Proceedings of the 3rd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/119/2025.TJ21647

Network Anomaly Traffic Detection Model Based on Spatio-Temporal Attention Feature Fusion

Xinpeng Chen[1,a,*,†], Jinan Shen[1,†], Fang Liang[1,†], Qiuyang Du[1,†]

[1]College of Intelligent Systems Science and Engineering, Hubei Minzu University, Enshi, 445000, China
a. 13277174467@163.com
**corresponding author*
[†]These authors contributed equally to this work.

Abstract: With the rapid proliferation of 5G, IoT devices are increasingly subjected to network attacks. Traditional Network Intrusion Detection Systems (NIDS) are becoming inadequate in the face of more complex network environments and massive data traffic. Deep learning-based intrusion detection algorithms have thus become a hot topic in cybersecurity research. However, existing NIDS have limitations in terms of accuracy, recall rates, false alarm rates, and generalization capabilities. The impact of data redundancy and data imbalance further degrades model performance. We have designed a new network that integrates Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, incorporating an attention mechanism to enhance learning capabilities. This allows the model to focus on both temporal and spatial features. Additionally, the introduction of the attention mechanism makes it easier to identify key anomalous data features amidst the redundant data. We also paid special attention to the issue of data imbalance. Our model and methods for balancing datasets were validated through binary and multi-class classification experiments on the two most commonly used datasets (UNSW_NB15, CICIDS2017). The results demonstrated good convergence and high accuracy. Compared to traditional models, our model shows significant improvements in detecting large-scale and multi-scenario network data attacks, making it suitable for network security detection in modern IoT environments.

Keywords: Network Intrusion Detection Systems, Convolutional Neural Networks, Long Short-Term Memory.

1. Introduction

As the Internet of Things (IoT) technology advances, it has been widely applied in various aspects of urban life. The IoT market is currently growing at a rapid rate of 16.7\%, and its net value is expected to exceed $300 billion. While these systems provide intelligent services, they also bring many security risks. Therefore, designing intelligent Network Intrusion Detection Systems (NIDS) to mitigate issues such as data leakage and theft has become a focal point for many researchers[1]. Intrusion detection methods can be classified into host-based intrusion detection systems and network-based intrusion detection systems based on the source of data. Additionally, they can be categorized into misuse intrusion detection and anomaly intrusion detection based on the detection method. Based on

DOI: 10.54254/2755-2721/119/2025.TJ21647

detection methods, IDS can be classified into Signature-based IDS (SIDS) and Anomaly-based IDS (AIDS)[2].

Traditional machine learning techniques have found extensive applications in the realm of detecting anomalous network traffic. Approaches such as naive Bayes[3,4], k-means[5,6], random forest[7,8], support vector machine[9,10], XGBoost[11,12], and decision tree[13,14] have demonstrated success in the detection of anomalous network traffic. In contemporary times, the scale of traffic has evolved significantly from the past. Traditional machine learning techniques prove inadequate when confronted with vast and intricate traffic data.

In current research on anomalous traffic detection, the proportion of hybrid models is increasing, and the experimental results are getting better, but there are still some problems. We conducted research on CNN-LSTM hybrid models and found that typically, the hybridization of CNN and LSTM models can lead to feature omission and poor learning capabilities. To address this, we first optimized the CNN and LSTM models to enhance their feature extraction capabilities. Additionally, we incorporated an attention mechanism to improve the model's learning and filtering capabilities, resulting in excellent experimental outcomes.

2. Methodology

Taking into consideration the characteristics of network traffic, we have designed a model for network traffic anomaly detection based on the fusion of spatio-temporal attention features. The overall structure of the model is illustrated in Fig 1.

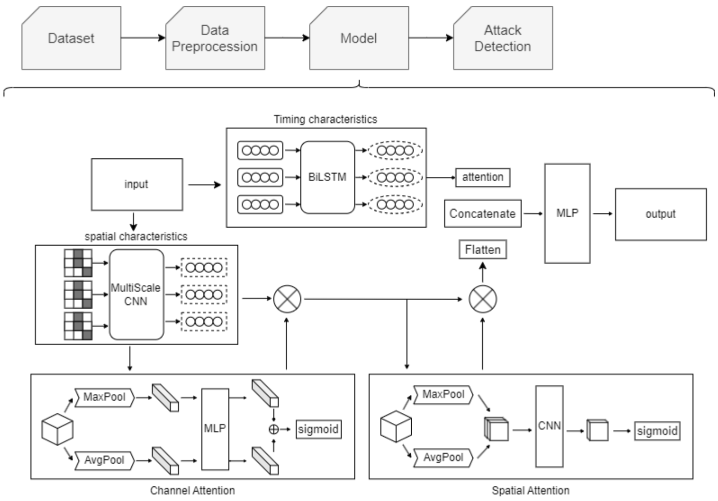

Figure 1: The overall structure of the mode.

