IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 March 2024, accepted 8 April 2024, date of publication 12 April 2024, date of current version 26 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3388149

RESEARCH ARTICLE

Open-Set Recognition in Unknown DDoS Attacks Detection With Reciprocal Points Learning

CHIN-SHIUH SHIEH[1], (Member, IEEE), FU-AN HO[1], MONG-FONG HORNG[1], (Member, IEEE), THANH-TUAN NGUYEN[2], AND PRASUN CHAKRABARTI[3], (Senior Member, IEEE)

[1]Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 82444, Taiwan
[2]Department of Electronics and Automation Engineering, Nha Trang University, Nha Trang 650000, Vietnam
[3]Department of Computer Science and Engineering, Sir Padampat Singhania University, Udaipur, Rajasthan 313601, India

Corresponding author: Mong-Fong Horng (mfhorng@nkust.edu.tw)

This work was supported in part by the National Science and Technology Council, Taiwan, under Grant NSTC 112-2221-E-992-045 and Grant NSTC 112-2221-E-992-057-MY3; and in part by Genie Networks Ltd.

ABSTRACT The internet, a cornerstone of modern life, has profound implications across personal, business, and society. However, its widespread use has posed challenges, especially concerning privacy and cybersecurity. Besides, the threats on the internet are increasing in terms of danger, intensity, and complexity. Distributed denial-of-service (DDoS) attacks have emerged as a common and dangerous cybersecurity threat capable of disabling the network systems of targeted organizations and services. Therefore, various security strategies, such as firewalls and intrusion detection systems (IDS), are employed to protect against DDoS attacks. Enhancing the defensive capabilities of IDS systems through machine learning (ML) and deep learning (DL) technologies is a significant trend nowadays. However, despite notable successes, detecting DDoS attacks using ML and DL technologies still faces challenges, especially with Unknown DDoS Attacks. In this research, the primary goal is to address the unknown DDoS detection problem through efficient and advanced techniques. Our proposed method, CNN-RPL, integrates Convolutional Neural Network (CNN) with Reciprocal Points Learning (RPL), a novel Open-Set Recognition (OSR) technology. This model can effectively handle both known and unknown attacks. The CNN-RPL model demonstrates excellent results, achieving an accuracy exceeding 99.93% against known attacks in the CICIDS2017 dataset. Simultaneously, the model achieves a commendable average accuracy of up to 98.51% against unknown attacks in the CICDDoS2019 dataset. In particular, the CNN-RPL model simplifies the architecture of the deep neural network by significantly reducing the number of training parameters without compromising defense capabilities. Therefore, our proposed method is genuinely efficient, particularly flexible, and lightweight compared to traditional methods. This can equip organizations and businesses with a highly applicable yet powerful security approach against the evolving complexities in the network space.

INDEX TERMS Cybersecurity, unknown attack detection, distributed denial-of-service (DDoS), open-set recognition (OSR), reciprocal points learning (RPL), machine learning, deep learning, incremental learning, convolutional neural networks (CNN).

I. INTRODUCTION

The DDoS attack overwhelms the target system's resources and network bandwidth by generating numerous service requests or traffic. The goal is to overwhelm the target system, rendering it incapable of managing the increased load and affecting its availability. Critical attributes of DDoS attacks encompass substantial decentralization, a notable surge in attack traffic, and the concealment of attack sources. These characteristics collectively amplify the detrimental effects

The associate editor coordinating the review of this manuscript and approving it for publication was Vicente Alarcon-Aquino.

of DDoS attacks on targeted systems, potentially leading to service unavailability, data breaches, and disruptions to business operations.

The fundamental principle behind DDoS attacks is to occupy the target system's resources, such as bandwidth, processor resources, and memory, beyond their normal operational limits. Attackers typically utilize multiple infected computers or other devices, forming a collective known as a ''botnet''. Through these botnets, attackers can centrally orchestrate and launch their attacks. The internet's thriving information activities and the continuous development of emerging services have led to the rapid evolution of DDoS attack methods. This poses a significant challenge to traditional IDS. Conventional defense mechanisms rely on experts for identification. Attack samples are required before identification. These attack samples also depend on investigative work by personnel.

According to the report by Yoachimik et al. [1], in the third quarter of 2023, Cloudflare faced the most complex and sustained DDoS attacks in its history. Furthermore, Cloudflare successfully mitigated thousands of high-capacity HTTP DDoS attacks, with 89 surpassing a rate of one billion requests per second (rps). The most significant attack peaked at 2.01 billion rps, tripling the previous record of 71 million rps. The overall HTTP DDoS attack traffic volume increased by 65% compared to the previous quarter. Similarly, L3/4 DDoS attacks also rose by 14%, with multiple attacks reaching the Tbps level per second. The most significant attack targeted Cloudflare's free DNS resolver 1.1.1.1, reaching a peak of 2.6 Tbps. In October 2023, Cloudflare, Google, and Amazon AWS disclosed a new zero-day vulnerability known as ''HTTP/2 Rapid Reset'' [2], [3], [4], [5]. This vulnerability exploits weaknesses in the HTTP/2 protocol, leading to the generation of massive and highly disruptive DDoS attacks. The HTTP/2 Rapid Reset vulnerability poses a significant threat to online services and highlights the ongoing challenges in securing internet protocols against evolving attack vectors. Security experts and service providers are actively working to address and mitigate the impact of this vulnerability on their platforms.

Thus, defense against DDoS is urgent, and researchers have proposed various methods for DDoS defense [6], [7], [8] including firewall filtering, load balancing, cloud defense, and intrusion detection methods that integrate machine learning and deep learning. Studies have shown that these intrusion detection methods exhibit excellent defense capabilities, combining machine learning and deep learning. Moreover, traditional deep learning and machine learning models in DDoS detection systems tend to deal with Close-Set training data. That means when the model in the training phase only encounters known attack types and normal traffic, its performance is usually quite good. Nevertheless, when these models encounter previously unseen attack forms or significantly different network traffic in real-world applications, their performance is often limited, especially when facing Open-Set data in real-life scenarios.

This phenomenon has created a demand for improving the generalization capability and practicality of DDoS attack detection systems. In real-world network environments, attackers continually innovate and develop new attack methods, which may make it difficult for traditional Close-Set training models to cope with unknown attack forms. Therefore, this research is motivated to enhance existing DDoS attack detection systems to make them more adaptable to unknown attacks, thereby improving the accuracy and performance of the models when dealing with Open-Set data.

A. RESEARCH OBJECTIVE

This research aims to propose an innovative model for unknown DDoS attacks within the framework of OSR to address novel forms of unknown DDoS attacks called CNN-RPL. We integrate CNN with RPL to achieve this objective, constructing an attack detection system with robust generalization capabilities. Through CNN, features are extracted from DDoS attack traffic, enhancing the model's accuracy in identifying known attacks and normal traffic. Simultaneously, this helps capture the underlying structure of attack traffic, enabling the model to better differentiate between different attack types. Introducing RPL, a OSR technology, equips the model with adaptability to unknown DDoS attack forms. Utilizing CNN deep features as samples, Reciprocal Points are identified for each known category, constraining the distribution of known categories to encourage the model to distinguish attack samples from their corresponding normal samples. This learning approach enhances the model's discriminate ability, allowing it to detect and respond effectively to new attacks in advance.

Our contributions mainly focus on the following:

- This research proposed a novel IDS model to classify known and detect unknown DDoS attacks with the CNN-RPL OSR method.
- The proposed model is designed to adapt dynamically, allowing it to learn incrementally from filtered-out unknown attack data labeled by security experts for expanding defense capabilities. This model not only benefits from the continuous integration of updated data but also retains core functionality with a high rate of defense against learned information.
- The proposed model has a relatively compact parameter set, which gives it significant operational flexibility while still ensuring a high defense capability with an average accuracy of up to 98.51%.

II. RELATED WORK

A. MACHINE LEARNING IN DDoS DETECTION

Bansal and Sanmeet [9] discuss the challenges of designing an efficient intrusion detection system in the fast-growing digital era. It emphasizes the importance of security in the IT sector and highlights machine learning algorithms, particularly XGBoost, for detecting DDoS attacks. The paper compares the performance of XGBoost with other classifiers

such as AdaBoost, Naive Bayes, MLP, and KNN. It reports that XGBoost outperforms these classifiers in detecting DDoS attacks. The document also provides details of the methods and materials used in model building, experimental results, and a comparison with other classifiers. It concludes by discussing future work in the field of intrusion detection systems. Furthermore, the XGBoost classifier outperforms other classifiers such as AdaBoost, Naive Bayes, MLP, and KNN in detecting DDoS attacks. The True Positive Rate (TPR) for XGBoost is reported to be 0.97, which is higher than the TPR rates of the other classifiers mentioned in the document. This indicates that XGBoost is more accurate in detecting DDoS attacks than traditional approaches.

Deep learning-based IDS differs from traditional methods in that it does not require a lot of attack signatures or a list of normal behaviors to generate detection rules. Instead, deep learning defines intrusion features by itself through training empirical data. This contrasts traditional methods that rely on predefined signatures or rules for detecting intrusions. The use of deep learning allows the system to learn its features and adapt to new attack patterns, addressing the limitations of traditional intrusion detection systems.

Kin et al. [10] designed a CNN model with two convolutional layers and two max-pooling layers. This architecture is suitable for processing images, and they converted the network traffic dataset into images for this purpose. The experimental results indicate that their CNN model achieved high accuracy in detecting benign and attack data in the CIC-2018 dataset. They have also evaluated the dataset using an RNN model for multi-class classification. Their CNN model outperformed the RNN model in terms of accuracy when applied to CIC-2018. This suggests that for this specific task and dataset, CNNs are more effective. In the end, Balancing the dataset and optimizing model architecture are essential steps in improving the robustness and accuracy of intrusion detection systems, especially given the evolving nature of cyber threats.

