Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795

ABNORMAL TRAFFIC DETECTION BASED ON ATTENTION AND BIG STEP CONVOLUTION

[1] Dr. Subba Reddy Borra,[2] M.Nikhila,[3] M. Manju Bhargavi,[4] P. Sushma
[1]Professor,[234]Students
Department Of CSE
Malla Reddy Engineering College for Women

ABSTRACT

The identification of abnormal traffic is essential to network security and service quality. A big-step convolutional neural network traffic detection model based on the attention mechanism is provided as a solution to the significant challenges in abnormal traffic identification caused by feature similarity and the detection model's single dimension. First, the raw traffic is preprocessed and mapped into a two-dimensional grayscale picture after the network traffic characteristics are examined. After that, histogram equalization is used to create multi-channel grayscale pictures. An attention mechanism is then added to give traffic characteristics varying weights in order to improve local features. In order to improve the flaws in convolutional neural networks, including local feature omission and overfitting, pooling-free convolutional neural networks are finally integrated to extract traffic characteristics of various depths. Both a real data collection and a balanced public data set were used for the simulation experiment. The suggested model is contrasted with ANN, CNN, RF, Bayes, and the two most recent models using the widely used method SVM as a baseline. 99.5% accuracy percentage with several classes is achieved experimentally. The best anomaly detection is found in the suggested model. Additionally, the suggested technique performs better in F1, recall, and accuracy than existing models. It is shown that the model is not only effective in detecting things, but also resilient to a variety of complicated contexts.

1. INTRODUCTION

Utilized by people from all walks of life, internet technology has greatly aided in the development of both the economy and society. However, due to the numerous flaws in the mainstream network security and defense technologies currently in use, the extensive application requirements also increase the complexity of the network's security configuration, leaving the entire system open to attack. Simultaneously, as a result of the TCP/IP network architecture's openness, computer viruses were able to proliferate via disguise and disrupt network functionality, leading to a decline in social and economic conditions. It is crucial to maintain network security to know how to evaluate data effectively in order to forecast

Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795

how the network will evolve going forward, identify anomalies, and take the necessary corrective action [1].

Network traffic categorization may assist in the discovery of anomalous traffic. The three primary techniques are port-based [2], deep packet detection based [3], and machine learning based [4], all of which are in line with its fundamental notion. Traditional machine learning and deep learning are the two types of machine learning. The first two algorithms had consistent performance and produced excellent classification results in the early days of the Internet, when it was tiny and the kinds of traffic were simple [5, 6, 7]. However, the categorization impact is diminished by the ongoing development of new Internet applications, which leads to an increase in the kinds of traffic and a rise in the complexity of traffic components. To overcome the shortcomings of the aforementioned techniques, machine learning enhancement techniques are suggested. The purpose of machine learning is to reliably efficiently and accurately categorize network traffic by extracting its statistical properties. It has several potential applications as well.

The collection of data sets, the creation of normalized data, data pre-processing, feature extraction, model training, and classification make up the whole process of classifying network traffic. Different techniques are used in traditional machine learning classification to choose the best subset of features that closely resemble the whole feature results for classification. This method depends on feature selection, which might have a direct impact on the classification outcomes and is unable to keep up with the changes in contemporary network traffic. Furthermore, the intricate interactions between distinct characteristics cannot be represented by conventional machine learning methods. Deep learning therefore emerges as the best technique for handling network traffic classification, exhibiting excellent performance in demanding and dynamic traffic categorization scenarios.

Numerous works on deep learning have focused on classifying network traffic in the last few years [8, 9, 10]. These findings show that deep learning approaches are feasible and can improve performance for traffic classification tasks, but they also show that deep learning research on network anomaly detection is still in its early stages. Unlike machine learning, deep learning reflects intricate nonlinear connections between variables and allows for the automated extraction of structured and complicated characteristics, which are then fed directly into a training classifier to enable the categorization of network traffic. In conclusion, both the feasibility and the quality of the anomalous traffic detection model for network security defense have increased. But there are still a lot of issues: First, for traffic data with comparable

Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795

attribute qualities, the categorization results are subpar. Second, the anomalous network detection model's rigid structure makes it impossible to extract features in a variety of dimensions and fields of view, which somewhat lowers the accuracy of classifying network data. Third, there is a chance of information loss when repeated pooling using convolution neural networks, which might lessen the significance of the sequence.

