Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.068673

ARTICLE

Towards Decentralized IoT Security: Optimized Detection of Zero-Day Multi-Class Cyber-Attacks Using Deep Federated Learning

Misbah Anwer[1,*], **Ghufran Ahmed**[1], **Maha Abdelhaq**[2], **Raed Alsaqour**[3], **Shahid Hussain**[4] **and Adnan Akhunzada**[5,*]

[1]Department of Computer Science, School of Computing, National University of Computer and Emerging Sciences (FAST-NUCES), Karachi, 75030, Pakistan

[2]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, 11671, Saudi Arabia

[3]Department of Information Technology, College of Computing and Informatics, Saudi Electronic University, Riyadh, 93499, Saudi Arabia

[4]Department of Computer Science, Penn State University, Behrend, PA 16563, USA

[5]Department of Data & Cybersecurity, College of Computing & IT, University of Doha for Science and Technology, Doha, 24449, Qatar

*Corresponding Authors: Misbah Anwer. Email: k200992@nu.edu.pk; Adnan Akhunzada. Email: adnan.adnan@udst.edu.qa

Received: 03 June 2025; Accepted: 05 August 2025; Published: 10 November 2025

ABSTRACT: The exponential growth of the Internet of Things (IoT) has introduced significant security challenges, with zero-day attacks emerging as one of the most critical and challenging threats. Traditional Machine Learning (ML) and Deep Learning (DL) techniques have demonstrated promising early detection capabilities. However, their effectiveness is limited when handling the vast volumes of IoT-generated data due to scalability constraints, high computational costs, and the costly time-intensive process of data labeling. To address these challenges, this study proposes a Federated Learning (FL) framework that leverages collaborative and hybrid supervised learning to enhance cyber threat detection in IoT networks. By employing Deep Neural Networks (DNNs) and decentralized model training, the approach reduces computational complexity while improving detection accuracy. The proposed model demonstrates robust performance, achieving accuracies of 94.34%, 99.95%, and 87.94% on the publicly available kitsune, Bot-IoT, and UNSW-NB15 datasets, respectively. Furthermore, its ability to detect zero-day attacks is validated through evaluations on two additional benchmark datasets, TON-IoT and IoT-23, using a Deep Federated Learning (DFL) framework, underscoring the generalization and effectiveness of the model in heterogeneous and decentralized IoT environments. Experimental results demonstrate superior performance over existing methods, establishing the proposed framework as an efficient and scalable solution for IoT security.

KEYWORDS: Cyber-attack; intrusion detection system (IDS); deep federated learning (DFL); zero-day attack; distributed denial of services (DDoS); multi-class; Internet of Things (IoT)

1 Introduction

Security has become a cornerstone of modern computing, underpinning the protection of sensitive information across industries. From corporate environments and financial institutions to healthcare systems and government agencies, the need for robust cybersecurity frameworks has never been more critical. As digital connectivity intensifies, particularly through the integration of Internet of Things (IoT) devices, the attack surface has significantly expanded. With the growing dependence on networked systems, ensuring the confidentiality, integrity, and availability of data is of paramount importance. Security is the key concern in computer science and in today's life.

It is a significant concern for any company, organization, business, etc. Security is adopted in all major areas to prevent destruction and loss [1]. Implementation and prevention of network or any software application from a zero-day attack is one of the challenges of security [2–5]. According to cybersecurity data, there are 2200 cyberattacks daily, or every 39 s on average. An average data breach in the US costs $9.44 million; by 2023, cybercrime is expected to cost $8 trillion [6]. Some statistics, facts, and figures show how top companies are affected and how attackers damage systems and exploit information [7–9].

A novel evaluation metric, Zero-day Detection Rate (Z-DR), measures the learning model's ability to identify zero-day invaders. The proposed zero-day short learning (ZSL) structure is divided into two primary components [10]. The models extract and map network data features to the distinguishing characteristics of popular attacks on visible classes during the attribute learning phase. During the inference phase, the model establishes linkages between known and unknown (zero-day) attacks to help identify and classify them as malicious [11]. Technically, FL is a distributed collaborative AI technique that uses a central server to manage multiple devices without using real datasets to train its models. Data privacy intensification, low-latency network communication, and improved learning quality are the significant features of FL for IoT applications [12]. Fig. 1 depicts the core architecture of federated learning with local and global models.