Kin et al. [11] developed a Convolutional Neural Network based model for the detection of DDoS attacks using the KDD [12] and CSE-CIC-IDS 2018 [13] datasets. Focusing on improving intrusion detection systems, the study addressed the challenge of distinguishing DoS attacks, including advanced types, from benign traffic. Unlike typical binary classifications, it encompassed both binary and multiclass classifications, enabling the identification of different attack categories within KDD. To enhance model performance, the researchers generated RGB and grayscale intrusion images and conducted extensive experiments with various hyperparameters. Notably, RGB images consistently outperformed grayscale ones, and the number of convolutional layers had a significant impact on accuracy. The study also compared the CNN model's performance with a RNN, with the CNN model demonstrating superior accuracy in both binary and multiclass classifications. This research contributes to the advancement of intrusion detection systems, particularly in identifying complex DoS attacks.

Hu et al. [14] aim to improve the recognition rate of the intrusion detection system. The problem addressed is the low average recognition rate of a multi-class intrusion detection system that uses CNNs for classification. Intrusion detection systems are used to identify and classify potentially malicious activities in computer networks. The proposed method uses the Fruit Fly Optimization Algorithm (FOA) during the pre-training process. FOA is used to help balance the representation of different classes in the dataset. Imbalance is a common issue in multi-class DDoS classification problems. The NSL-KDD [15] dataset is used to test the model. The NSL-KDD dataset is a commonly used dataset for evaluating intrusion detection systems. Testing the model on this dataset allows for a comparison of the proposed approach's performance with a CNN method that does not include data equalization. The proposed method is more accurate when compared to a CNN method that does not address class imbalance or data equalization. This suggests that the combination of CNNs, FOA, and the resampling method has resulted in an improved recognition rate for the intrusion detection system.

B. OPEN-SET RECOGNITION

In Yang et al.'s survey [16], OSR is a sub-topic of Out-of-Distribution (OOD) detection that focuses on recognizing known classes while rejecting unknown or out-of-distribution samples. In other words, OSR is concerned with recognizing samples from a fixed set of classes while being able to reject samples that do not belong to any of those classes. The main challenge in OSR is that the model must distinguish between known and unknown samples, even if it has not seen any examples of unknown samples during training. This requires the model to learn to identify the boundaries between the known and unknown regions of the input space and to make decisions based on the confidence or uncertainty of its predictions. OSR has many practical applications, such as in image classification, object detection, natural language processing, and DDoS attack detection, where it is important to recognize known classes while rejecting unknown or irrelevant samples.

The following is a list of the main categories of OSR methods along with a brief description of each:

1) Classification-based methods: These methods use a threshold on the output of a classifier to distinguish between known and unknown classes. Examples include the Extreme Value Theory (EVT)-based [17] uncertainty calibration, addressing neural network overconfidence with Compact Abating Probability (CAP) and EVT methods. EVT-Free [18] Confidence Enhancement methods provide alternative empirical successes. Furthermore, Unknown Generation [19] employs image synthesis and boundary adjustment strategies.
2) Distance-based methods: These methods use a distance metric to measure the similarity between a test sample

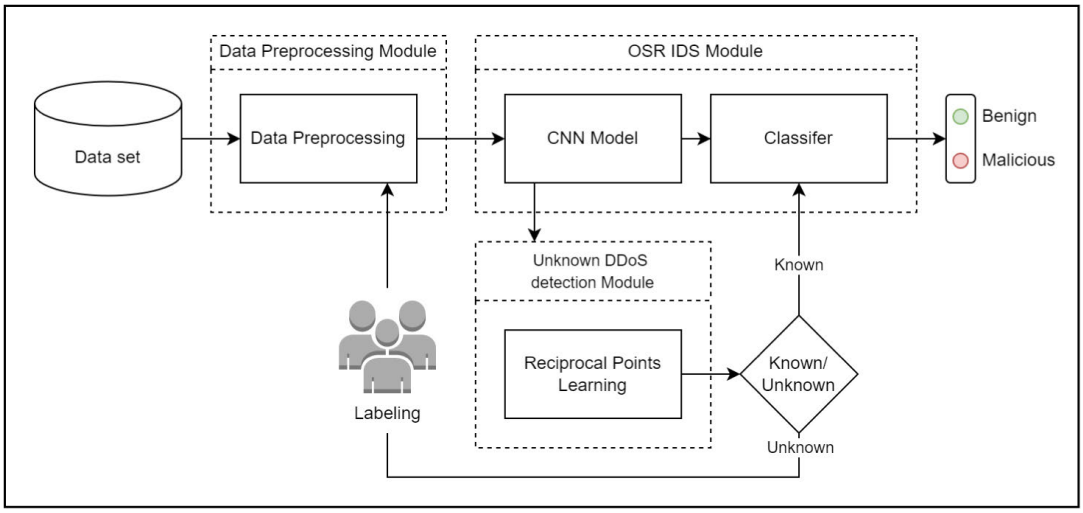


FIGURE 1. The diagram of CNN-RPL.

and the training data. The CROSR's [20] fusion of visual embeddings adds depth to distance computation. GMVAE's [21] Gaussian mixture model and the adaptation of nearest neighbors for OSR exemplify practical and efficient techniques.

3) Reconstruction-based models: Sparse Representation Method [22] assume that normal data can be accurately reconstructed using a limited set of basis functions, whereas anomalous data should suffer from higher reconstruction costs, thus generating a dense representation. Reconstruction-Error Method [23] assume that a reconstruction model trained on the normal data will produce higher-quality outcomes for normal test samples than anomalies. Deep reconstruction models include AEs, VAEs, GANs, and U-Net, which can all be used as the backbone for this method.
4) Hybrid methods: These methods combine multiple techniques to improve the performance of OSR. Examples include the Hybrid Threshold-Distance method and the Hybrid GAN-Distance method.

C. OPEN-SET RECOGNITION ON UNKNOWN DDOS DETECTION

In recent years, various OSR-based methods of detecting unknown DDoS attacks have been proposed with the increasing prevalence of DDoS attacks. These methods can be categorized into two classes: one based on integrating deep learning and machine learning and the other relying on Generative Adversarial Network (GAN) identification methods.

On the deep learning and machine learning side, a novel framework was proposed in 2021 [24], integrating BI-LSTM, Gaussian Mixture Model (GMM), and incremental learning. The GMM captures unknown traffic, which traffic engineers discern and label. These labeled instances are subsequently incorporated into the framework as additional training samples. In 2022, another novel framework was proposed [25], leveraging reconstruction error and distributing hidden layer characteristics to detect unknown DDoS attacks. The architecture employs DHRNet. It is reimagined as a 1D integrated neural network, incorporating a loss function with a Spatial Location Constraint Prototype Loss (SLCPL) to address Open-Set risks. In the subsequent stage, a One-Class SVM based on a random gradient descent approximation was employed to recognize unknown patterns.

On the GAN side, in 2022, the study proposes a novel DDoS detection framework featuring GAN with Dual Discriminators (GANDD) [26]. The additional discriminator is specifically designed to identify adversarial DDoS traffic. Experimental results indicate that GANDD can effectively solve adversarial DDoS attacks. The model is trained using adversarial DDoS traffic synthesized by GP-WGAN and is compared with three other deep learning technologies: DNN, LSTM, and GAN. The GANDD model outperforms the other deep learning models, demonstrating its efficacy with a TPR of 84.3%.

Moreover, a novel GAN with a symmetrically constructed generator and discriminator defense system (SDGAN) [27] was proposed in the same year. Both symmetric discriminators aim to identify adversarial DDoS traffic simultaneously. Experimental results demonstrate that the suggested SDGAN is effective against adversarial DDoS attacks. While training on adversarial DDoS data generated by CycleGAN, SDGAN outperforms other machine learning models, achieving a TPR of 87.2%. Furthermore, a comprehensive test evaluates SDGAN's ability to defend against unseen adversarial threats, where it remains effective with a TPR of around 70.9%, compared to RF's 9.4%.

Finally, in 2023, the CNN-Geo [28] was proposed. The CNN-Geo framework leverages a CNN construction focusing on Geometrical metrics, employing deep learning techniques to enhance accuracy in identifying DDoS attacks.

Additionally, the authors have incorporated an incremental learning module capable of efficiently integrating novel, unknown traffic identified by telecommunication experts during the monitoring process.

D. RECIPROCAL POINTS LEARNING AND COMPARISION WITH CONTEMPORARY APPROACH

This section delves into comparing our RPL integration model with state-of-the-art methods. Using an approach that specifically emphasizes separating and identifying unknown attack patterns through innovative learning mechanisms, RPL marks an advancement in cybersecurity efforts. This comparative analysis will highlight the advantages of RPL in navigating the complexity of cyber threats but also set the stage for discussion when comparing it against other state-of-the-art approaches. For convenience, we have compiled a table of comparisons of recent research studies in machine learning and deep learning. A summary of the methodologies employed, the extent of the challenges encountered, and certain constraints of the research are provided in Table 1.

III. PROPOSED METHODOLOGY

Our proposed model The CNN-RPL is designed to detect both known and unknown attack features simultaneously, enhancing the capabilities of existing Intrusion Detection Systems. The model comprises three distinct architectures: the Data Pre-processing Module, the OSR IDS Module, and the Unknown DDoS Detection Module, as illustrated in Fig. 1. By harnessing the capabilities of these three components, our proposed model excels in data pre-processing and identifies known and previously unseen DDoS attacks through OSR techniques. While Fig. 1 may suggest a structural with standard 1D CNN models, the ingenuity of proposed model lies in its subtle components. Central to its innovation is the Open Set Recognition IDS module, which integrates a unique RPL algorithm within the Unknown DDoS Detection Module. This algorithm adeptly identifies and learns from new DDoS patterns, thereby boosting the model's adaptability. Additionally, our classifier surpasses conventional designs with domain-specific optimizations that sharpen its ability to discern between benign and malicious traffic.