In order to address the aforementioned obstacles and issues, this study offers the following contributions:

• Based on the attention mechanism, we provide an Attention and Big Step Convolution Neural Network (ABS-CNN) model in this research [11]. Assigning attention weights to data sequences might help the attention mechanism differentiate subtle aspects and address issues like comparable features producing inferior classification results. Assigning attention weights to data sequences might help the attention mechanism differentiate subtle aspects and address issues like comparable features producing inferior classification results. Experiments demonstrate that the upgraded feature model is more resilient and has a greater classification accuracy. In this study, we tackle the single model dimensionality issue by means of histogram equalization. After converting the traffic data into grayscale pictures, the images undergo histogram equalization. enhanced multi-channel convolution in conjunction with it to automatically extract and merge fine-grained multi-field features. The results of the trials demonstrate that very well-defined traffic is produced when histogram equalization is used, leading to improved resilience and model detection performance.

• Big-step convolution is combined to extract traffic characteristics in order to compensate for the decreased correlation of traffic sequences caused by pooling. Furthermore, stepwise convolution is another name for big-step convolution. Stepwise convolution lessens the damage caused by accuracy loss owing to information loss while preserving the sequence-related properties that the convolution layer extracted.

There are five sections to this study. The research background and primary contributions of this study are briefly described in Section I. Section II will provide an overview and description of the state of the research. The introduction of the model and the implementation of the algorithm are completed in Section III. In Section IV, comprehensive experiments and a thorough analysis of the findings will be conducted. In conclusion, Section V will provide an analysis and summary of the model presented in this work as well as suggest some potential avenues for further study.

2. EXISTING SYSTEM

Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795

A cost-sensitive SVM (CMSVM) was suggested by Shi et al. [16] as a solution to the network traffic imbalance issue. Adaptive weights are used by the model to handle the imbalance issue for various applications via the use of a multi-class SVM and an active learning method. A real-time network classification model using SPPSVM was suggested by Cao et al. [17]. The model reduces the dimensionality of the original data using the principle component analysis (PCA) feature selection approach, and it applies an enhanced particle swarm optimization process to determine the ideal parameters. When compared to the conventional SVM model, the classification accuracy is better. To identify anomalous traffic while removing duplicate characteristics from the traffic data, Farid et al. [18] used naive bayes and decision trees. The detection rate is increased by the suggested algorithm. The majority of machine learning-based categorization techniques rely on human feature design and selection, which is insufficient given how networks are evolving today.

A unique autoencoder-based deep neural network was suggested by Gianni et al. [19]. In order to extract the fundamental characteristics of interest, the model embeds many autoencoders into convolutional and recurrent neural networks. Stacked fully connected neural networks are used to classify network traffic.

A tree-structured recurrent neural network was introduced by Ren et al. [20] to split big classification tasks into smaller ones using a tree structure. With a superior classification effect, the model can automatically discover the nonlinear connection between the input and output data. A novel technique for classifying encrypted communications was presented by Tal et al. [21]. In order to classify traffic, the approach first transforms traffic data into comprehensible visuals. Next, it integrates convolutional neural networks to classify the images. In order to address the issue that recurrent neural networks are prone to gradient expansion or disappearance, Li et al. [22] suggested a bidirectional independent recurrent neural network with parallel operations and configurable gradients. By using both forward and backward inputs, the model captures the bi-directional structural aspects of network traffic and combines global attention to highlight its key components.