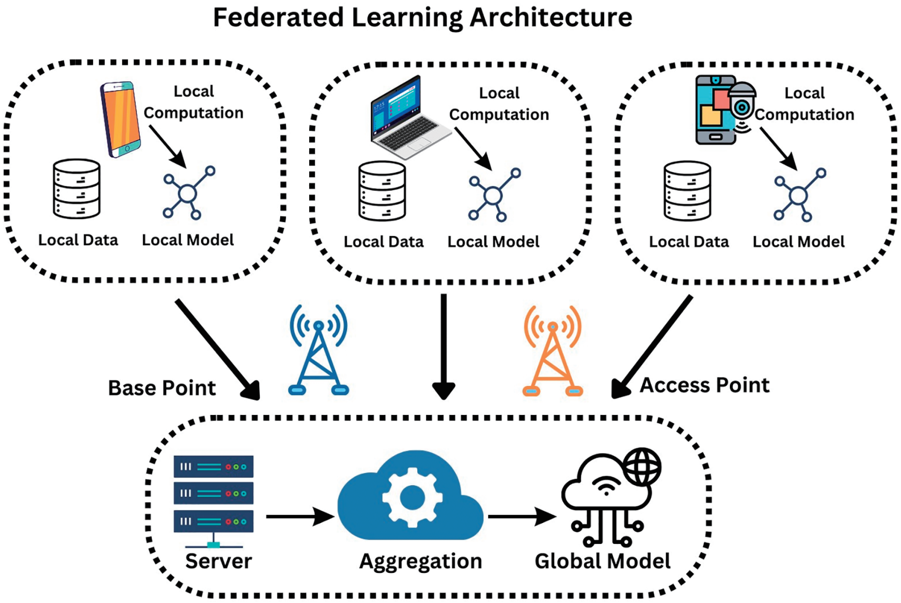

Figure 1: Enhancing security using federated learning architecture

This paper proposes a novel deep learning-based federated architecture for detecting zero-day attacks in IoT networks. The key contributions are (1) the design of the ZerodayX (Zeroday eXtended) model within a federated learning framework, enabling distributed and privacy-preserving training across edge nodes; (2) the integration of deep neural networks with collaborative learning to classify benign and malicious traffic; and (3) extensive evaluation on four benchmark datasets IoT-23, Kitsune ARP-MitM, Bot-IoT, and UNSW-NB15 demonstrating superior performance over existing methods in detection accuracy, efficiency, and scalability. The research accomplished in this work proposes a novel deep-learning model that identifies malicious and benign attacks and categorizes them as benign attacks. Collaborative and deep learning is used to identify and detect botnet attacks and train locally. It sends information to an aggregate server that sends a global model to reduce complexities and enhance performance using deep neural networks (DNN). This paper is organized as follows: Section 2, related work, presents the literature, and identifies research gaps and limitations. Section 3 discusses the proposed methodology and its algorithms. Section 4 discusses the results and the discussion. Finally, Section 5 concludes with a future direction.

2 Related Work

Rapid improvement in technology consequently increases the security threats. Technology is advancing day by day in daily life, and the concept of connecting all devices generates a massive volume of data through the internet. The network of networks, i.e., the internet, raised several security and privacy issues, increasing exponentially. Table 1 shows the previous efforts made by different authors with the datasets, the type of IDS, and their contributions.

Table 1: Related work contribution with different algorithms

Ref.	Dataset	Type of IDS	Attack mitigation	Algo./Model	Contributions
[13]	MAWI	CIDS		NEW	Port Centric CIDS.
	KDDCup99	NIDS	Anomaly detection	DAGMM	Network Anomaly.
	UNSW-NB15	NIDS	APT detection	CNN-LSTM	FL techniques using edge computing for APT detection.
	Bot-IoT, N-BaIoT	IoT network	FL	DNN	Detect zero-day botnet attack.
[14]	Historical web request and real time http requests	Web attacks		Encoder decoder RNN	Zero-day web attack detection.
[15]	CTU-13	DNN		ANN, NN, DNN	Design hidden layers on CTU-13 multiple NN efficiently.
[16]	NSL-KDD, CICids-17	Autoencoders		Unsupervised	Zero-day detection through autoencoder outlier.
[17]	NSL-KDD, CIDD	IDS		DNN	Transfer learning to evade ZDA by identifying new attack packet from normal packet.
[18]	ADFA-Linux	HADS		FDM and ML	(ADFA-WD and ADFA-WD: SAA) with illuminating descriptions were assembled with ML algorithms and the frequency distribution procedure.
[19]	Kitsune ARP-MITM	IDS		LSTM, CuDNNLSTM	Intrusion detection using LSTM and CuDNNLSTM.

Any vulnerability in the system leads to malicious activity, threats,and attacks, and damages the Confidentiality, Integrity, and Availability (CIA) triad. With expeditious enhancement and innovation in technology and online services, IoT connects and manages everything over the internet [13,20]. Cyberattacks, specifically persevering Internet of Things, are increasing and originate advanced security threats concerning zero-day attacks. To prevent IoT applications from intruding, there is a need for a mechanism that can protect IoT applications from unknown zero-day attacks to secure the end user's privacy [21]. Table 2 shows the presence of IoT, IIoT, and zero-day attacks and the contribution of the proposed method.

Table 2: Depicts the presence of ZDA with IoT and IIoT

Ref.	Contribution	Zero-day	IoT	IIoT
[22]	In this paper, a survey is conducted for a complete year in which authors discuss the advancement of IoT and a survey on IDS's ability for real-time analysis, alerts, and feedback.	✓	✓	x
[23]	The authors discussed two important features in this research. The first one is about the security threats in IoT networks and a few tools like packet sniffers, network analyzers, etc. The second feature is to compare many IDS solutions implemented through ML for validation, detection approaches, and algorithms.	x	x	x
[24]	The research focuses on the IoT Security domain and its threats. ML and DL approaches are discussed in it to identify the intrusions in IoT devices.	✓	✓	x
[25]	In this research, the authors discussed IoT networks' security problems, requirements, and solutions. Many machines and deep learning implementations are discussed in it.	✓	✓	x
[26]	The research is based on a comparison of IDS in IoT-based networks. A few characteristics are the focus of the research, namely, machine learning and hybrid learning mitigation, information detection, and protocol-based detection with statistics as well.	✓	x	x
[27]	Properties of IDPS are discussed with the deployment plan, possible attacks, different datasets, and techniques with methods of data processing.	✓	✓	x
Proposed Approach	ZerodayX	✓	✓	✓

During the attribute learning process, the learning models convert aspects of the network data to semantic attributes that distinguish between known attacks and benign behavior. During inference, the models create correlations between known and zero-day threats to detect malicious attempts [28]. Anomalous data points are those that do not fit into any of the pre-existing clusters or that do not satisfy certain inclusion requirements. In [29], the authors apply DRL using a supervised dataset containing labels without interacting with a live environment, as in classical DRL, and outline the modifications required to carry out these adaptations thoroughly. To achieve this, initially assimilate the states with the network properties and the actions with the intrusion labels.

