, no. 2, pp. 90–125. 2025.

[6] A. Alqazzaz, "SecuFL-IoT: An adaptive privacy-preserving federated learning framework for anomaly detection in smart industrial networks," *Sci. Rep*., vol. 16, no. 1, 4107, 2026.

[7] M. Sajid *et al*., "Enhancing intrusion detection: A hybrid machine and deep learning approach," *Journal of Cloud Computing*, vol. 13, no. 1, 123, 2024.

[8] A. Vyas, P. C. Lin, R.-H. Hwang, and M. Tripathi, "Privacy-preserving federated learning for intrusion detection in IoT environments: A survey," *IEEE Access*, vol. 12, pp. 127018–127050, 2024.

[9] F. Mosaiyebzadeh *et al*., "Privacy-preserving federated learning-based intrusion detection system for IoHT devices," *Electronics*, vol. 14, no. 1, 67, Dec. 2024.

[10] G. Rampone, T. Ivaniv, and S. Rampone, "A hybrid federated learning framework for privacy-preserving near-real-time Intrusion detection in IoT environments," *Electronics*, vol. 14, no. 7, 1430, 2025.
[11] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Machine Learning Research*, pp. 1273–1282. 2017.