This model concurrently inputs data into both the CNN and BiLSTM networks, facilitating the extraction of spatial and temporal features simultaneously. In comparison to models employing a single class of feature extraction, this approach yields richer and more complex data features. To ensure the model can effectively learn these intricate features without struggling due to their complexity, we incorporate a channel spatial attention module and a general attention module. These modules assign weights to temporal and spatial data features, preventing difficulties in learning caused by the high complexity of the data. This enhancement enables the model to focus on capturing the essential parts of the data features.

In the BLoCNet literature, the model initially inputs data into the CNN for compressed feature extraction. Subsequently, it passes through two layers of BiLSTM to learn data features. However, due to the recursive nature of feature extraction, some data features are lost in this process. To address this issue, this paper adopts a novel approach. It juxtaposes the number of spatial and temporal

DOI: 10.54254/2755-2721/119/2025.TJ21647

features extracted through both the CNN layer and BiLSTM layer. This innovative method significantly reduces the loss of features during the recursive extraction of data features.

2.1. Bi-LSTM Extraction Of Temporal Features

The extraction of temporal features is primarily achieved through the BiLSTM network. Compared to the traditional LSTM network, it can simultaneously capture data features extracted by both the forward and reverse LSTMs.

This capability renders it more comprehensive and effective than traditional LSTM in learning the relationships within data sequence features.

2.2. Multi-scale convolution to extract spatial features

CNN are highly effective in reducing dimensionality and compressing image parameters. They inherently excel at extracting spatial feature relationships in data. Hence, we integrate them into our model for the purpose of extracting spatial features. Compared to the traditional CNN, we have enhanced it for traffic data by designing three convolutional layers with kernel sizes of 3x3, 5x5, and 7x7, respectively. After converting the traffic data into grayscale maps and inputting them into three distinct convolutional layers, the extraction of spatial features at different scales becomes more comprehensive compared to using a single convolutional layer with a uniform window size. We combine the output of the three feature datasets into a 3x3 convolutional layer. Additionally, we incorporate a dropout layer and a batch normalization layer to accelerate model training and mitigate potential overfitting issues arising from multi-scale feature extraction. As evident, our convolutional layer deliberately excludes a pooling layer. This decision is based on the low complexity observed in grayscale maps. Introducing a pooling layer may not efficiently compress the data and could result in feature loss. Furthermore, this approach is designed to preserve feature data for enhanced learning in subsequent attention modules.

2.3. Attention Mechanisms

Given that our model incorporates both Bidirectional Long and BiLSTM and Multi-Scale Convolutional Neural Networks, it becomes more comprehensive in extracting traffic data and less prone to overlooking crucial features. However, this design choice introduces the challenge of parameter explosion. If no further processing is applied and the data is directly passed into the fully connected layer for prediction output, the model's learning time will be prolonged, and the fitting process becomes notably challenging.

3. Experiment Configuration

In order to assess the effectiveness of the model, we employed six metrics: Accuracy, Precision, Recall, F1 score, and Confusion Matrix. The acc metrics appear in the text as abbreviations for accuracy. The formulas for these metrics are as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

DOI: 10.54254/2755-2721/119/2025.TJ21647

These metrics are calculated using the following variables: True Positives (TP): Positive samples correctly predicted as positive by the model. False Positives (FP): Negative samples predicted as positive by the model. False Negatives (FN): Positive samples predicted as negative by the model. True Negatives (TN): Negative samples correctly predicted as negative by the model.

We have presented the accuracy rates of the datasets before and after balancing, and listed the performance of other models, as shown in Table Table.1 It can be seen that there are significant differences in model performance before and after balancing the datasets. The accuracy of the UNSW-NB15 dataset after balancing is lower than before balancing because we significantly reduced the amount of normal traffic data. This makes it easier to detect the model's performance on minority class traffic. According to the confusion matrix mentioned earlier, the recognition rate of our minority class labels has greatly improved.

Table 1: The performance of other models.

Algorithm	Year	Data	ACC
CNN-BLSTM-Attention	Proposed Model	UNSW-NB15 The Balanced UNSW-NB15 CICIDS2017 The Balanced CICIDS2017	92.9% 85.1% 98.5% 99.36%
CNN-BLSTM	Proposed Model	UNSW-NB15 CICIDS2017	90% 94%
BLoCNet[18]	2023	CICIDS2017	93.2%
CNN-LSTM[19]	2022	UNSW-NB15	82.41%
Vision Transformer[31]	2022	CICIDS2017 UNSW-NB15	96.4% 89.8%
Multi-Head Attention + BiLSTM[32]	2023	UNSW-NB15	99.08%
Transformer[32]	2023	UNSW-NB15	97.94%
Tree-Based Ensemble Network[33]	2023	CICIDS2017	96.5%
DBN[34]	2023	UNSW-NB15	66.6%
CNN-BiLSTM[36]	2023	CICIDS2017	84.23%
CNN-BiLSTM[37]	2020	UNSW-NB15	77.16%
DNN[38]	2019	CICIDS2017 UNSW-NB15	96% 66%
DNN[39]	2021	UNSW-NB15	81.7%
ELM[40]		UNSW-NB15	66.33%
Negative selection algorithm and classifiers [41]	2018	CICIDS2017	97%
MLP[42]	2019	CICIDS2017	87.2%
DeepGFL[43]	2018	CICIDS2017	53.1%
DBN[44]	2022	CICIDS2017	94%
CNN-LSTM[45]	2023	CICIDS2017	99.27%
AdaBoost-EFS[46]	2019	CICIDS2017	81.83%
CLAIRE[47]	2023	CICIDS2017	98%
LSTM[48]	2020	UNSW-NB15	70%
HC-DTTSVM[49]	2023	UNSW-NB15	81.21%
ANN-MLP[50]	2020	UNSW-NB15	76.96%