Federated DL (collaborative learning, a decentralized machine learning approach that trains locally and sends information to an aggregate server, which then creates a global model) was introduced to reduce complexity and enhance performance using a deep neural network (DNN) to monitor and analyze network traffic. However, some of the threats are model poisoning, inference attacks, and communication security. In model poisoning the targeted model is a global model, in which malicious clients intentionally upload manipulated model updates, directly effect to the performance or the insertion of a backdoor. While in inference attacks, adversaries exploit shared model parameters to reconstruct sensitive information even without accessing raw data. While in a communication security model updates through the network, if the data is not encrypted so it is easy to intercept and an attacker can tempered packet easily, as a result parameter manipulation and man in the middle attack can implement. By distributing the learning process and enabling secure collaboration across devices, the proposed approach ensures high scalability, real-time responsiveness, and robustness, making it a viable solution for next-generation IoT security infrastructures.

In conclusion, deep learning and deep neural networks are particularly noteworthy machine learning approaches to be aware of (DNN). High-computational GPUs have changed several fields of computer science and engineering, including AI. In the area of cybersecurity, it is to be assumed that it will be comparable. There are numerous instances of DNNs being used in cybersecurity, even though IDS approaches are still in their infancy [30]. To examine and train the model, different datasets are used with different splits and ratios using 10-fold cross-validation. Different authors [20,31] put efforts into the literature to detect IDS, DoS, and zero-day. Still, existing techniques are not promising in detecting real-time attacks on networks and applications. Moreover, the literature has not discussed the identification of intruders using IoT devices and networks. To overcome the issue, there is a need for a mechanism to identify real-time detection of unknown zero-day attacks on IoT networks.

3 ZerodayX (Zeroday eXtended) Methodology

This study proposes a novel deep learning-based federated learning (FL) framework to detect multi-class cyberattacks, including zero-day threats, in IoT environments. The approach employs a Zero-day Short Learning (ZSL) model that learns distinguishing attributes from known attack classes during the training phase and maps these to unseen threats during inference. It also evaluates ML-based NIDS and detects new and unseen attacks. The attribute learning stage maps the network data feature to semantic attributes that differentiate the benign and known attacks. IoT applications and devices face the severe challenge of botnet attacks and zero-day attacks. When devices are vulnerable and compromised, attacks like DDoS and data pilferage in a network are called botnet attacks [32]. The paper proposed the implementation of FL to detect unknown, unseen, and unmatched attacks to identify the attacks. To ensure efficient and scalable training across distributed IoT devices, the proposed federated learning framework incorporates adaptive communication strategies tailored for resource-constrained environments. Specifically, gradient compression techniques are applied to minimize the size of model updates transmitted during each communication round. By reducing the dimensionality and precision of gradient vectors, this approach significantly lowers bandwidth requirements while preserving model convergence. Additionally, asynchronous update protocols are implemented to allow edge devices to send updates independently, without waiting for global synchronization. This design choice mitigates the impact of straggler devices and reduces communication delays, enhancing training throughput in heterogeneous networks. Together, these communication-efficient mechanisms support real-time collaboration among IoT nodes while maintaining the performance of the intrusion detection model trained on the IoT23 dataset. The proposed model Zeroday eXtended (ZerodayX) identifies the overall architecture to identify the attack using the trained model of FL, and issues the notification. It gives early notification of a zero-day attack before exploiting the system and damage it. Fig. 2 identifies the proposed novel architecture of ZerodayX.

The capability to detect suspicious activity in the system and generate an alert is known as intrusion detection, while intrusion reaction encompasses the remedies when detecting intrusion. An ID identifies the attacker on network and host-based systems. Data Collection and Pre-processing are mentioned in Algorithm 1.

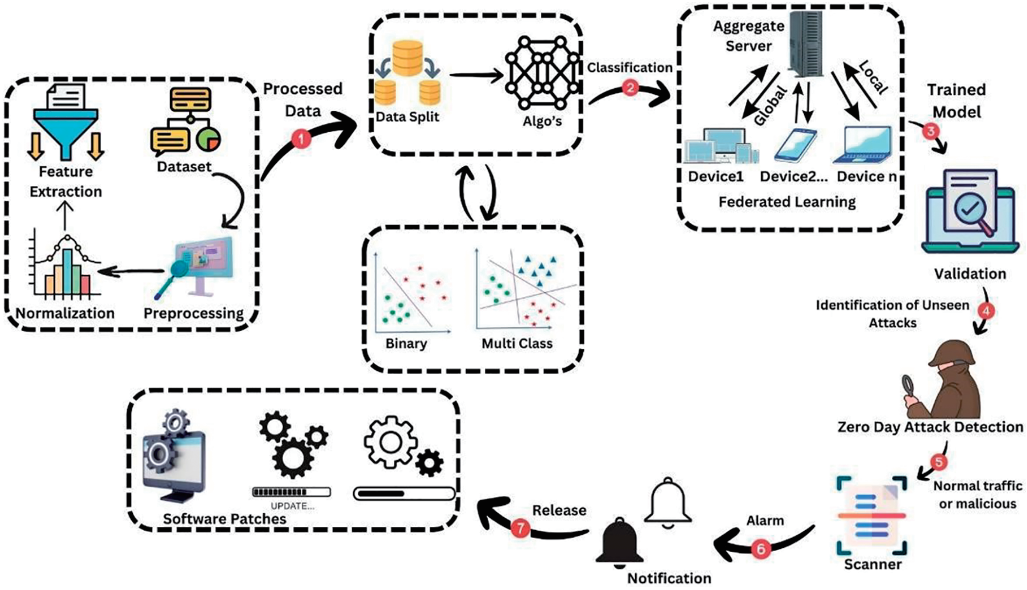

Figure 2: ZerodayX architecture determines the traffic as normal or an attack

Algorithm 1: Data collection and Pre-processing.

```
Input: Data obtained from Ton-IoT23
Output: Pre-processed IDS data
Parameters: Attacks, protocols (IP, TCP, UDP), packets, flag
Initialization: $DIDS = \varnothing$
$DIDS \leftarrow$ Parse Pre-processed file
Begin
for each $DOIDS$ in $CR$ do
    $D \leftarrow$ Load($DOIDS$)
    DIDS ← D['Text']
end for
for each $DOIDS$ in $DIDS$ do
    $RV \leftarrow L.\text{CleanAllM}(DIDS)$
    $RV \leftarrow L.\text{CleanAllD}(DIDS)$
end for
for each $DOIDS$ in $RV$ do
    $RV \leftarrow L.\text{DataImbalance}(RV)$
end for
End
```

To improve the efficiency of the model and curtail the dependence on supervised labeled data, integrate the self-supervised learning (SSL) technique, precisely contrastive pretraining in our FL architecture. Due to the inadequacy of annotated data in IoT environments and the uncertainty of zero-day attacks, SSL facilitates the model to learn robust and vigorous feature representations from unsupervised data by improving the compliance among distinguished views of the identical dataset. The detection of ZA and its notification is discussed in Algorithm 2.

