Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Supplementary Materials for

Zhihui LI, Congyuan XU, Kun DENG, Chunyuan LIU, 2025. A subspace-based few-shot intrusion detection system for the Internet of Things. *Front Inform Technol Electron Eng*, 26(6):862-876.
https://doi.org/10.1631/FITEE.2400556

1 Simulations and analysis

1.1 Datasets

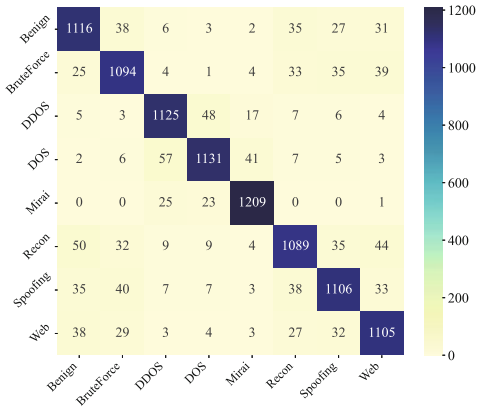
A DDoS is a distributed denial-of-service attack that utilizes many legitimate distributed servers to send requests to a target, thereby preventing normal legitimate users from obtaining services. A DoS is a denial-of-service attack in which an attacker, through various means, prevents the target system or network resources from providing regular services to legitimate users. Mirai is a type of malware that creates a botnet by infecting IoT devices such as cameras, routers, and other smart devices, and a network can be used to execute large-scale DDoS attacks. Spoofing attacks enable malicious actors to operate under the identity of the victim's system to gain illegal access to network traffic. Such attacks are widely used for system access, data theft, and malware distribution. Recon attacks are typically used to gather information regarding a target network or system as a preparatory step for an attacker to access deeply. Web attacks are designed to exploit web service vulnerabilities in IoT devices to gain unauthorized access, steal sensitive information, and perform malicious actions. The Brute force attack is a repeated attempt to gain access to a system by making repeated attempts at a password or password phrase.

Table S1 Information on the CICIDS2017 dataset

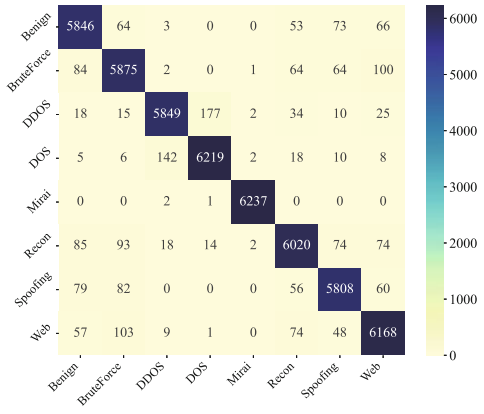
Categories	Training set	Test set
Benign	35065	14935
Bot	1379	564
DDoS	6992	3008
Dos	21797	9459
FTP-Patator	5554	2340
PortScan	6959	3041
SSH-Patator	4128	1733
Web	1533	667

Table S2 Information on the CICIDS2018 dataset

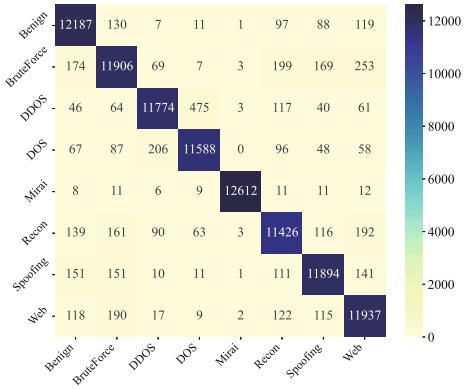
Categories	Training set	Test set
Benign	62936	27055
Bot	6910	3090
BruteForce	14571	6270
DDoS	15242	6487
Dos	14084	5961
Web	14110	5977



(a) 5-way 1-shot



(b) 5-way 5-shot



(c) 5-way 10-shot

Fig. S1 Confusion matrices in different settings: (a) 5-way 1-shot; (b) 5-way 5-shot; (c) 5-way 10-shot

1.2 Simulation results

1.2.1 Confusion matrices

Fig. S1 shows the confusion matrices under different settings, where the highest number of misclassified samples is observed for DDoS being incorrectly identified as DoS, and vice versa. Similarly, for Brute Force attacks, the highest number of misclassifications occurs with Web attacks, and the same is valid for Web attacks being incorrectly identified as Brute force. By analyzing the numerical features of these two groups of attacks, we found that their distributions were strikingly similar, which led to the propensity of the model for misclassification.

1.2.2 Simulation results on the CICIDS2017 and CICIDS2018 datasets

We conducted additional simulations on the classic datasets CICIDS2017 and CICIDS2018 to evaluate our model better. On these two datasets, we performed binary classification and multi-class classification simulations. For binary classification, similar to previously, we only considered the cases with K= 5 and K= 10. We compared the model's detection capabilities under settings with K = 1, 5, and 10 for multi-class classification. Similar to the simulations set up in Section 4.2, we selected one category as the unknown category for parallel simulations each time to verify the model's generalization ability to unknown category traffic.