Traditional and some recent CNN architectures, while proficient in pattern recognition, are typically constrained by their reliance on known data distributions, often leading to a local optima performance when encountering novel or sophisticated DDoS attacks. The proposed model stems from the need to address the gap in identifying novel attack patterns that conventional models often miss. By integrating RPL, the model adeptly discerns unknown traffic types, enhancing its predictive capabilities. This approach is rooted in a thorough analysis of existing CNN architectures, where the RPL component refines classification boundaries for better generalization to unknown data. These improvements, although not visually distinct in the schematic, significantly elevate the model's detection capabilities beyond state-of-the-art methods.

A. CNN CLASSIFIER

The model comprises three convolutional layers, each followed by a PReLU activation function and interconnected with max-pooling operations as shown in Fig. 2 and detailed as Fig. 3. Convolutional layers are used to extract the features from the input data, and PReLU (Parametric Rectified Linear Unit) is an activation function that introduces learnable parameters to enhance the model's capacity. Max-pooling helps reduce the dimensionality of the data while retaining the most important information. The flattened features are connected to a fully connected layer. The model has two output branches. One output is for the classification of known features, which is typical in Close-Set classification tasks. The other output is designed to process input features associated with Reciprocal Points Learning. These are the two-dimensional embedded features, obtained after being down sampled by the output of the fully connected layer.

B. LOSS FUNCTION FOR OPEN-SET RECOGNITION

Adapting the loss function of a Convolutional Neural Network model is essential when working with OSR. The Softmax loss function is often used in Close-Set multiclass classification problems, where the goal is to assign an input data point to one of several possible classes. It is particularly associated with neural networks and deep learning models for tasks like image classification and natural language processing. In OSR or anomaly detection tasks, using the Softmax loss alone may not be suitable because it does not inherently handle unknown or out-of-distribution data. The defection of Softmax can be expressed by $\sigma(z_i)$:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}, \text{ for } i = 1, 2, \ldots, K \tag{1}$$

where the z_i is the input vector of CNN output, the e^{z_i} is the standard exponential function for input vector, the K is the number of classes in the multi-class classifier, the e^{z_j} is standard exponential function for output vector. To address these limitations and make models suitable for OSR or anomaly detection tasks, modifying the loss function, model architecture, or incorporating additional techniques is common. This can involve techniques like using a ''reject'' class, margin-based losses, calibration methods, or leveraging auto-encoders and uncertainty estimation.

C. RECIPROCAL POINTS LEARNING

Reciprocal Points Learning (RPL) [37], [38] is a technique used to address OSR problems. The RPL calculates the distance between deep features extracted from the model's feature space. This feature space can be the output of a neural network layer or any other embedding space used in your model. For each feature, the RPL calculates

TABLE 1. Comparative analysis of recent advances in DDoS detection models.

Author	Dataset	Problem Scope	Technical	Limitation
Beitollahi et al. (2022) [29]	NSL-KDD	CSR	A RBF system using the CSA technique.	The study's exclusive reliance on the NSL-KDD dataset may restrict the generalizability of the model, particularly in the context of OSR.
Neto et al. (2022) [30]	CICDDoS2019	CSR	A federated learning approach for collaborative, privacy-preserving DDoS detection.	Future work need to focus on model optimization and real-time application improvements.
Najafimehr et al. (2022) [31]	CICIDS2017, CICDDoS 2019	CSR, OSR	DBSCAN labels traffic, while statistical measures define the ML framework for DDoS detection.	The proposed method may be excessively computationally intensive for real-world use.
Zhao et al. (2023) [32]	CICDDoS2019 CICIDS2017, CICIDS2018, KDD_CUP99 NSL-KDD, UNSW	CSR	Genetic algorithms for automatic DNN network generation.	The main emphasis of this research is the development of models for CSR
Sharif et al. (2023) [33]	CICIDS2017	CSR	The study attempts to detect DDoS attacks from various tools using MLP.	The study's focus on the CICIDS2017 dataset may limit its applicability in OSR.
Shieh et al. (2023) [25]	CICIDS2017, CICDDoS2019	CSR, OSR	The DDoS defense model uses OC-SVM with SGD.	The proposed method's complexity may limit computational resources and real-time applicability.
Yonas et al. (2023) [34]	IoTID20 CICIoT2023	CSR	The framework employs an ensemble method, AUWPAE, combining various models to adaptively detect DDoS attack.	Low throughput and high latency indicate a trade-off between accuracy and speed, with little exploration of other attack scenarios or network conditions.
Nguyen et al. (2024) [35]	CIC-IDS-2017, CIC-IDS-2018, BoT-IoT	CSR, OSR	A hybrid model combining SOCNN, LOF, and iNNE detect known and unknown DDoS attacks.	The study notes the need for balancing real-time detection efficiency and adapting to evolving attack strategies.
Lam et al. (2024) [36]	CICIDS2017, CICDDoS 2019	CSR, OSR	CNN-based algorithm incorporating OSR and FCM for unknown attack detection.	Due to its many parameters, the model may be difficult to implement,especially in computationally intensive environments.
Our	CICIDS2017, CICDDoS2019	CSR, OSR	An IDS incorporated CNN with Reciprocal Points Learning makes the model flexible and effective	N/A

the reciprocal (the inverse) between the feature and its corresponding center point. This reciprocal term represents how far the feature is from its associated class center. The loss encourages the model to push deep features far away from their reciprocal points, particularly those reciprocal points associated with out-of-distribution data. This step ensures a clear separation between known and unknown features.

The prototype loss [39], integrated into the RPL framework, is an adjustment of the Softmax loss [40]. Its objective is to encourage the model to acquire prototype representations for each designated class. The learning process involves minimizing the loss associated with classifying reciprocal points by using the negative log-probability of the true class K through Stochastic Gradient Descent (SGD) as:

$$L_c(x;\ \theta, o) = -\log \frac{e^{-d(\theta(x),\, o^i)}}{\sum_{i=1}^{K} e^{-d(\theta(x),\, o^i)}},\ \text{for } i = 1, 2, \ldots, K \tag{2}$$

where $d\left(\theta\left(x\right), o^i\right)$ is the Euclidean distance between $\theta\left(x\right)$ and o^i, the $\theta\left(x\right)$ is the denote as the embedding function which is the output of CNN and o^i is the center of each classes. Minimizing Equation (2) that maximizes the dissimilarity between known data and the set of reciprocal points, facilitating the expansion of the gap between the closed space and the open space. This outcome aligns with our initial objective.

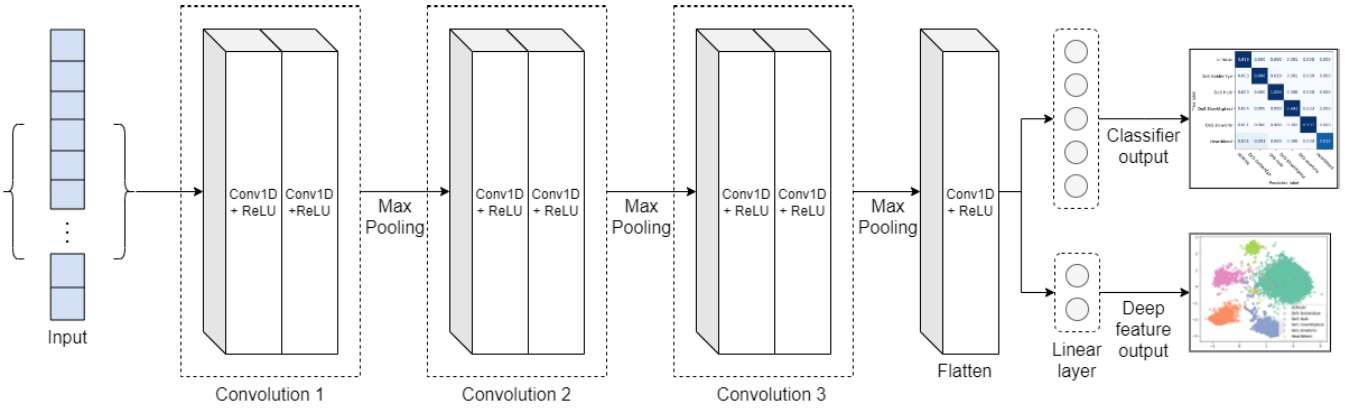


FIGURE 2. The CNN model of propose model.

```
==========================================================================
Layer (type:depth-idx)                   Output Shape              Param #
==========================================================================
CNNModel                                 [512, 6]                  6,424
├─Conv1d: 1-1                            [512, 16, 73]             64
├─PReLU: 1-2                             [512, 16, 73]             1
├─Conv1d: 1-3                            [512, 16, 73]             784
├─PReLU: 1-4                             [512, 16, 73]             1
├─Conv1d: 1-5                            [512, 32, 72]             1,568
├─PReLU: 1-6                             [512, 32, 72]             1
├─Conv1d: 1-7                            [512, 32, 72]             3,104
├─PReLU: 1-8                             [512, 32, 72]             1
├─Conv1d: 1-9                            [512, 32, 71]             3,104
├─PReLU: 1-10                            [512, 32, 71]             1
├─Conv1d: 1-11                           [512, 32, 71]             3,104
├─PReLU: 1-12                            [512, 32, 71]             1
├─Conv1d: 1-13                           [512, 4, 70]              132
├─PReLU: 1-14                            [512, 4, 70]              1
├─Linear: 1-15                           [512, 2]                  562
├─PReLU: 1-16                            [512, 2]                  1
├─Linear: 1-17                           [512, 6]                  18
==========================================================================
Total params: 18,872
Trainable params: 18,872
Non-trainable params: 0
Total mult-adds (M): 434.63
==========================================================================
```

FIGURE 3. Architecture of CNN model.