Algorithm 2: Detection of ZA and notification.

```
Input: Take cleaned data from Algorithm 1
Output: Service violation which causes zero-day, Vulnerabilities of zero-day
Parameters: Same as parameters used in Algorithm 1
Initialization: $DIDS$ = Pre-processed file
Begin
$DO \leftarrow$ Split($DIDS$)
$FDO \leftarrow$ FL($DO$)
if $DO.Pattern == new$ then
    if $DO.Behavior == malicious$ then
        Show('its zero day')
        CreatePatches($DO.Pattern$)
    end if
end if
End
```

Implementation of multiclass classification **algorithm** is discussed in Algorithm 3.

Algorithm 3: Implementation of multi-class.

```
procedure MULTICLASS($L, D, F$)
    Input: Dataset in packets ($L, D, F$), where $L$ represents local, global, and aggregate, respectively
    Output: Type of attacks for the detection of zero-day cyber attacks
    Parameters: $L, D, F$
    Initialization: $L, D, F = \varnothing$
    Begin
    while devices $\leq n$ do
        if Maximum Transmission Unit (MTU) = 1 for all protocols then
            for $i = 0$ to data.length do
                $a[F] \leftarrow FL(D, B, F, H)$
            end for
            for $j = 0$ to $a[F]$.length do
                $a[D] \leftarrow a[F]$
            end for
        end if
    end while
    Return $a[D]$
    End
end procedure
```

This pretraining method allows the model to recognize underlying patterns and semantic structures before downstream fine-tuning on a limited number of labeled data. Contrastive pretraining enhanced the model's ability to detect unseen Zero-day attacks that had not yet been noticed by making it more sensitive to anomalous behaviors on the IoT23 dataset. This approach offers a promising direction for scalable, label-efficient intrusion detection systems that can adapt to evolving threat landscapes in real-world IoT deployments. Fig. 3 shows the accuracy and loss, and Fig. 4 depicts the precision and recall.

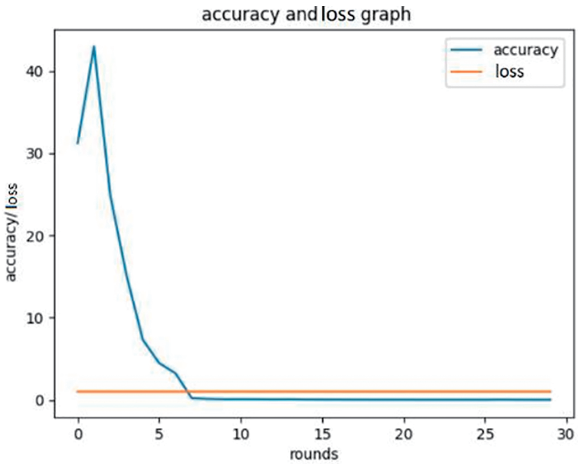

Figure 3: Loss and accuracy

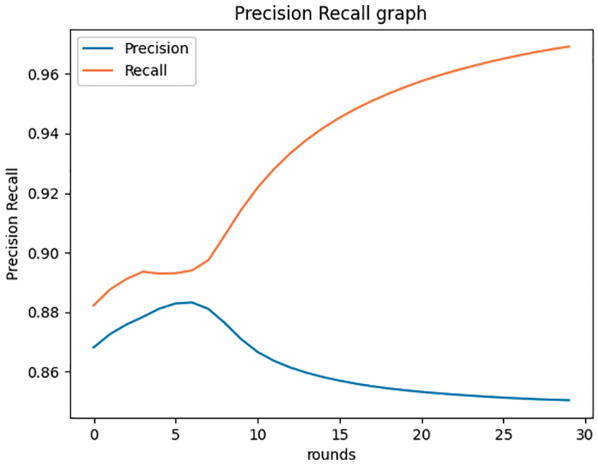

Figure 4: Precision and Recall

The weight scaling factor technique is used to compute how much client's training data is in comparison to the general training of the relative multitude of clients in the organization. The client's batch size to have the option to work out the quantity of data points. We then, at that point, continue to acquire the worldwide training data size which acts as the server's training data. finally, determined the scaling factor as a small portion. Scale model weights strategy is used to scale the local model's weights based on the previously determined values of their scaling factor determined in technique. Sum scaled weights strategies sum up the entirety of clients' scaled weights recently determined. We trained 30 global training circles by the comms round and tested the prepared global model after every correspondence round with our test data. Fig. 5 shows the analysis of the comparison.