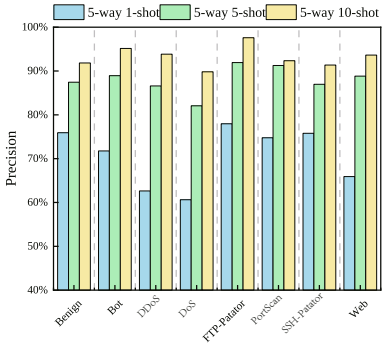


Fig. S2 Detection precision on the CICIDS2017 dataset

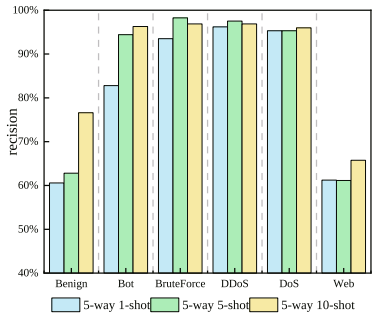


Fig. S3 Detection precision on the CICIDS2018 dataset

The results of the binary classification simulations are shown in Table 4. The simulation results are 99.16% and 99.26% on the CICIDS2017 dataset and are 99.06% and 98.57% on the CICIDS2018 dataset. From the simulation results, it can be seen that the model has a very high detection accuracy for both benign and malicious traffic, both fluctuating around 99%.

The results of the multi-class classification simulations are shown in Fig. S2 and S3 and Table 5. As shown in Fig. S2, the model's detection precision for various types of traffic in CICIDS2017 under different settings is displayed, and it is evident that the model's detection precision increases with the increase in the number of samples (K value). Fig. S3 shows the model's detection precision for various types of traffic in CICIDS2018 under different settings, and similar to the previous case, the simulation conclusion is that the detection precision increases with the increase in the number of samples. Among these, the detection effects for Bot, Brute force, DDoS, and DoS attacks are the best, while the detection effects for benign traffic and Web attacks are the worst. In this simulations, we labelled the Infiltration and SQL Injection attacks in the CICIDS2018 dataset as Web attacks. SQL Injection attacks may have features very similar to normal traffic, mainly when attackers use advanced techniques to hide their attack behavior. These attacks may not contain threatening keywords or irregular structural features, making them difficult to distinguish based solely on traffic features. This is why the detection effects for benign traffic and Web attacks are relatively poor. Table 5 also provides the detection accuracy rates on these two datasets. The detection accuracy rates under different settings are 83.88%, 92.66%, and 95.58% on CICIDS2017 and are 91%, 94.16%, and 94.7% on CICIDS2018.

2 Discussion

2.1 Visualization analysis

Network traffic is abstract, and T-SNE is used to downsize and visualize the dataset to facilitate the observation of connections between different flows. In Fig. S4, web attacks, spoofing attacks, and benign traffic are visualized, and the distance between the categories is compact. Through continuous training, our model performed well in the detection task. By downscaling and visualizing the feature vectors of the last layer of the neural network, it can be visualized that our model accurately detects samples of each class and increases the interclass distance. The visualization results are presented in Fig. S5.

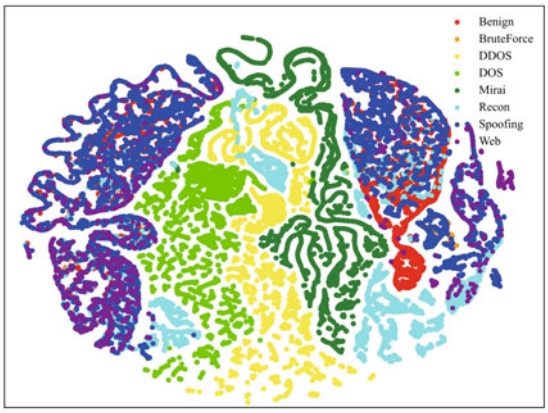


Fig. S4 Visualization of dataset before model training

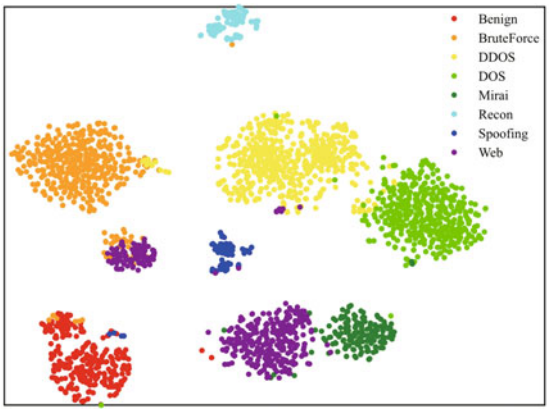


Fig. S5 Classification visualization after model training