We acknowledge that trying to set hard boundaries or constraints on the open space itself can be problematic, especially when dealing with scenarios where a significant number of unknown or out-of-distribution samples are present. Rather than directly bounding the open space, we propose to manage open space risk indirectly. This is achieved by controlling the distance between the samples in the known class space and their respective reciprocal points as $L_o(x;\ \theta, p)$:

$$L_o(x;\ \theta, p) = \frac{1}{M}\sum_{j=1}^{M} d\left(\theta\left(x\right) - p_j^i\right),\ \text{for } i = 1, 2, \ldots, K \tag{3}$$

where $d\left(\theta\left(x\right) - p_j^i\right)$ is the Euclidean distance between $\theta\left(x\right)$ and p^i, the $\theta\left(x\right)$ is the denote as the embedding function which is the output of CNN and p^i is the reciprocal of each classes, M represents the number of reciprocals for each

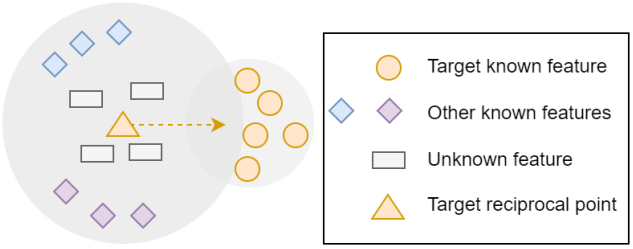


FIGURE 4. A closer look at embedded feature shifting in RPL.

classes. In essence, as Fig. 4 shows, this approach encourages the model to have known class samples clustered around their respective class's reciprocal point while keeping them away from the more uncertain and potentially risky regions of the open space.

The overall loss function of Reciprocal Points Learning combines two key components: the Classification Loss (L_c) and the reduction of Open Space Loss (L_o). The classification loss measures the error associated with classifying known data accurately. The open space loss focuses on managing open space risk by encouraging a separation between known classes and the open space, which may contain out-of-distribution or unknown samples.

$$L(x;\ \theta, o, p) = L_c(x;\ \theta, o) + \lambda L_o(x;\ \theta, p) \tag{4}$$

where the λ is the Hyperparameter to control the constrain scale. With the Open-Set capable Loss function, the training data set can be used to train the model. This method improves the model's ability to distinguish between known and unknown data, making it more robust for OSR tasks.

D. UNKNOWN DDoS ATTACKS DETECTION METHOD

To identify unknown DDoS attacks, we have devised a method for computing the probability of the target feature. This technique employs the Exponential function to determine the probability, relying on the RPL's Euclidean distance between the target feature and the center, denoted as $P(x)$:

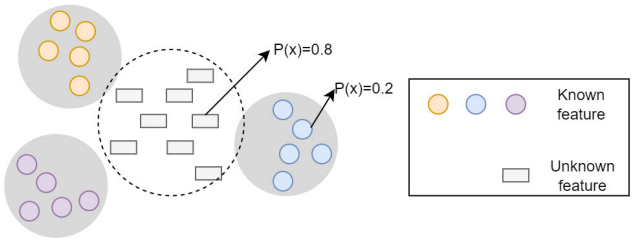


FIGURE 5. Unknown detection diagram.

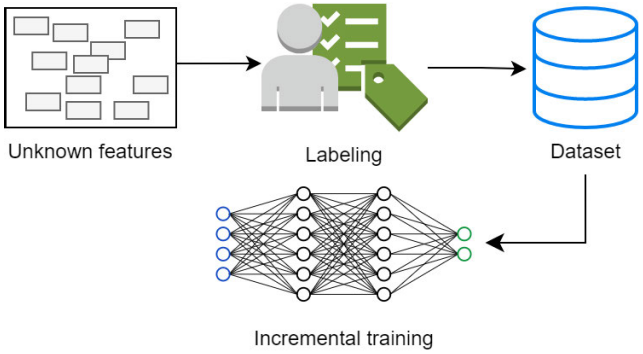


FIGURE 6. Unknown feature labeling and incremental training.

To identify unknown DDoS attacks, we have devised a method for computing the probability of the target feature. This technique employs the Exponential function to determine the probability, relying on the RPL's Euclidean distance between the target feature and the center, denoted as $P(x)$:

$$P(x) = e^{-\lambda d(\theta(x^i)-o^i)}, \text{ for } i = 1, 2, \ldots, K \tag{5}$$

where $d(\theta(x^i) - o^i)$ is the is the Euclidean distance between $\theta\,(x)$ deep feature and o^i center, λ is the constrained term of probability, it adjusts the falling trend of Exponential. Consequently, this function is adaptable to various distributions.

As shown in Fig. 5, in this study, Equation (5) is employed to compute the $P(x)$ for each embedded feature. This dynamic adjustment helps in defining the boundaries between known and unknown classes.

E. INCREMENTAL LEARNING FOR UNKNOWN DDoS ATTACKS DETECTION

Features labeled as unknown will be annotated by the human engineer as either attack or benign class and reintroduced into the training dataset for incremental training as shown in Fig. 6. The retraining model is capable of identifying labeled unknown attacks after training and enhances the model's recognition capabilities.

IV. EXPERIMENT

This experiment was conducted using the Ubuntu 22.04 LTS operating system, equipped with an 11th Gen Intel i7-11700 processor running at 4.800GHz, 96GB of DDR4 system memory, and an NVIDIA GeForce RTX 3090 as the machine learning accelerator. The setup included the NVIDIA Driver version 535.113.01 and CUDA Version 12.1, managed through the Anaconda development environment. Python version 3.8.17 was used, and the model framework employed libraries such as PyTorch 2.0.1, numpy 1.24.3, scikit-learn 1.2.0, among others.

A. DATASET

The Canadian Institute for Cybersecurity (CIC), an institution dedicated to cutting-edge research, training, and innovation in cybersecurity. The institute engages in research covering various cybersecurity domains, including intrusion detection, information security management, data privacy, encryption techniques, vulnerability analysis, and defense strategies. The network flow datasets produced by them are also widely used in the field of cybersecurity research. Therefore, this study utilizes two datasets created by this organization, namely CICIDS 2017 [41] and CICDDoS 2019 [42], as the training and testing datasets for the model:

- CICIDS 2017 dataset [41], a labeled collection distinguishing benign from attack traffic, captured through a mirror port and covering all common protocols. This dataset includes more than 80 network flow features extracted and presented in a CSV file.
- CICDDoS 2019 dataset [42], which addresses the shortcomings of existing DDoS attack datasets by including a comprehensive set of 11 representative DDoS attacks. The dataset is completely labeled with 80 network traffic features, and the authors provide the most important feature sets to detect different types of DDoS attacks. Additionally, a new taxonomy for DDoS attacks is proposed [43], [44], categorizing attacks based on their characteristics and including a new category for ''hybrid attacks''.

Table 2 shows the utilization of the CICIDS 2017 and CICDDoS 2019 datasets in this study, including analysis of attack categories, quantities, and whether they serve as training data.

1) CLOSE-SET LABELS

During the Close-Set training phase, as detailed in Table 3, two primary categories are identified: Benign and Malicious. The Benign category encompasses all benign network traffic and is denoted solely as BENIGN. In contrast, the Malicious category encompasses all forms of known malicious traffic, with distinct labels assigned to each specific type of malicious traffic. Consequently, when subjected to testing, the model's challenges are multiple. It must discern between benign and malicious traffic and exhibit the capability to discriminate among the various types of malicious traffic, effectively identifying and categorizing them based on their specific characteristics.

TABLE 2. Dataset usage analysis.

Dataset	Label name	Quantity	Training
CICIDS 2017 Wednesday	Benign	319178	Yes
	DoS GoldenEye	159049	Yes
	DoS Hulk	7647	Yes
	DoS Slowhttptest	5707	Yes
	DoS slowloris	5109	Yes
	Heartbleed	11	Yes
CICIDS 2017 Friday	DDoS	95114	No
CICDDoS 2019	LDAP	2179928	No
	MSSQL	4522489	No
	DNS	5071002	No
	NetBIOS	4093273	No
	NTP	1202639	No
	UDP	3134643	No
	SNMP	5159863	No
	SSDP	2610610	No
	SYN	1380015	No

TABLE 3. Labels of close-set classification.

Benign	Malicious
BENIGN	DoS slowloris, DoS Slowhttptest, DoS Hulk, DoS GoldenEye, Heartbleed, DDoS LOIT, DrDoS LDAP, DrDoS MSSQL, DrDoS DNS, DrDoS NetBIOS, DrDoS NTP, DrDoS UDP, DrDoS SNMP, DrDoS SSDP, SYN data

2) OPEN-SET LABELS

In the OSR phase, as delineated in Table 4, the dataset is divided into two primary categories: Known and Unknown. The Known category comprises traffic originating from the CICIDS 2017 Wednesday dataset, which was previously utilized in the Close-Set training phase. On the other hand, the Unknown category encompasses data that was not incorporated into the training dataset. This includes data from CICIDS 2017 Friday and DDoS attack data sourced from CICDDoS 2019.

TABLE 4. Labels of open-set classification.

Known	Unknown
BENIGH, DoS slowloris, DoS Slowhttptest, DoS Hulk, DoS GoldenEye, Heartbleed	DDoS LOIT, DrDoS LDAP, DrDoS MSSQL, DrDoS DNS, DrDoS NetBIOS, DrDoS NTP, DrDoS UDP, DrDoS SNMP, DrDoS SSDP, SYN data

B. TRAINING PARAMETERS

The training parameters, as outlined in Table 5, provide a comprehensive overview of the training process. Throughout training, 100 epochs are executed, with an initial learning rate of 0.003. The MultiStepLR method adjusts the learning rate as training progresses dynamically. A batch size of 512 is chosen, and the optimization algorithm used is Adam. To enhance the robustness of the model, training is systematically conducted with a fixed set of 10 random seeds, effectively reducing the variability associated with seed values. This approach ensures the model's performance is not overly dependent on specific random initializations. Furthermore, the training-to-testing data ratio is established at 8:2, allowing for a substantial amount of data to be used for training while reserving a significant portion for testing and evaluation. This balance helps assess the model's generalization and performance on unseen data.

TABLE 5. Training parameters.