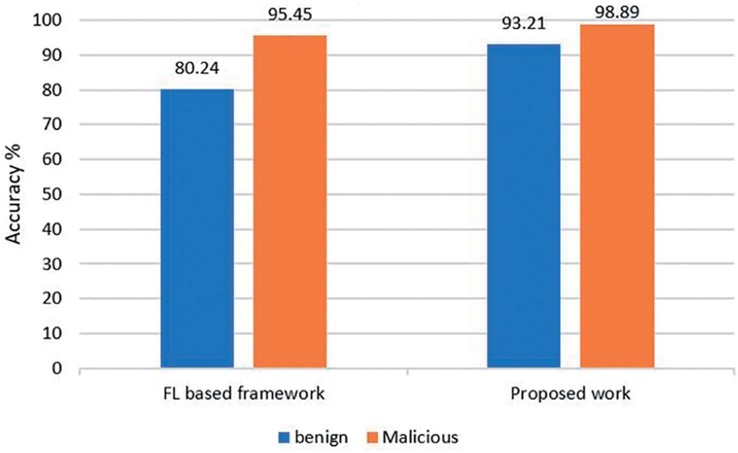

Figure 5: Comparative performance analysis of proposed work

4 Results and Discussions

We trained the proposed ensemble deep learning models using an efficient python library called Keras with python version 3.12.3. We used the following libraries in our performance evaluation tests: Numpy, Tensorflow, Scikit-learn, and Pandas. All three datasets were implemented on an FL-based IDS model with different train and test splits to evaluate the proposed method. A FL-based IDS system is proposed with the ability to identify the traffic flows as malicious or non-malicious. Fig. 6 shows the confusion matrix for federated learning, and Fig. 7 depicts the ROC curve.

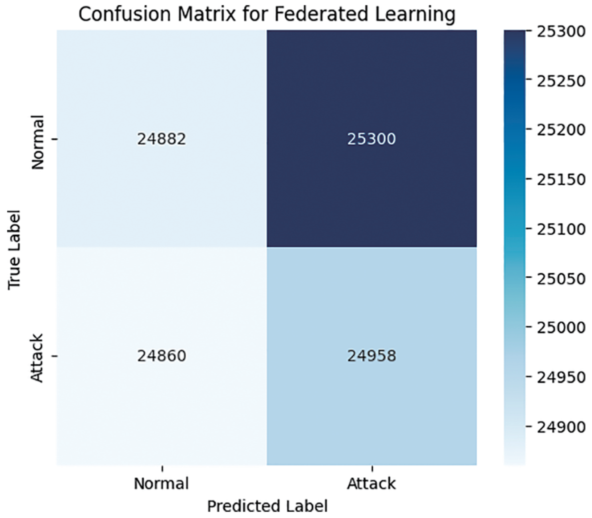

Figure 6: Confusion matrix for federated learning

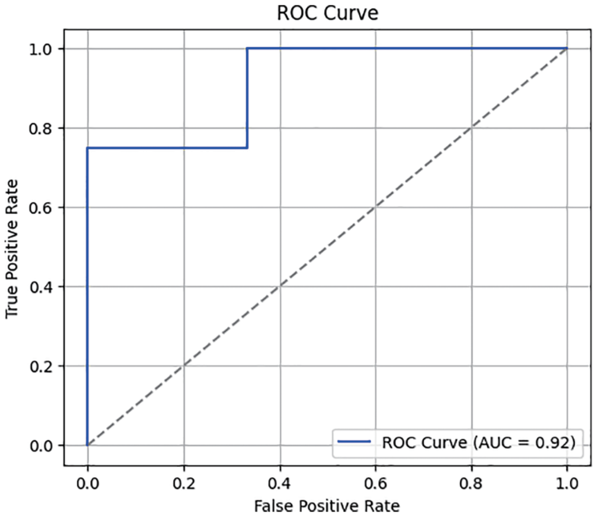

Figure 7: RoC Curve

In [33], a novel incorporated model with an underlying deep autoencoder (AE) for intrusion detection was proposed. Initially, AE was trained on normal traffic, then on handling anomalies. In order to extract valuable low-dimensional features for anomalous data without requiring a feature extraction training stage, the trained AE model is then used once more. The proposed model used the UNSW dataset N-BaIoT and determined 89% accuracy. In paper [19], the author used the state-of-the-art Kitsune ARP MitM dataset with feature extraction of 0 for non-malicious and 1 for malicious. LSTM and CuDNNLSTM were used in detecting and identifying intrusions with the following results, as mentioned in Table 3.

Table 3: Train and test accuracies with different layers with 50 epochs on 2500 k records [19]

Data size	Epoch	Batch	Layers	Loss	Train accuracy	Test accuracy	Time
2500 k	1	32	3	0.0153	99.46%	99.46%	22
2500 k	1	64	4	0.0091	92.00%	99.96%	40
2500 k	10	500	1	0.2502	92.4%	99.96%	22
2500 k	30	1000	2	0.221	91.00%	99.96%	45
2500 k	50	1000	3	0.183	99.90%	99.99%	71

The significance of assessing lightweight model versions in the context of federated learning (FL). As a result, we improved our experimental setup by using model compression methods, particularly quantization and structured pruning, to lower the communication overhead and computing complexity of FL. Table 4 identifies the performance evaluation metrics and details of the packet and its attack experiments conducted on the Bot-IoT.

Table 4: BoT IoT dataset

Split	Category	Precision	Recall	F1-score
70:30	DoS UDP	1.00	1.00	1.00
	DDoS TCP	1.00	1.00	1.00
	DDoS UDP	1.00	1.00	1.00
	Service Scan	1.00	1.00	1.00
	Os Fingerprint	1.00	1.00	1.00
	DoS TCP	1.00	1.00	1.00
	DoS HTTP	1.00	1.00	1.00
	DDoS HTTP	1.00	1.00	1.00
	Normal	1.00	1.00	1.00
	Macro Avg	1.00	1.00	1.00
	Weighted Avg	1.00	1.00	1.00
	Accuracy	1.00	1.00	0.98
80:20	DoS UDP	1.00	1.00	1.00
	DDoS TCP	1.00	1.00	1.00
	DDoS UDP	1.00	1.00	1.00
	Service Scan	1.00	1.00	1.00
	Os Fingerprint	1.00	1.00	1.00
	DoS TCP	1.00	1.00	1.00
	DoS HTTP	1.00	1.00	1.00
	DDoS HTTP	0.99	1.00	1.00
	Normal	1.00	1.00	1.00
	Macro Avg	1.00	1.00	1.00
	Weighted Avg	1.00	1.00	1.00
	Accuracy	1.00	1.00	0.978