A multi-level feature fusion model was presented by Lin et al. [23] as a solution to the data imbalance issue. For improved efficiency, the model incorporates statistical, byte, and data temporal aspects. A traffic categorization model called TSCRNN based on spatial and temporal variables was suggested by Lin et al. [24]. To accomplish effective traffic categorization, the model first preprocesses the original data and then uses CNN and bi-directional RNN to learn the traffic's spatial and temporal properties.

Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795

An integrated model for deep learning was presented by Saadat et al. [25]. In order to classify network traffic, the model first automatically extracts traffic characteristics using a one-dimensional convolutional neural network. This is followed by the efficient feature selection of ALO and the clustering of SOM-based data.

Disadvantages

- To enhance the efficacy and efficiency of Abnormal Traffic Detection Generation, the current system does not use hybrid deep learning or a good ML model detection strategy.
- The attention-based big-step convolutional neural network (ABS-CNN) model is more accurate and efficient than an existing system.

3. PROPOSED SYSTEM

• Utilizing the attention mechanism as its foundation, this research presents an Attention and Big Step Convolutional Neural Network (ABS-CNN) model [11]. Assigning attention weights to data sequences helps differentiate subtle characteristics, which is useful for solving issues like comparable features leading to inferior classification results. Assigning attention weights to data sequences helps differentiate subtle characteristics, which is useful for solving issues like comparable features leading to inferior classification results. Both the classification accuracy and resilience of the model with improved features are shown experimentally to be greater.

• To address the issue of one-dimensional models, histogram equalization is used in this work. After converting the traffic data to grayscale, the photos are histogram equalized. In conjunction with enhanced multi-channel convolution, it can automatically extract and combine fine-grained features from several fields. There is an improvement in model detection efficiency and resilience as a consequence of the generally well-defined traffic seen in the studies after histogram equalization. The traffic characteristics are retrieved using merging big-step convolution in order to tackle the decreased correlation of traffic sequences caused by pooling. Another name for big-step convolution is stepwise convolution. By reducing the impact of accuracy loss caused by information loss, stepwise convolution keeps the sequence-related characteristics that the convolution layer extracted.

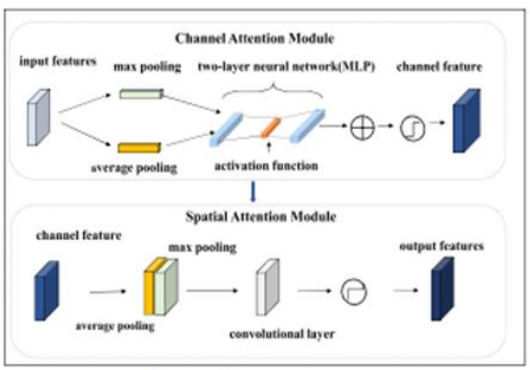
Advantages

- The ABS-CNN model has an input layer, three convolutional layers, a fully connected layer, and an output layer. To improve convolution's capability to extract traffic information, a convolutional attention mechanism is included.

Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795

- To test the influence of each component on the model, the ablation research is carried out in the proposed system by deleting each component from the ABS-CNN one by one and comparing it with the ABS-CNN of the whole pair. In order to investigate how the attention mechanism, histogram equalization, and large-step convolution impact the performance of the model and.

4. SYSTEM ARCHITECTURE



5. ALGORITHIMS

Gradient boosting

A machine learning approach known as gradient boosting is used for a variety of applications, including classification and regression analysis. It provides a paradigm for making predictions by combining several weak prediction models, most often decision trees.(1) and (2) A technique known as gradient-boosted trees is produced when a decision tree is used as the weak learner. In most cases, this approach achieves better results than random forest.While other boosting approaches use a stage-wise construction, gradient-boosted trees take it a step further by enabling optimization of any differentiable loss function.

Logistic regression Classifiers

The relationship between a group of independent factors and a categorical dependent variable is examined in logistic regression analysis. When the dependant variable may only take on two values—for example, yes or no—the method is known as logistic regression. When the dependent variable may take on three or more distinct values, such "Married," "Single," "Divorced," or "Widowed," the method is known as multinomial logistic regression. While the dependent variable data format differs from multiple regression, the procedure's practical application is comparable.