4. Conclusion and Future Work

We have developed a network anomaly traffic detection model that uses spatio-temporal attention feature fusion, which is suitable for the characteristics of traffic data. When analyzing traffic data, we observed a clear data imbalance phenomenon, which may have a negative impact on model training. To address this issue, we used oversampling and undersampling methods to ensure data balance and prevent the model from overfitting data features. The preprocessed data is input into a multi-scale convolution and bidirectional BiLSTM network to extract spatial and temporal features simultaneously. Given the complexity and richness of the extracted features, we introduced an attention mechanism to calculate the attention weights of spatial and temporal features.

In our experiments, we thoroughly compared the results before and after data balancing, proving that preprocessing is crucial for the model's performance in the case of data imbalance. We also compared our model with many other models, demonstrating that our model performs well in both binary and multi-class classification scenarios.

However, the experiments also revealed some limitations. For example, some minority class samples are still difficult to learn, and the detection performance is not ideal. We speculate that the features of these types of samples are not obvious, and the model has not learned them. Additionally, different datasets provide different types of traffic data features. Some features are present in some datasets but not in others.

References

[1] *Bataev, A.V.; Zhuzhoma, I.; Bulatova, N.N. Digital Transformation of the World Economy: Evaluation of the Global and Russian Internet of Things Markets. In Proceedings of the 2020 9th International Conference on Industrial Technology and Management (ICITM). IEEE, 2020, pp. 274–278.*

[2] *Aldweesh, A.; Derhab, A.; Emam, A.Z. Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. Knowledge-Based Systems 2020, 189, 105124.*

[3] *Shitharth, S.; Kshirsagar, P.R.; Balachandran, P.K.; Alyoubi, K.H.; Khadidos, A.O. An innovative perceptual pigeon galvanized optimization (PPGO) based likelihood Naïve Bayes (LNB) classification approach for network intrusion detection system. IEEE Access 2022, 10, 46424–46441.*

[4] *Islam, R.; Devnath, M.K.; Samad, M.D.; Al Kadry, S.M.J. GGNB: Graph-based Gaussian naive Bayes intrusion detection system for CAN bus. Vehicular Communications 2022, 33, 100442.*

[5] *Saheed, Y.K.; Arowolo, M.O.; Tosho, A.U. An efficient hybridization of k-means and genetic algorithm based on support vector machine for cyber intrusion detection system. International Journal on Electrical Engineering and Informatics 2022, 14, 426–442.*

[6] *Jain, M.; Kaur, G.; Saxena, V. A K-Means clustering and SVM based hybrid concept drift detection technique for network anomaly detection. Expert Systems with Applications 2022, 193, 116510.*

[7] *Waskle, S.; Parashar, L.; Singh, U. Intrusion detection system using PCA with random forest approach. In Proceedings of the 2020 International Conference on Electronics and Sustainable 445 Communication Systems (ICESC). IEEE, 2020,pp. 803–808.*

[8] *Liu, C.; Gu, Z.; Wang, J. A hybrid intrusion detection system based on scalable K-means+random forest and deep learning. Ieee Access 2021, 9, 75729–75740.*

[9] *Almaiah, M.A.; Almomani, O.; Alsaaidah, A.; Al-Otaibi, S.; Bani-Hani, N.; Hwaitat, A.K.A.; Al-Zahrani, A.; Lutfi, A.; Awad, A.B.; Aldhyani, T.H. Performance investigation of principal component analysis for intrusion detection system using different support vector machine kernels. Electronics 2022, 11, 3571.*

[10] *Kumari,A.; Mehta,A.K. A hybrid intrusion detection system based on decision tree and support vector machine. In Proceedings of the 2020 IEEE 5th International conference on computing communication and automation (ICCCA). IEEE, 2020,pp. 396–400.*

[11] *Devan, P.; Khare, N. An efficient XGBoost–DNN-based classification model for network intrusion detection system. Neural Computing and Applications 2020, 32, 12499–12514.*

[12] *Bhati, B.S.; Chugh, G.; Al-Turjman, F.; Bhati, N.S. An improved ensemble based intrusion detection technique using XGBoost. Transactions on emerging telecommunications technologies 2021, 32, e4076.*

[13] *Louk, M.H.L.; Tama, B.A. Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system. Expert Systems with Applications 2023, 213, 119030.*

[14] *Kasongo, S.M. An advanced intrusion detection system for IIoT based on GA and tree based algorithms. IEEE Access 2021, 9, 113199–113212.*