Federated learning aims to jointly train a single global model without transferring sensitive information between clients. FL goes via the round-by-round exchange of training models between clients and the server to achieve this. The server first distributes a single global model to a group of sampled clients for each round, after which participating clients execute local optimization on the data. After the optimization, the server aggregates all locally trained models to update the model globally. The previously mentioned procedure for updating the global model can be expressed as follows: FL model is articulated in Eq. (1), where $w_{(t+1)}$ is the updated global model at round $t + 1$, k is the number of participating clients, n_{k} is the number of data samples held by client k, $\sum_{k=1}^{k} n_k$ is the total number of data samples across all clients, and w_k^t is the local model update from client k at round t. Loss function in FL is mentioned in Eq. (2).

$$w_{(t+1)} = \sum_{k=1}^{k} \frac{nk}{n} w_k^t \tag{1}$$

$$\min_{w} F(w) = \sum_{k=1}^{k} \frac{n_k}{n} F_k(w) \tag{2}$$

Table 5 identifies the performance evaluation metrics and details of the packet and its attack experiments conducted on the Arp-MitM.

Table 5: ARP-MITM dataset

Split	Type	Category	Precision	Recall	F1-score
70:30	Training set	Not-malicious	1.00	1.00	1.00
		Malicious	1.00	1.00	1.00
		Accuracy	1.00	–	–
		Macro avg	1.00	1.00	1.00
		Weighted avg	1.00	1.00	1.00
	Test set	Not-malicious	1.00	1.00	1.00
		Malicious	1.00	1.00	1.00
		Macro avg	1.00	1.00	1.00
		Weighted avg	1.00	1.00	1.00
		Accuracy	1.00	–	–
80:20	Training set	Not-malicious	1.00	1.00	1.00
		Malicious	1.00	1.00	1.00
		Accuracy	1.00	–	–
		Macro avg	1.00	1.00	1.00
		Weighted avg	1.00	1.00	1.00
	Test set	Not-malicious	1.00	1.00	1.00
		Malicious	1.00	1.00	1.00
		Macro avg	1.00	1.00	1.00
		Weighted avg	1.00	1.00	1.00
		Accuracy	1.00	–	–

The resulting lightweight FL model showed its efficacy in edge situations by maintaining a good detection performance with 98% accuracy while reducing model size and inference time. These results validate the model's suitability for resource-aware, real-time intrusion detection in federated IoT settings. Although several studies have documented good detection performance using the IoT-23 dataset, frequently surpassing 95% accuracy or F1-score, the majority of these assessments are carried out on known classes using data that has been randomly split. Statistical measures like standard deviation and confidence intervals are rarely reported in studies, and hypothesis testing is rarely done. In particular, evaluating the model's generalizability to zero-day attacks is challenging due to the lack of such statistical precision. Table 6 identifies the performance evaluation metrics and details of the packet and its attack experiments conducted on the UNSW datasets.

Table 6: UNSW dataset

Split	Type	Category	Precision	Recall	F1-score
70:30	Training set	Not-malicious	1.00	1.00	1.00
		Malicious	1.00	1.00	1.00
		Accuracy	1.00	–	–
		Macro avg	1.00	1.00	1.00
		Weighted avg	1.00	1.00	1.00
	Test set	Not-malicious	1.00	1.00	1.00
		Malicious	1.00	1.00	1.00
		Weighted avg	1.00	1.00	1.00
		Macro avg	1.00	1.00	1.00
		Accuracy	1.00	–	–
80:20	Training set	Not-malicious	1.00	1.00	1.00
		Malicious	1.00	1.00	1.00
		Accuracy	1.00	–	–
		Macro avg	1.00	1.00	1.00
		Weighted avg	1.00	1.00	1.00
	Test set	Not-malicious	1.00	1.00	1.00
		Malicious	1.00	1.00	1.00
		Weighted avg	1.00	1.00	1.00
		Macro avg	1.00	1.00	1.00
		Accuracy	1.00	–	–

Loss of federated learning rounds on different datasets is mentioned in Fig. 8, and the time complexity of ZerodayX is illustrated in Fig. 9.

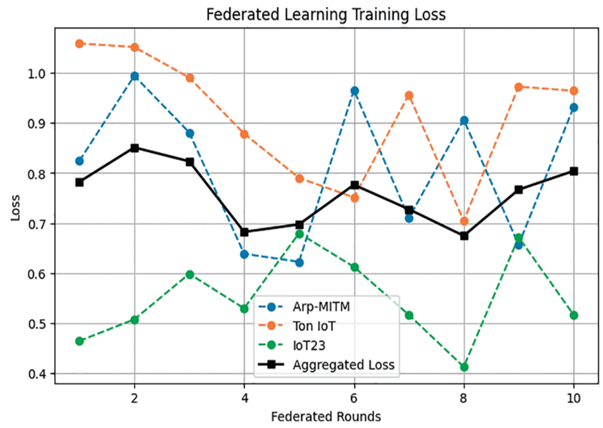

Figure 8: Loss of federated learning

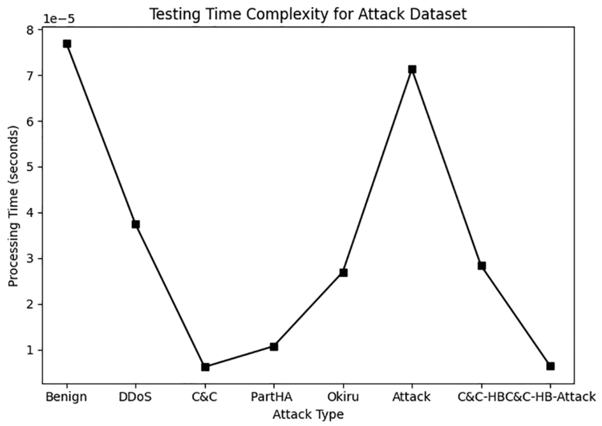

Figure 9: Time complexity of the proposed architecture

Table 7 shows the overall performance on the Bot-IoT, Arp-MitM, and UNSW datasets. Table 8 mentions the overall evaluation metrics, where C&C-Heartbeat-Attack shows the highest precision, recall, and F1 score and confirms the model's applicability for real-time intrusion detection on edge devices, thereby addressing concerns related to deployment feasibility in constrained environments.