Parameter	Value
Training epoch	100
Learning rate*	0.003
Batch size	512
Optimizer	Adam
Fixed random seeds	0, 42, 123, 222, 419, 844, 918, 1344, 65536, 815149
Dataset split ratio	Training 80%, Testing 20%

*The learning rate is adjusted using the MultiStepLR function, which decreases the learning rate of each parameter group by a factor of gamma when the number of epochs hits one of the specified milestones.

C. VALIDATION METRICS

The confusion matrix evaluates our proposed model and others' performance by comparing predicted and actual classes. It is commonly used in supervised learning tasks in machine learning and data mining. The confusion matrix is used to calculate various metrics such as accuracy, precision, recall, and F1 score, which are important evaluation metrics in machine learning.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN}, \tag{8}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

where True Positives (TP) represents the number of positive cases that are correctly identified as positive, while False Positives (FP) represents the number of negative cases that are incorrectly identified as positive. True Negatives (TN) represents the number of negative cases that are correctly identified as negative, while False Negatives (FN) represents the number of positive cases that are incorrectly identified as negative. Those metrics gave a better understanding of how

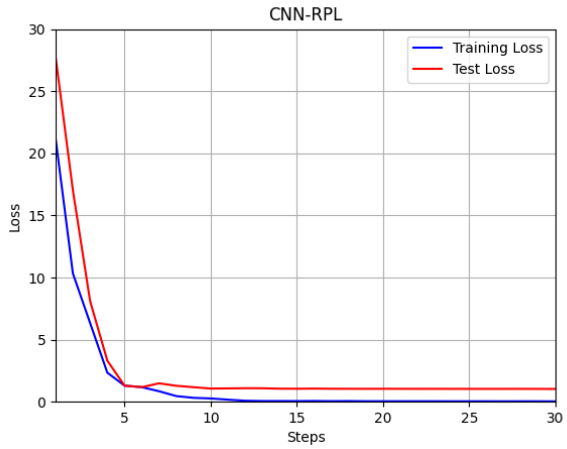


FIGURE 7. Training steps plot of the CNN-RPL model.

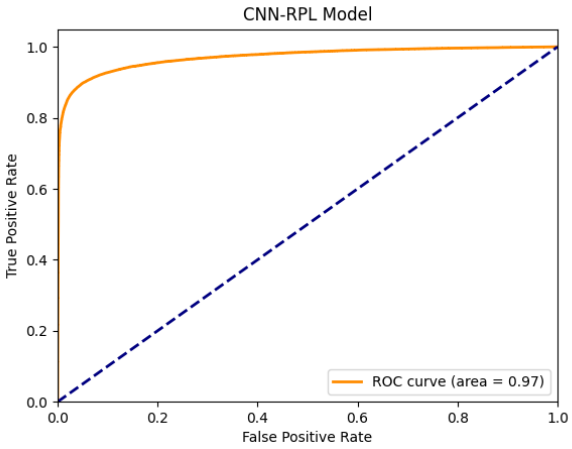


FIGURE 8. The ROC curve of the CNN-RPL model.

well the model performs and identified areas where it may need improvement.

D. THE RESULT OF KNOWN DDoS ATTACK DETECTION

The first evaluation involves Close-Set DDoS attack detection, serving as a fundamental evaluation of the model's proficiency in recognizing known attacks. The model tries to classify attacks into two distinct groups: benign and malicious. Within the malicious category, it further challenges the model to differentiate and categorize the various known attack types, showcasing its capacity to provide detailed attack classification.

The training dataset chosen is CICIDS 2017 Wednesday, split into an 80:20 ratio for training and evaluation. Fig. 7 is the training steps plot of the model, showing the change in the loss metrics through the training process.

Fig. 8 is the ROC curve showing the performance of the CNN-RPL model at different classification thresholds.

Besides the CNN-RPL model, we have employed and evaluated three other ML or DL methods, including XGBoost, KNN, and SVM, for basic comparison.

TABLE 6. The result of known DDoS attacks.

Dataset	Method	Accuracy	Precision	Recall	F1
CICIDS 2017 Wednesday	XGBoost	0.9996	0.9996	0.9996	0.9996
	KNN	0.9931	0.9931	0.9931	0.9931
	SVM	0.9897	0.9897	0.9897	0.9897
	CNN-RPL	0.9993	0.9993	0.9993	0.9993

The results in Table 6 show the performance of four different models in this experiment. The baseline comparison highlights the performance of CNN-PRL compared to conventional methods. The CNN-RPL method stands out with an impressive average score of 0.9993. These results emphasize the effectiveness of the CNN-RPL method in recognizing known attack patterns. It demonstrates the model can efficiently distinguish between benign and malicious traffic, maintaining clear differentiation without ambiguity or confusion.

E. THE LIMITATION OF CLOSE-SET CLASSIFICATION

As Table 6 shows, the CNN-RPL model performs well at Close-Set DDoS attack detection. However, real-world DDoS attacks are often very diverse, with different attack scenarios and techniques depending on the hackers. With new attack methods, traditional ML or DL systems usually face difficulty detecting and defending effectively. The systems are constrained by their inability to recognize, classify, or respond to attack patterns that they haven't encountered during the training process. Table 7 provides insights into the performance of CNN-RPL on unknown DDoS attack detection when OSR techniques are omitted.

TABLE 7. The result of unknown DDoS attacks without OSR of CNN-RPL.

Dataset	Accuracy	Precision	Recall	F1
CICIDS 2017 Friday	0.1510	0.3010	0.2315	0.2617
CICDDoS 2019 MSSQL	0.0989	0.0395	0.0989	0.0432
CICDDoS 2019 LDAP	0.1854	0.0814	0.1854	0.0932
CICDDoS 2019 DNS	0.0892	0.0352	0.0892	0.0382
CICDDoS 2019 NetBIOS	0.1081	0.0436	0.1081	0.0480
CICDDoS 2019 NTP	0.2921	0.1434	0.2921	0.1693
CICDDoS 2019 UDP	0.1367	0.0569	0.1367	0.0637
CICDDoS 2019 SNMP	0.0878	0.0346	0.0878	0.0375
CICDDoS 2019 SSDP	0.1597	0.0682	0.1597	0.0772
CICDDoS 2019 SYN	0.2645	0.1071	0.2645	0.1381

In this experiment, we utilized two datasets: CICIDS 2017 Friday and CICDDoS 2019, including various unknown DDoS attack patterns. Table 7 results show performance with consistently low evaluation metrics. This demonstrates that these attacks are new to the trained model. Without the OSR module, the system mostly misclassifies attacks

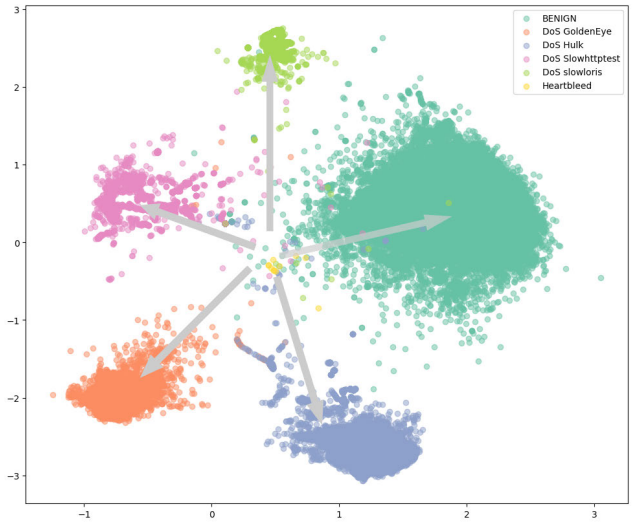


FIGURE 9. The embedded features distribution.

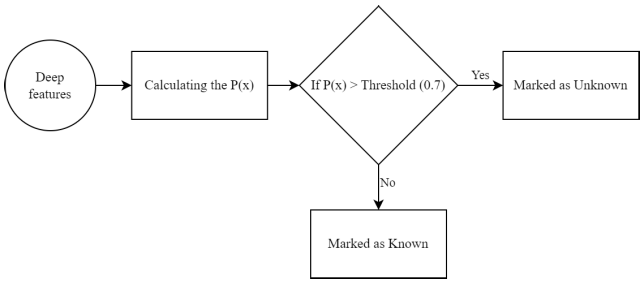


FIGURE 10. Unknown detection diagram.

as regular access. This result emphasizes that incorporating OSR techniques to handle and organize unknown attacks is necessary.

In this research, the RPL technique is employed as the primary one to address the challenge of unknown DDoS attacks. The RPL calculates the loss function for Reciprocal Points of each known class, as denoted in Equation (4), which instigates a shift in the distribution of known features, illustrated in Fig. 9. This transformation involves strategically displacing each known feature outward in alignment with the directional arrows. This process effectively carves out a dedicated space in the center, ready to accommodate the unknown features, making it a robust method for handling unknown DDoS attacks.

F. THE RESULT OF UNKNOWN DDoS ATTACKS DETECTION

The evaluation procedure is shown in Fig. 10. Firstly, the embedding features produced by the CNN layers and processed by the RPL algorithm are input into the unknown detection module. Equation (5) calculates the probability that an attack is unknown or not. Unknown attacks are those in which the probability exceeds the set threshold. On the contrary, known attacks are those with a probability below the set threshold.

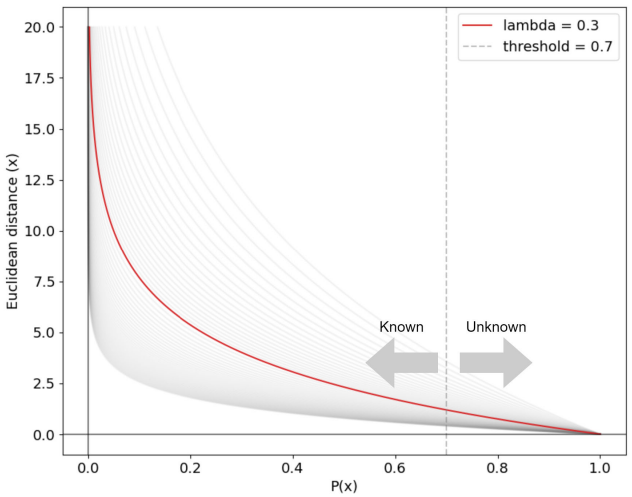


FIGURE 11. Unknown detection visualization.