In the realm of categorical-response variable analysis, logistic regression is in direct competition with discriminant analysis. Logistic regression, according to many statisticians, is more flexible and appropriate for modeling the majority of cases than discriminant analysis. Reason being, in contrast to discriminant analysis, logistic regression does not presume regularly distributed independent variables.

Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795

Using both numerical and categorical independent variables, this application calculates multinomial logistic regression and binary logistic regression. All of the following are reported: likelihood, deviance, odds ratios, confidence limits, and quality of fit for the regression equation. A full residual analysis, including diagnostic residual reports and graphs, is executed by it. So that it can find the optimal regression model with the fewest independent variables, it may do an independent variable subset selection search. It helps find the optimal classification cutoff point and gives confidence intervals on expected values using ROC curves. By automatically categorizing rows that aren't utilized in the study, it lets you verify your findings.

SVM

The goal of discriminant machine learning in classification problems is to identify a discriminant function that can accurately predict labels for newly acquired instances using a iid (independent and identically distributed) training dataset. Instead of computing conditional probability distributions, as is necessary in generative machine learning methods, a discriminant classification function simply assigns a data point x to one of the classes involved in the classification job. While generative methods are often used for outlier identification in predictions, discriminant methods are more efficient in terms of training data and processing resources, particularly when dealing with a multidimensional feature space and just requiring posterior probabilities. Learning a classifier is geometrically similar to solving for the equation of a multidimensional surface that optimally divides the feature space into its constituent classes.

When compared to other popular machine learning classification methods like perceptrons or genetic algorithms (GAs), support vector machines (SVMs) always provide the same ideal hyperplane value due to their analytical solution of the convex optimization issue. The initiation and termination criteria have a significant impact on the solutions for perceptrons. Training produces uniquely specified SVM model parameters for a particular training set for a certain kernel that translates the data from the input space to the feature space. In contrast, the perceptron and GA classifier models are modified each time training is initiated. A number of hyperplanes will satisfy this criterion as GAs and perceptrons just attempt to reduce training errors.

Convolutional Neural Network (CNN)

For picture processing and identification jobs, a deep learning technique known as a Convolutional Neural Network (CNN) is ideal. Convolutional neural networks (CNNs) build hierarchical feature representations from raw input pictures,

Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795

requiring less preparation than other classification models. They are quite good at using convolutional layers, which use filters to find local patterns in pictures, to prioritize different objects and characteristics.

Convolutional neural networks (CNNs) take their connection pattern from the human visual cortex, where neurons have particular areas of visual space that they react to. Due to its design, CNNs are able to successfully detect patterns and correlations in pictures' spatial dimensions. High accuracy in tasks like object identification, segmentation, and picture classification is achieved by CNNs by stacking many convolutional and pooling layers, which allow them to learn more complicated information.

6. IMPLEMENTATION

Modules

Service Provider

See the downloaded predicted data sets, the trained and tested accuracy results in a bar chart, the predicted traffic type, the predicted traffic type ratio, the view of all remote users, and the results of the training and testing accuracy.

View and Authorize Users

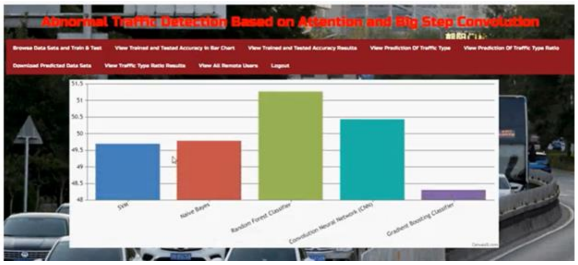
All users who have registered for this module may be seen by the administrator. Users' names, email addresses, and physical addresses are viewable to the administrator, who may also authorize users.