Table 7: Federated Learning metrics on ARP-MITM, Bot-IoT and UNSW

Datasets	Split	Accuracy	Recall	Precision	F1 score
Metric (%)					
Bot-IoT	70:30	99.95	99.67	99.92	99.28
Bot-IoT	80:20	99.94	99.45	99.62	99.45
ARP-MitM	70:30	87.93	62.81	79.56	70.01
ARP-MitM	80:20	94.34	56.78	82.67	66.45
UNSW-NB15	70:30	87.94	97.50	85.45	96.78
UNSW-NB15	80:20	88.07	99.85	85.78	95.43

Table 8: Shows the overall evaluation metrics of the IoT23 dataset with a variety of attacks

Class	Precision	Recall	F1-Score
Benign	79.67	78.56	83.76
DDoS	98.56	68.89	89.65
C&C	45.56	75.45	68.56
PartOfAHorizontalportScan	49.34	64.34	69.89
Okiru	79.82	92.76	90.87
Attack	73.78	46.67	62.56
C&C-HeartBeat	42.87	23.45	92.67
C&C-Heartbeat-Attack	100.00	100.00	100.00

5 Conclusion

This study proposes multi-class zero-day attack detection in IIoT devices using federated learning. The proposed model uses several datasets, ARP-MITM, Bot-IoT, and UNSW-NB15 with an accuracy of 94.34%, 99.94%, and 87.94%, respectively. Ton-IoT and IoT 23 datasets are also used to evaluate the identification of zero-day attacks using deep Federated learning (DFL). The federated learning model is implemented using locally trained data. Then, the model will reach the aggregated server of multiple devices using a multiclass implementation. Furthermore, validation is also performed after FL and detection. Finally, patches are implemented after notification to resume the already exploited system. Hence, the model is very efficient and can detect and identify zero-day attacks on the Internet of Things and its devices. In the future, implementing advanced federated learning can enhance the performance and classification of zero-day attack detection.

Acknowledgement: Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2025R97), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding Statement: This research was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2025R97), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contributions: Study conception and design: Misbah Anwer, Ghufran Ahmed; Data collection: Misbah Anwer, Maha Abdelhaq; Analysis and interpretation of results: Misbah Anwer, Raed Alsaqour, Shahid Hussain, Adnan Akhunzada; Draft manuscript preparation: Misbah Anwer, Maha Abdelhaq; Supervision, review, and editing: Ghufran

Ahmed, Adnan Akhunzada; Validation: Ghufran Ahmed, Shahid Hussain; Funding acquisition: Ghufran Ahmed, Adnan Akhunzada, Maha Abdelhaq, Raed Alsaqour. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The datasets utilized in this study are publicly accessible and can be obtained from the following sources: IoT23 Dataset, Available at: https://www.stratosphereips.org/datasets-iot23 (accessed on 04 August 2025); ARP-MITM Dataset, Available at: https://www.kaggle.com/datasets/ymirsky/network-attack-dataset-kitsune (accessed on 04 August 2025); Bot-IoT Dataset, Available at: https://research.unsw.edu.au/projects/bot-iot-dataset (accessed on 04 August 2025); UNSW-NB15 Dataset, Available at: https://research.unsw.edu.au/projects/unsw-nb15-dataset (accessed on 04 August 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest to report regarding the present study.

References

1. Zhang J, Ling Y, Fu X, Yang X, Xiong G, Zhang R. Model of the intrusion detection system based on the integration of spatial-temporal features. Comput Secur. 2020;89:101681.
2. Sowah RA, Ofori-Amanfo KB, Mills GA, Koumadi KM. Detection and prevention of man-in-the-middle spoofing attacks in MANETs using predictive techniques in artificial neural networks (ANN). J Comput Netw Commun. 2019;2019(11):e4683982–14. doi:10.1155/2019/4683982.
3. Korba AA, Boualouache A, Ghamri-Doudane Y. Zero-x: a blockchain-enabled open-set federated learning framework for zero-day attack detection in IoV. IEEE Trans Vehicular Technol. 2024;73(9):12399–414. doi:10.1109/tvt.2024.3385916.
4. Sharma A, Singh UK. Cloud computing security through detection & mitigation of zero-day attack using machine learning techniques. In: Natural Language Processing for Software Engineering. Beverly, MA, USA: Scrivener Publishing LLC.; 2025. p. 357–88.
5. Bikila DD, Čapek J. Machine learning-based attack detection for the internet of things. Future Gener Comput Syst. 2025;166(13):107630. doi:10.1016/j.future.2024.107630.
6. Touré A, Imine Y, Semnont A, Delot T, Gallais A. A framework for detecting zero-day exploits in network flows. Comput Netw. 2024;248(1):110476. doi:10.1016/j.comnet.2024.110476.
7. 160 Cyber Security S. Cybersecurity statistics: updated report 2025 [cited 2025 Aug 16]. Available from: https://www.getastra.com/blog/security-audit/cyber-security-statistics/.
8. Diro A, Chilamkurti N. Leveraging LSTM networks for attack detection in fog-to-things communications. IEEE Commun Mag. 2018;56(9):124–30. doi:10.1109/mcom.2018.1701270.
9. Sebbar A, Zkik K, Baddi Y, Boulmalf M, Kettani MDECE. MitM detection and defense mechanism CBNA-RF based on machine learning for large-scale SDN context. J Ambient Intel Humaniz Comput. 2020;11(12):5875–94. doi:10.1007/s12652-020-02099-4.
10. Khan LU, Saad W, Han Z, Hossain E, Hong CS. Federated learning for internet of things: recent advances, taxonomy, and open challenges. IEEE Commun Surv Tutor. 2021;23(3):1759–99.
11. Venkatraman S, Surendiran B. Adaptive hybrid intrusion detection system for crowd sourced multimedia internet of things systems. Multimed Tools Appl. 2020;79(5–6):3993–4010. doi:10.1007/s11042-019-7495-6.
12. Lopez-Martin M, Carro B, Sanchez-Esguevillas A. Application of deep reinforcement learning to intrusion detection for supervised problems. Expert Syst Appl. 2020;141(6):112963. doi:10.1016/j.eswa.2019.112963.
13. Blaise A, Bouet M, Conan V, Secci S. Detection of zero-day attacks: an unsupervised port-based approach. Comput Netw. 2020;180(1):107391. doi:10.1016/j.comnet.2020.107391.
14. Bouacida N, Mohapatra P. Vulnerabilities in federated learning. IEEE Access. 2021;9:63229–49. doi:10.1109/access.2021.3075203.
15. Cao Z, Zhao Z, Shang W, Ai S, Shen S. Using the ToN-IoT dataset to develop a new intrusion detection system for industrial IoT devices. Multimed Tools Appl. 2025;84(16):16425–53. doi:10.1007/s11042-024-19695-7.