1) EXPERIMENTAL PARAMETERS

In Equation (5) of the unknown detection module, there are two key parameters to consider: *lambda* and *threshold*. These parameters play a pivotal role in shaping the behavior of the algorithm. The *lambda* influences the curve of the algorithm's descent, while *threshold* sets the standard for distinguishing between known and unknown elements. The *lambda* is configured at 0.3, and the *threshold* is 0.7 in this specific study. These parameter values have been selected to optimize the model's ability to identify unknown elements. It is important to note that these values can be adjusted to accommodate different distribution scenarios, tailoring the model's performance for optimal unknown recognition. For a visual understanding, Fig. 11 provides a clear description of the descent curves under various *lambda* values and the placement of the *threshold*. In this context, the red line corresponds to *lambda* as 0.3, while the dashed line represents a *threshold* value of 0.7.

2) RECIPROCAL POINTS LEARNING ON UNKNOWN DDoS ATTACKS DETECTION

Reciprocal Points Learning is the foundational technique in detecting unknown DDoS attacks, characterized by unknown features that often yield similar confidence scores across each known class. In simpler terms, the likelihood of classifying an unknown feature into any of the known classes is roughly equal. This phenomenon is clearly shown in Fig. 12, which depicts the distribution of known and unknown samples in the two-dimensional feature space. Unknown features tend to congregate towards the center, while known features tend to cluster closer to the outer edges. This visual representation demonstrates an apparent clustering between the known and unknown classes in the feature space.

This unique characteristic emphasizes the practicality and effectiveness of the RPL method. Utilizing the distance of each feature from the central, as illustrated in Fig. 13,

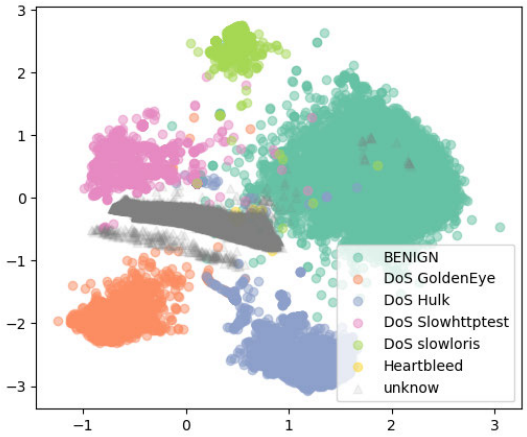


FIGURE 12. The embedded features distribution with unknown features.

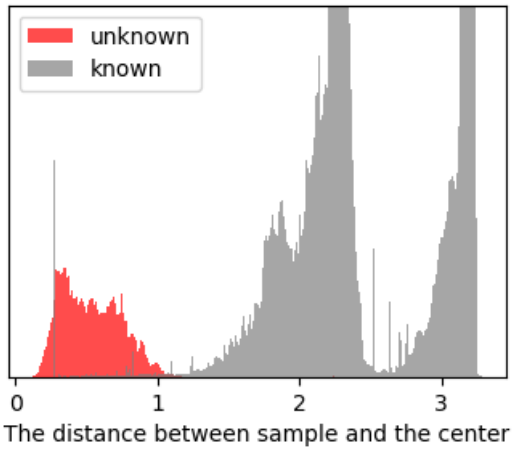


FIGURE 13. The visualization of embedded features distance distribution.

visualizes the distribution of known and unknown features along a number line. Notably, there is minimal overlap between known and unknown features, indicating the model's ability to segregate them distinctly. Moreover, employing the Equation (5) algorithm constructs the model's confusion matrix specifically for unknown features, as shown in Fig. 14. This visualization helps us to estimate the model's effectiveness in recognizing unknown DDoS attacks with precision and accuracy.

After integrating the OSR module, Table 8 showcases CNN-RPL's detection performance concerning various unknown DDoS attacks, encompassing datasets such as CICIDS 2017 Friday and CICDDoS 2019. The results emphasize the model's ability to consistently attain an average detection rate of 98% across ten different unknown DDoS attacks. This noteworthy achievement underscores CNN-RPL's effectiveness in accurately identifying and differentiating unknown DDoS attacks.

G. INCREMENTAL LEARNING

By utilizing the CNN-RPL method to detect unknown attacks, we re-label the attack data and incorporate it into the

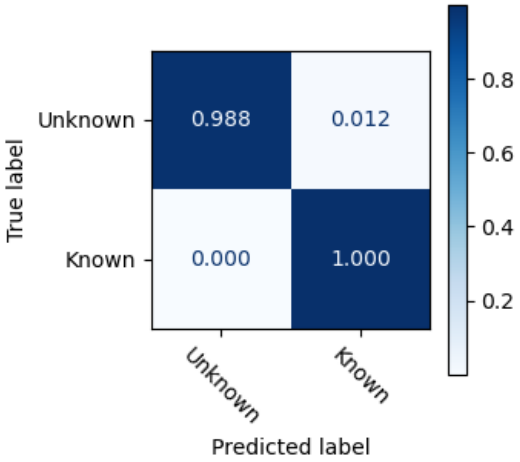


FIGURE 14. Confusion matrix when the threshold set for the distance distribution.

TABLE 8. The result of unknown DDoS attacks with OSR.

Dataset	Accuracy	Precision	Recall	F1
CICIDS 2017 Friday	0.9901	0.9907	0.9901	0.9903
CICDDoS 2019 MSSQL	0.9879	0.9889	0.9879	0.9881
CICDDoS 2019 LDAP	0.9971	0.9971	0.9971	0.9971
CICDDoS 2019 DNS	0.9687	0.9762	0.9687	0.9707
CICDDoS 2019 NetBIOS	0.9941	0.9941	0.9941	0.9941
CICDDoS 2019 NTP	0.9840	0.9842	0.9840	0.9839
CICDDoS 2019 UDP	0.9937	0.9937	0.9937	0.9936
CICDDoS 2019 SNMP	0.9631	0.9732	0.9631	0.9658
CICDDoS 2019 SSDP	0.9926	0.9927	0.9926	0.9925
CICDDoS 2019 SYN	0.9802	0.9802	0.9802	0.9802

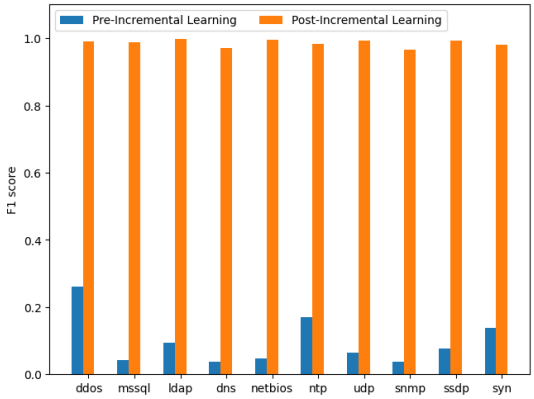


FIGURE 15. Before and after incremental learning comparison.

training dataset for another round of learning. After the re-labeled data, CNN-RPL can effectively identify the specific attack, enabling a robust defense against known attacks.

Fig. 15 compares F1 scores before and after incremental learning for unknown DDoS attacks. After re-labeling and

TABLE 9. The comparison with other methods on CICIDS2017.

Dataset	Method	Accuracy	Precision	Recall	F1
CICIDS 2017 Wednesday	DNN-GA [32]	0.9906	0.9896	0.9915	0.9905
	CNN-Geo [28]	0.9979	0.9962	0.9944	0.9953
	Alexnet-FCM [36]	0.9976	0.9944	0.9991	0.9967
	CNN-RPL	0.9993	0.9993	0.9993	0.9993

training, the F1 scores have recovered to satisfactory ranks for each type of unknown attack. This confirms the capability of CNN-RPL to identify attack traffic after incremental training.

H. COMPARISON WITH OTHERS RESULTS

In this section, we will compare our method with recent studies. First, we will evaluate the effectiveness of algorithms in detecting known attacks. The recent algorithms selected include Deep Neural Network with Genetic Algorithms (2023) [32], Convolutional Neural Network featuring Geometrical Metric (2023) [28], and AlexNet with Fuzzy C-Means clustering (2024) [36]. The effectiveness of these algorithms is assessed on the same traditional dataset, CICIDS 2017 Wednesday. Table 9 shows the comparison with these algorithms.

To better evaluate CNN-RPL, this research also compares experimental results with three recent methods designed to detect unknown DDoS attacks: DHR-Reconstruct error-OCSVM [25], AlexNet-FCM [36], and CNN-Geo [28]. These three models tailored for unknown DDoS attack recognition have exhibited notable performance. DHR-Reconstruct error-OCSVM employs a combination of reconstruction error and One-Class SVM algorithms to distinguish unknown attacks effectively. AlexNet-FCM leverages the powerful AlexNet architecture commonly used in image recognition tasks along with the Fuzzy C-Means clustering method. On the other hand, CNN-Geo relies on the Geometrical metric method to excel in discerning unknown attacks. It is crucial to note that the comparison employs identical unknown datasets, CICIDS 2017 Friday and CICDDoS 2019, across all methodologies. Those make a remarkable achievement in unknown DDoS detection domains.