Remote User

In this module, you will find n users. The user must register before they may begin any activity. Upon registration, the user's details are stored in the database. He will be prompted to enter his approved username and password after he successfully registers. A user's profile, traffic type predictions, and the ability to register and log in are all accessible after a successful login.

7. SCREEN SHORTS

Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795



8. CONCLUSION

This study provides a detection model based on attention and big-step convolution to overcome the challenges created by comparable characteristics and single model structure on anomalous traffic identification. Both publicly accessible datasets and datasets from actual environment crawls were used for the experiments. Performance analysis shows how effective the model is.

- In terms of accuracy, precision, recall, and F1-Score, ABS-CNN outperforms classical models. It has been shown that ABS-CNN produces predictions with excellent accuracy and a strong detection effect. Furthermore, the classification accuracy of many kinds of traffic is 100% based on the confusion matrix, indicating the excellent sensitivity of ABS-CNN in aberrant traffic detection.
- When compared to other CNN model variations, ABS-CNN operates more effectively and requires less testing and training time. Furthermore, ABS-CNN exhibits the greatest classification results with unmatched advantages in accuracy, precision, recall, and F1-Score.
- The ablation analysis's findings demonstrate that ABS-CNN adds an attention mechanism to allocate attention weights to various features, improving feature distinction and mitigating the challenges brought on by feature similarity. Histogram equalization is a data preparation technique introduced by ABSCNN that improves the single channel structure of the model and boosts its detection performance. Simultaneously, the sequence-related characteristics are retained when the pooling layer is removed, lowering training parameters, increasing operational efficiency, and achieving effective anomalous traffic detection.
- ABS-CNN exhibits exceptional detection results when tested on traffic that is crawled by a real environment. Encrypted

Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795

traffic is what the actual environment records. ABS-CNN shows the fine-grained capacity to encrypt harmful communication in addition to efficiently classifying encrypted traffic application kinds. This shows that the ABS-CNN has some resilience and can adjust to situations with varying levels of complexity. In addition to offering a potential fix for the issues with comparable characteristics and single model dimension on abnormal traffic recognition, the suggested method expands on the use of attention mechanisms and histogram equalization in abnormal traffic detection. Here are some ideas for future study directions:

• To collect samples for data pre-processing, it is still necessary to partition the data using the available network tools, which leads to a small number of samples being lost. Furthermore, the five-tuple sequence produced incorrect samples that were duplicated and lacked labels. In the future, do further study to identify more appropriate pre-processing steps and instruments.

• Examine the temporal and spatial correlations between various packets to investigate anomalous traffic identification in temporal and spatial mining.

REFERENCES

[1] O. Salman, I. H. Elhajj, A. Kayssi, and A. Chehab, ‘‘A review on machine learning–based approaches for internet traffic classification,’’ *Ann. Telecommun.*, vol. 75, nos. 11–12, pp. 673–710, Dec. 2020.

[2] A. Madhukar and C. Williamson, ‘‘A longitudinal study of P2P traffic classification,’’ in *Proc. 14th IEEE Int. Symp. Modeling, Anal., Simulation*, Monterey, CA, USA, Sep. 2006, pp. 179–188, doi: 10.1109/MASCOTS. 2006.6.

[3] S. Sen, O. Spatscheck, and D.Wang, ‘‘Accurate, scalable in-network identification of P2P P2P traffic using application signatures,’’ in *Proc. 13th Int. Conf. World Wide Web*, New York, MY, USA, May 2004, pp. 512–521.

[4] L. Ding, J. Liu, T. Qin, and H. Li, ‘‘Internet traffic classification based on expanding vector of flow,’’ *Comput. Netw.*, vol. 129, pp. 178–192,
Dec. 2017.

[5] T. Liu, Y. Sun, and L. Guo, ‘‘Fast and memory-efficient traffic classification with deep packet inspection in CMP architecture,’’ in *Proc. IEEE 5th Int. Conf. Netw., Archit., Storage*, Macau, China, Jul. 2010, pp. 208–217, doi: 10.1109/NAS.2010.43.