16. Dutta V, Choraś M, Pawlicki M, Kozik R. A deep learning ensemble for network anomaly and cyber-attack detection. Sensors. 2020;20(16):4583. doi:10.3390/s20164583.
17. Ghelani D. Cyber security, cyber threats, implications and future perspectives: a review. Am J Sci Engi Technol. 2022;3(6):12–9.
18. Vu L, Nguyen QU, Nguyen DN, Hoang DT, Dutkiewicz E. Deep transfer learning for IoT attack detection. IEEE Access. 2020;8:107335–44. doi:10.1109/access.2020.3000476.
19. Anwer M, Ahmed G, Akhunzada A, Siddiqui S. Intrusion detection using deep learning. In: 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME); 2021 Oct 7–8; Flic en Flac, Mauritius. p. 1–6.
20. Popoola SI, Ande R, Adebisi B, Gui G, Hammoudeh M, Jogunola O. Federated deep learning for zero-day botnet attack detection in IoT-Edge devices. IEEE Internet Things J. 2022;9(5):3930–44. doi:10.1109/jiot.2021.3100755.
21. Tang R, Yang Z, Li Z, Meng W, Wang H, Li Q, et al. ZeroWall: detecting zero-day web attacks through encoder-decoder recurrent neural networks. In: IEEE INFOCOM 2020-IEEE Conference on Computer Communications; 2020 Jul 6–9; Toronto, ON, Canada. p. 2479–88.
22. Haider W, Creech G, Xie Y, Hu J. Windows based data sets for evaluation of robustness of host based intrusion detection systems (IDS) to zero-day and stealth attacks. Future Internet. 2016;8(4):29. doi:10.3390/fi8030029.
23. Mahmood M, Shafi Q. A smart IDS in IoT system to detect zero-day intrusions using automated signature update. Forthcoming. 2023. doi:10.21203/rs.3.rs-3014508/v1.
24. Sarhan M, Layeghy S, Moustafa N, Portmann M. Cyber threat intelligence sharing scheme based on federated learning for network intrusion detection. J Netw Syst Manage. 2023;31(1):3. doi:10.21203/rs.3.rs-1631421/v1.
25. Yao W, Hu L, Hou Y, Li X. A lightweight intelligent network intrusion detection system using one-class autoencoder and ensemble learning for IoT. Sensors. 2023;23(8):4141. doi:10.3390/s23084141.
26. Ali S, Rehman SU, Imran A, Adeem G, Iqbal Z, Kim KI. Comparative evaluation of AI-based techniques for zero-day attacks detection. Electronics. 2022;11(23):3934. doi:10.3390/electronics11233934.
27. Guo Y. A review of machine learning-based zero-day attack detection: challenges and future directions. Comput Commun. 2023;198(10):175–85. doi:10.1016/j.comcom.2022.11.001.
28. Sarhan M, Layeghy S, Gallagher M, Portmann M. From zero-shot machine learning to zero-day attack detection. Int J Inform Secur. 2023;22(4):947–59. doi:10.21203/rs.3.rs-2097775/v1.
29. Kumar V, Sinha D. A robust intelligent zero-day cyber-attack detection technique. Complex Intel Syst. 2021;7(5):2211–34.
30. AL-Shatnawi A, Al-Saqqar F, Alhusban S. A holistic model for recognition of handwritten arabic text based on the local binary pattern technique. Int J Interac Mobile Technol (iJIM). 2020;14(16):20–34. doi:10.3991/ijim.v14i16.16005.
31. Zoppi T, Ceccarelli A, Bondavalli A. Unsupervised algorithms to detect zero-day attacks: strategy and application. IEEE Access. 2021;9:90603–15. doi:10.1109/access.2021.3090957.
32. Ali M, Siddique A, Hussain A, Hassan F, Ijaz A, Mehmood A. A sustainable framework for preventing IoT systems from zero day DDoS attacks by machine learning. Int J Emerg Technol. 2021;12(1):116–21.
33. Alaghbari KA, Lim HS, Hanif Md Saad M, Yong YS. Deep autoencoder-based integrated model for anomaly detection and efficient feature extraction in IoT networks. IoT. 2023;4(3):345–65. doi:10.3390/iot4030016.