The experiments entailed a comparison of five key indicators: Parameter, Avg. Accuracy, Avg. Precision, Avg. Recall, and Avg. F1:

- Parameter signifies the count of parameters present in the model. The learning process revolves around fine-tuning these parameters to enhance the network's proficiency in delivering precise predictions or classifications for a given task. The number of model parameters also directly affects model size, memory usage, and training time. Achieving and maintaining accuracy while reducing the number of parameters can optimize the computing resources.
- Avg. Accuracy, Avg. Precision, Avg. Recall, and Avg. F1 pertain to the averaging of results obtained through unknown DDoS detection on dataset containing elements from CICIDS 2017 Friday and CICDDoS 2019. Averaging these metrics simplifies the comparison process, offering a more intuitive means of assessing and contrasting the performance of various models.

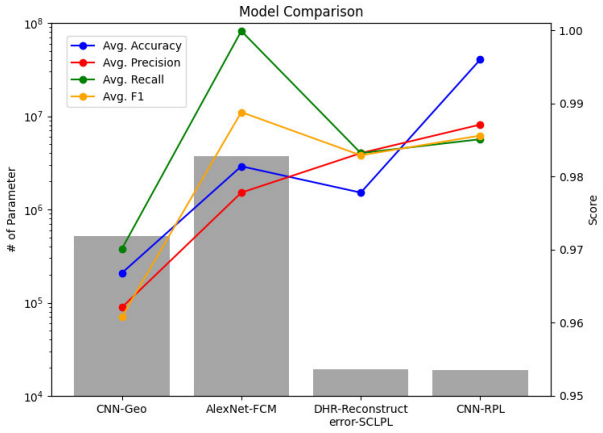


FIGURE 16. Charts of comparison with others.

TABLE 10. Comparison with others methods.

Model	Parameter	Avg. Accuracy	Avg. Precision	Avg. Recall	Avg. F1
CNN-Geo[28]	521834	0.9668	0.9621	0.9701	0.9608
AlexNet-FCM[36]	3745138	0.9814	0.9778	0.9999	0.9888
DHR-Reconstruct error-OCSVM[25]	19159	0.9836	0.9832	0.9832	0.9829
CNN-RPL	18872	0.9851	0.9871	0.9851	0.9856

The test results are presented in Table 10, and a precise observation emerges. Despite employing the fewest parameters, CNN-RPL has attained highly commendable outcomes in critical metrics such as Avg. Accuracy, Avg. Precision, Avg. Recall, and Avg. F1. which substantiates that CNN-RPL has effectively maintained a high unknown DDoS attack defense performance while conserving computational resources.

Fig. 16 shows a comparative chart among the four models, and from the comparison of the Parameter, it is evident that each of the four models has its strengths and weaknesses in different indicators:

- DHR-Reconstruct error-OCSVM: This model uses a similar number of parameters but has slightly lower accuracy compared to CNN-RPL. The DHR-Reconstruct error-OCSVM method employs a two-layer filtering process, Reconstruct error and One Class SVM, which reduces the issue of misclassifying unknown attacks but also leads to decreased accuracy simultaneously.

- AlexNet-FCM: This model uses a large number of parameters, resulting in the highest accuracy among the four models. Regardless, it also exhibits the highest variability in accuracy. Using a large model can increase accuracy, but it can also introduce issues like vanishing gradient and overfitting, reducing the model's reliability.
- CNN-Geo [28]: Although CNN-Geo has the lowest accuracy among the four models and uses a considerable number of parameters, it still achieves an accuracy of over 95%. Additionally, in their research, this model also possesses the ability to detect GAN attacks, a feature that the other three models lack.

The results in Table 10 and Fig. 16 indicate that, compared to models in the same research domain, CNN-RPL offers several advantages, including a more lightweight model design, consistent performance in detecting unknown attacks, and a straightforward model architecture. Therefore, CNN-RPL stands out as an outstanding model for detecting unknown DDoS attacks.

V. CONCLUSION AND FUTURE WORK

This comprehensive and in-depth study explores methods to conduct unknown DDoS attacks, addressing the detection challenges associated with such attacks. The proposed CNN-RPL model, based on Open-Set recognition, is introduced to tackle the problem of detecting unknown DDoS attacks. The study emphasizes that the model performs well in defending against known DDoS attacks and exhibits superior performance in handling unknown ones.

CNN-RPL achieves defense rates of over 98% for known and unknown attacks in CICIDS 2017 and CICDDoS 2019, which is the model's validation data for DDoS attacks. Compared with existing models, CNN-RPL reduces the number of layers while maintaining the defense rate, a negligible decline of 0.324% with the best model. Additionally, the CNN-RPL significantly reduces the training parameters, accelerating the training and deployment costs and enhancing the practical feasibility of real-world applications.

As the network environment continually evolves and attack techniques constantly develop, the CNN-RPL model, which is an advanced tool for detecting unknown DDoS attacks, faces the ongoing challenge of enhancing its performance and adaptability. With attacks like HTTP/2 attacks, hackers continuously exploit vulnerabilities in the network to launch attacks, and DDoS attacks are growing at an incredible speed. Therefore, the possible directions for future improvements are as follows:

- Expanding the Diversity of the Dataset: It is essential to represent different aspects of the studied problem or scenario comprehensively and extensively. This includes methods such as introducing data from different categories, considering various features and attributes, accounting for duration and spatial variations, considering extreme scenarios, and balancing dataset distribution. By expanding the diversity of the dataset, the model's generalization ability can be enhanced, enabling it to better adapt to various scenarios and challenges.
- Deepening Open-Set Recognition Techniques: Future research will concentrate on further optimizing Open-Set recognition methods. By introducing higher-level feature extraction and more intelligent model selection, we seek to improve CNN-RPL's detection accuracy and sensitivity to unknown attacks.
- Feasibility for Practical Deployment: We will further study the practical deployment feasibility of the model, considering computational resource requirements, operational efficiency, and real-world applicability. This ensures that the model can operate efficiently in real-world environments.
- Automate Data Labeling: This process is crucial in developing machine learning models, as it helps to train the model and determine its accuracy. Assigning labels to datasets manually can be time-consuming and tedious, especially for large datasets. Automated data labeling can significantly reduce the time and effort required for manual data labeling, resulting in increased efficiency and speed. Gebrye et al. [45] proposes an intelligent raw network data extractor and labeler tool by incorporating the limitations of the tools that are available to transform PCAP to CSV. The authors employed several data preprocessing operations on the selected network intrusion dataset to generate and process a high-quality DDoS attack dataset suitable for machine learning models.

Achieving these goals will contribute to maintaining CNN-RPL's leading in unknown DDoS attack detection and better addressing the evolving challenges in network security.

ACKNOWLEDGMENT

The authors would like to thank the use of ChatGPT 3.5 (https://chat.openai.com/) to provide content writing and grammar improvement in the drafting of this article. The following prompts were entered into ChatGPT and the results have been used directly in their research:

- Improve the writing of the content.
- Check the grammar of the content.

REFERENCES

[1] O. Yoachimik and J. Pacheco. (2023). *DDoS Threat Report for 2023 Q3*. [Online]. Available: https://blog.cloudflare.com/ddos-threat-report-2023-q3/
[2] MSRC. (2023). *Microsoft Response To Distributed Denial of Service (DDoS) Attacks Against HTTP/2*. [Online]. Available: https://msrc.microsoft.com/blog/2023/10/microsoft-response-to-distributed-denial-of-service-ddos-attacks-against-http/2/
[3] G. Bourzikas. (2023). *HTTP/2 Zero-Day Vulnerability Results in Record-breaking DDoS Attacks*. [Online]. Available: https://blog.cloudflare.com/zero-day-rapid-reset-http2-record-breaking-ddos-attack/
[4] G. Cloud. (2023). *How It Works: The Novel HTTP/2 'Rapid Reset' DDoS Attack*. [Online]. Available: https://cloud.google.com/blog/products/identity-security/how-it-works-the-novel-http2-rapid-reset-ddos-attack
[5] CVE. (2023). *CVE-2023-44487*. [Online]. Available: https://www.cve.org/CVERecord?id=CVE-2023-44487

[6] N. Ahuja, G. Singal, D. Mukhopadhyay, and N. Kumar, "Automated DDOS attack detection in software defined networking," *J. Netw. Comput. Appl.*, vol. 187, Aug. 2021, Art. no. 103108.

[7] M. J. Awan, U. Farooq, H. M. A. Babar, A. Yasin, H. Nobanee, M. Hussain, O. Hakeem, and A. M. Zain, "Real-time DDoS attack detection system using big data approach," *Sustainability*, vol. 13, no. 19, p. 10743, Sep. 2021.

[8] R. Doriguzzi-Corin, S. Millar, S. Scott-Hayward, J. Martínez-del-Rincón, and D. Siracusa, "LUCID: A practical, lightweight deep learning solution for DDoS attack detection," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 2, pp. 876–889, Jun. 2020.

[9] A. Bansal and S. Kaur, "Extreme gradient boosting based tuning for classification in intrusion detection systems," in *Proc. Int. Conf. Adv. Comput. Data Sci.*, Oct. 2018, pp. 372–380.

[10] J. Kim, Y. Shin, and E. Choi, "An intrusion detection model based on a convolutional neural network," *J. Multimedia Inf. Syst.*, vol. 6, no. 4, pp. 165–172, Dec. 2019.

[11] J. Kim, J. Kim, H. Kim, M. Shim, and E. Choi, "CNN-based network intrusion detection against denial-of-service attacks," *Electronics*, vol. 9, no. 6, p. 916, Jun. 2020.

[12] C. Elkan, "Results of the KDD'99 classifier learning," *ACM SIGKDD Explorations Newslett.*, vol. 1, no. 2, pp. 63–64, Jan. 2000.

[13] L. Liu, G. Engelen, T. Lynar, D. Essam, and W. Joosen, "Error prevalence in NIDS datasets: A case study on CIC-IDS-2017 and CSE-CIC-IDS-2018," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Oct. 2022, pp. 254–262.

[14] J. Hu, C. Liu, and Y. Cui, "An improved CNN approach for network intrusion detection system," *Int. J. Netw. Secur.*, vol. 23, no. 4, pp. 569–575, 2021.