[6] N. Cascarano, L. Ciminiera, and F. Risso, ‘‘Optimizing deep packet inspection for high-speed traffic analysis,’’ *J. Netw. Syst. Manage.*, vol. 19, no. 1, pp. 7–31, Mar. 2011.

Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795

[7] G. Aceto, A. Dainotti, W. de Donato, and A. Pescape, ‘‘PortLoad: Taking the best of two worlds in traffic classification,’’ in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM)*, San Diego, CA, USA, Mar. 2010, pp. 1–5, doi: 10.1109/INFCOMW.2010.5466645.

[8] L. Vu, C. T. Bui, and Q. U. Nguyen, ‘‘A deep learning based method for handling imbalanced problem in network traffic classification,’’ in *Proc. 8th Int. Symp. Inf. Commun. Technol.*, Dec. 2017, pp. 333–339.

[9] P. Wang, F. Ye, X. Chen, and Y. Qian, ‘‘Datanet: Deep learning based encrypted network traffic classification in SDN home gateway,’’ *IEEE Access*, vol. 6, pp. 55380–55391, 2018.

[10] J. H. Shu, J. Jiang, and J. X. Sun, ‘‘Network traffic classification based on deep learning,’’ *J. Phys., Conf. Ser.*, vol. 1087, Sep. 2018, Art. no. 062021.

[11] D. Bahdanau, K. H. Cho, and Y. Bengio, ‘‘Neural machine translation by jointly learning to align and translate,’’ 2014, *arXiv:1409.0473*.

[12] C.Wang, T. Xu, and X. Qin, ‘‘Network traffic classification with improved random forest,’’ in *Proc. 11th Int. Conf. Comput. Intell. Secur. (CIS)*, Shenzhen, China, Dec. 2015, pp. 78–81, doi: 10.1109/CIS.2015.27.

[13] Z. Yuan and C. Wang, ‘‘An improved network traffic classification algorithm based on Hadoop decision tree,’’ in *Proc. IEEE Int. Conf. Online Anal. Comput. Sci. (ICOACS)*, Chongqing, China, May 2016, pp. 53–56, doi: 10.1109/ICOACS.2016.7563047.

[14] A. V. Phan, M. L. Nguyen, and L. T. Bui, ‘‘Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems,’’ *Appl. Intell.*, vol. 46, no. 2, pp. 455–469, Mar. 2017.

[15] B. Schmidt, A. Al-Fuqaha, A. Gupta, and D. Kountanis, ‘‘Optimizing an artificial immune system algorithm in support of flow-based internet traffic classification,’’ *Appl. Soft Comput.*, vol. 54, pp. 1–22, May 2017.

[16] S. Dong, ‘‘Multi class SVM algorithm with active learning for network traffic classification,’’ *Expert Syst. Appl.*, vol. 176, Aug. 2021, Art. no. 114885.

[17] J. Cao, Z. Fang, G. Qu, H. Sun, and D. Zhang, ‘‘An accurate traffic classification model based on support vector machines,’’ *Int. J. Netw. Manage.*, vol. 27, no. 1, Jan. 2017, Art. no. e1962.

[18] D. Md. Farid, N. Harbi, and M. Zahidur Rahman, ‘‘Combining Naive Bayes and decision tree for adaptive intrusion detection,’’ 2010, *arXiv:1005.4496*.

[19] G. D’Angelo and F. Palmieri, ‘‘Network traffic classification using deep convolutional recurrent autoencoder neural networks for spatial–temporal features extraction,’’ *J. Netw. Comput. Appl.*, vol. 173, Jan. 2021, Art. no. 102890.

[20] X. Ren, H. Gu, and W. Wei, ‘‘Tree-RNN: Tree structural recurrent neural network for network traffic classification,’’

Vol. 15 No. 3 (2024):218-229
DOI: https://doi.org/10.61841/turcomat.v15i3.14795

Expert Syst. Appl., vol. 167, Apr. 2021, Art. no. 114363.