[15] L. Dhanabal and S. P. Shantharajah, "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 6, pp. 446–452, 2015.

[16] J. Yang, K. Zhou, Y. Li, and Z. Liu, "Generalized out-of-distribution detection: A survey," 2021, *arXiv:2110.11334*.

[17] W. J. Scheirer, L. P. Jain, and T. E. Boult, "Probability models for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2317–2324, Nov. 2014.

[18] P. Perera and V. M. Patel, "Deep transfer learning for multiple class novelty detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11536–11544.

[19] Z. Ge, S. Demyanov, Z. Chen, and R. Garnavi, "Generative OpenMax for multi-class open set classification," 2017, *arXiv:1707.07418*.

[20] R. Yoshihashi, W. Shao, R. Kawakami, S. You, M. Iida, and T. Naemura, "Classification-reconstruction learning for open-set recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4011–4020.

[21] A. Cao, Y. Luo, and D. Klabjan, "Open-set recognition with Gaussian mixture variational autoencoders," in *Proc. Int. Joint Conf. Artif. Intell.*, vol. 35, no. 8, May 2021, pp. 6877–6884.

[22] H. Zhang and V. M. Patel, "Sparse representation-based open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1690–1696Aug. 2017.

[23] P. Oza and V. M. Patel, "C2AE: Class conditioned auto-encoder for open-set recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2302–2311.

[24] C.-S. Shieh, W.-W. Lin, T.-T. Nguyen, C.-H. Chen, M.-F. Horng, and D. Miu, "Detection of unknown DDoS attacks with deep learning and Gaussian mixture model," *Appl. Sci.*, vol. 11, no. 11, p. 5213, Jun. 2021.

[25] C.-S. Shieh, T.-T. Nguyen, C.-Y. Chen, and M.-F. Horng, "Detection of unknown DDoS attack using reconstruct error and one-class SVM featuring stochastic gradient descent," *Mathematics*, vol. 11, no. 1, p. 108, Dec. 2022.

[26] C.-S. Shieh, T.-T. Nguyen, W.-W. Lin, Y.-L. Huang, M.-F. Horng, T.-F. Lee, and D. Miu, "Detection of adversarial DDoS attacks using generative adversarial networks with dual discriminators," *Symmetry*, vol. 14, no. 1, p. 66, Jan. 2022.

[27] C.-S. Shieh, T.-T. Nguyen, W.-W. Lin, W. K. Lai, M.-F. Horng, and D. Miu, "Detection of adversarial DDoS attacks using symmetric defense generative adversarial networks," *Electronics*, vol. 11, no. 13, p. 1977, Jun. 2022.

[28] C.-S. Shieh, T.-T. Nguyen, and M.-F. Horng, "Detection of unknown DDoS attack using convolutional neural networks featuring geometrical metric," *Mathematics*, vol. 11, no. 9, p. 2145, May 2023.

[29] H. Beitollahi, D. M. Sharif, and M. Fazeli, "Application layer DDoS attack detection using cuckoo search algorithm-trained radial basis function," *IEEE Access*, vol. 10, pp. 63844–63854, 2022.

[30] E. C. P. Neto, S. Dadkhah, and A. A. Ghorbani, "Collaborative DDoS detection in distributed multi-tenant IoT using federated learning," in *Proc. 19th Annu. Int. Conf. Privacy, Secur. Trust (PST)*, Aug. 2022, pp. 1–10.

[31] M. Najafimehr, S. Zarifzadeh, and S. Mostafavi, "A hybrid machine learning approach for detecting unprecedented DDoS attacks," *J. Supercomput.*, vol. 78, no. 6, pp. 8106–8136, Apr. 2022.

[32] J. Zhao, M. Xu, Y. Chen, and G. Xu, "A DNN architecture generation method for DDoS detection via genetic alogrithm," *Future Internet*, vol. 15, no. 4, p. 122, Mar. 2023.

[33] D. Mohammed Sharif, H. Beitollahi, and M. Fazeli, "Detection of application-layer DDoS attacks produced by various freely accessible toolkits using machine learning," *IEEE Access*, vol. 11, pp. 51810–51819, 2023.

[34] Y. K. Beshah, S. L. Abebe, and H. M. Melaku, "Drift adaptive online DDoS attack detection framework for IoT system," *Electronics*, vol. 13, no. 6, p. 1004, Mar. 2024.

[35] X.-H. Nguyen and K.-H. Le, "Robust detection of unknown DoS/DDoS attacks in IoT networks using a hybrid learning model," *Internet Things*, vol. 23, Oct. 2023, Art. no. 100851.

[36] T.-L. Nguyen, H. Kao, T.-T. Nguyen, M.-F. Horng, and C.-S. Shieh, "Unknown DDoS attack detection with fuzzy C-means clustering and spatial location constraint prototype loss," *Comput., Mater. Continua*, vol. 78, no. 2, pp. 2181–2205, 2024.

[37] G. Chen, L. Qiao, Y. Shi, P. Peng, J. Li, T. Huang, S. Pu, and Y. Tian, "Learning open set network with discriminative reciprocal points," in *Computer Vision–ECCV 2020*. Cham, Switzerland: Springer, 2020, pp. 507–522.

[38] G. Chen, P. Peng, X. Wang, and Y. Tian, "Adversarial reciprocal points learning for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8065–8081, Nov. 2022.

[39] H.-M. Yang, X.-Y. Zhang, F. Yin, and C.-L. Liu, "Robust classification with convolutional prototype learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3474–3482.

[40] R. Ranjan, C. D. Castillo, and R. Chellappa, "L2-constrained softmax loss for discriminative face verification," 2017, *arXiv:1703.09507*.

[41] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proc. 4th Int. Conf. Inf. Syst. Secur. Privacy*, 2018, pp. 108–116.

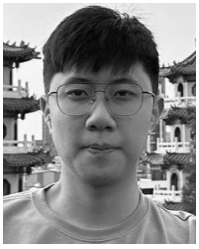
[42] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy," in *Proc. Int. Carnahan Conf. Secur. Technol. (ICCST)*, Oct. 2019, pp. 1–8.

[43] J. Mirkovic and P. Reiher, "A taxonomy of DDoS attack and DDoS defense mechanisms," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 2, pp. 39–53, Apr. 2004.

[44] A. Asosheh and N. Ramezani, "A comprehensive taxonomy of DDOS attacks and defense mechanism applying in a smart classification," *WSEAS Trans. Comput.*, vol. 7, no. 4, pp. 281–290, 2008.

[45] H. Gebrye, Y. Wang, and F. Li, "Traffic data extraction and labeling for machine learning based attack detection in IoT networks," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 7, pp. 2317–2332, Jul. 2023.

CHIN-SHIUH SHIEH (Member, IEEE) received the M.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1991, and the Ph.D. degree from the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan, in 2009. He joined as a Faculty Member of the Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, in August 1991, where he is currently a Professor. His research interests include wireless networks and handover techniques.



FU-AN HO received the M.S. degree from the National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan, in 2024. His research interests include cybersecurity and open-set recognition.

MONG-FONG HORNG (Member, IEEE) received the B.S. degree from National Chiao Tung University, in 1989, and the M.S. and Ph.D. degrees from National Cheng Kung University, Taiwan, in 1991 and 2003, respectively. He is currently a jointly-appointed Professor with the Department of Electronics Engineering, National Kaohsiung University of Science and Technology, and Kaohsiung Medical University, Taiwan. He was the President of Taiwanese Association of Consumer Electronics (TACE) and the Chair of Tainan Chapter, IEEE Signal Processing Society. His research interests include the Internet of Things, machine learning, computer networks, and medical informatics and related industrial cooperation.

THANH-TUAN NGUYEN was born in Khánh Hòa, Vietnam. He received the M.S. degree in electronics engineering from Ho Chi Minh City University of Technology, Vietnam, in 2014, and the Ph.D. degree in electronics engineering from the National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan, in 2023. Since 2014, he has been a Faculty Member of the Faculty of Electrical and Electronics Engineering, Nha Trang University, Vietnam. His research interests include information security, intelligent computation, and heuristic optimization.

PRASUN CHAKRABARTI (Senior Member, IEEE) received the Ph.D. (Engg.) degree in computer science and engineering from Jadavpur University in 2009, the higher doctoral (D.Litt) degree from Sambalpur University in 2022. In 2022, he registered for another higher doctoral degree (D.Sc.). He is working as the Director, the Directorate of Research and Publication, the Dean (International Affairs), a Senior Professor with the Department of Computer Science and Engineering, and the Officiating Dean with the School of Engineering, Sir Padampat Singhania University, India. He became a Full Professor at only 34 years and is one of the youngest Deputy Vice Chancellors of Gujarat at only 40 years. He has 127 SCI/Scopus indexed publications (SCI -73), 11 books, 70 granted International patents, four granted Indian Design patents, and 16 granted Indian copyrights. He has supervised 11 Ph.D. candidates successfully. On various research assignments, he has visited Waseda University Japan (2012) availing prestigious INSA-CICS travel grant, University of Mauritius (2015), Nanyang Technological University, Singapore (2015, 2016, and 2019), Lincoln University College Malaysia (2018), National University of Singapore (2019), Asian Institute of Technology Bangkok Thailand (2019), and ISI Delhi (2019). He is a fellow of Institution of Engineers (India), IET (U.K.), Royal Society of Arts London, Iranian Neuroscience Society, IETE, ISRD (U.K.), IAER (London), Nikhil Bharat Shiksha Parisad (Government of West Bengal), and Scientific Communications Research Group Egypt. He is also a Visiting Senior Scientist with the National Kaohsiung University of Science and Technology, Taiwan; a Visiting Distinguished Research Fellow with Wales Institute of Digital Information, U.K.; a Honorary Adjunct Distinguished Professor with Don State Technical University, Russia; a Visiting Professor with Shiraz University of Medical Sciences, Iran; and a Honorary Visiting Distinguished Scientist with PLANET Laboratory, Politecnico di Torino, Italy.