Research on Malicious TLS Traffic Detection Based on Spatiotemporal Feature Fusion

Mingyue Qin[1], Mei Nian[1(✉)], Jun Zhang[1,2], and Bingcai Chen[1]

[1] School of Computer Science and Technology, Xinjiang Normal University, No. 102, Xinyi Road, Urumqi 830054, Xinjiang, China
2468830639@qq.com

[2] Xinjiang Institute of Physical and Chemical Technology, Chinese Academy of Sciences, Urumqi 830011, China

Abstract. Aiming at the problem that traditional machine methods rely on expert experience and the effect of malicious traffic identification is not ideal, a deep learning hybrid model is proposed to detect malicious TLS traffic. The model combines one-dimensional convolutional neural network and two-way long-term and short-term memory network to compress and extract network traffic features from two dimensions of space and time series. At the same time, attention score of output information is extracted by attention mechanism, and traffic identification is carried out by using mixed features obtained by fully connected neural network. Based on the open data set, the experimental results show that the accuracy, recall and F1 value of the model on the test set reach 94.67%, 89.66% and 91.08% respectively, which has good recognition effect.

Keywords: Malicious TLS traffic · Convolution neural network · BiLSTM · Attention mechanism

1 Introduction

In recent years, with the wide use of encryption technology, traffic encryption has become a standard practice. In 2015, 21% of the website traffic was encrypted. By 2019, more than 80% of the website traffic was encrypted, with a year-on-year growth of more than 90%. Malicious software uses encrypted channel and traffic encryption technology to achieve deep hiding and frequent variation of malicious behavior, resulting in a large number of malicious traffic characteristics in the Internet have not been found. Cisco released a security report in 2018, which pointed out that more than 70% of malicious software used TLS (Transport Layer Security) encryption technology to avoid exposure of attack behavior [1]. Traffic encryption not only protects users' privacy, but also provides convenience for many malware to hide their attacks. As a basic project of network defense, port obfuscation and port hopping technology used in encrypted traffic lead to a sharp decline in the accuracy of the traditional DPI and DFI methods based on plaintext [2], and the decryption behavior consumes a lot of computing resources and time. Therefore, how to accurately identify and quickly classify malicious TLS traffic without decrypting it has become a challenge.

Q. Liang et al. (Eds.): Artificial Intelligence in China, LNEE 854, pp. 318–325, 2022.
https://doi.org/10.1007/978-981-16-9423-3_40

2 Related Work

In view of the new challenges brought by the abuse of encryption technology to network management and security, academia and industry have turned their research focus to the use of machine learning technology based on load or behavior [3]. The workflow is as follows: first, design features manually (such as traffic features or grouping features), then extract and select appropriate features from the original traffic, and finally use classifiers (decision tree, naive Bayes and random forest, etc. [4–6]) to classify traffic.

Recently, researchers pay more attention to deep learning methods. Wang [7] proposed for the first time to convert the traffic into gray image and use two-dimensional CNN to extract the spatial characteristics of the traffic, but this method only uses the first 784 bytes of the whole flow, does not fully combine with other information of the traffic, and lacks the anonymization of the IP address. Cheng Hua et al. [8] proposed to transform the traffic load into sentence vector using word2vec model, and realized the identification of malicious encrypted C&C traffic through CNN.

The existing research on the identification and classification of encrypted traffic mainly focuses on the temporal or spatial characteristics of traffic. Usually, only one feature dimension of encrypted traffic is studied, which leads to the lack of robustness of the model. In the face of complex network traffic, the recognition effect may decline seriously.

To solve this problem, this paper proposes a malicious TLS traffic detection model based on CNN-BiLSTM-Attention. The model uses one-dimensional CNN and BiLSTM-Attention to compress and extract the spatial and temporal features of the traffic, and stitches the processed temporal and spatial features together to get a mixed feature vector, The full connected neural network is used to complete the recognition task. Experiments show that the effect of the model has been significantly improved compared with the existing research.

3 Model Design

3.1 Model Overview

In this paper, we use deep learning algorithm to automatically extract the temporal and spatial features from malicious encrypted traffic and train the classifier. In the spatial dimension feature learning module, one-dimensional CNN algorithm is used to extract the spatial features. The processed TLS streams are converted into two-dimensional gray image, and then the images are converted into a byte sequence of CSV file. 1D-CNN is used to learn the spatial features; In the time dimension feature learning module, BiLSTM-attention algorithm is used to learn time features in the field of time series classification, and attention mechanism is used to extract the attention score of the output information of BiLSTM. Finally, the features mined by the two deep neural networks are spliced and input into the softmax classifier to complete the identification and classification of malicious TLS traffic. The model architecture is shown in Fig. 1.

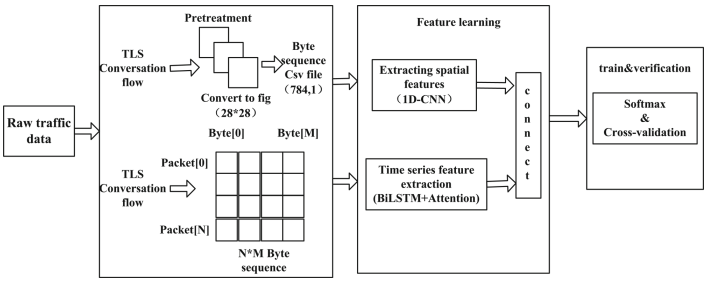

Fig. 1. Model architecture diagram

3.2 Data Preprocessing

The original data set file is in pcap format, which is composed of multiple data packets. The feature learning object of this model is data stream, so we first preprocess the data set and aggregate it into data stream. Firstly the streamdump tool is used to aggregate the packets into a data stream according to the packet quintuple. Secondly, the special MAC address, IP address and other specific information in the packet that interfere with the classification results are deleted.

3.3 Spatial Feature Extraction Model Based on 1D-CNN

Literature [7] shows that 1D-CNN, which is suitable for sequence data classification, can achieve better classification effect for traffic classification. Therefore, 1D-CNN model is used to compress and extract spatial features in traffic dataset.

CNN requires that the input dimension size is the same, and the data stream connects the information and content exchange part of the front part, which can better reflect the main characteristics of the whole data stream. Therefore, the first 784 bytes of data of each data stream are intercepted, and the data stream whose length is less than 784 bytes is supplemented with 0x00, and its category is marked.

The flow chart of one-dimensional CNN model in this paper is shown in Fig. 2, which is divided into two parts: 1) a cyclic structure composed of a convolution layer and a maximum pooling layer, which repeat two rounds; 2) Dropout layer, flatten layer and their connected fully connected neural networks.

3.4 Time Series Feature Extraction Model Based on BiLSTM-Attention

Network traffic has obvious hierarchical characteristics. The chain structure of bytes, packets and data streams is almost the same as the composition of words, sentences and paragraphs in natural language. Therefore, this paper uses the variant model of LSTM, bi-directional LSTM (BiLSTM), which is excellent in the field of natural language processing, to process the data. At the same time, considering the different importance of each packet in the session traffic, in order to highlight this difference and further improve the recognition effect of the model, attention mechanism is used to calculate the weight of the hidden layer output and weighted sum on the basis of BiLSTM model. The structure of the whole BiLSTM-Attention model is shown in Fig. 3.

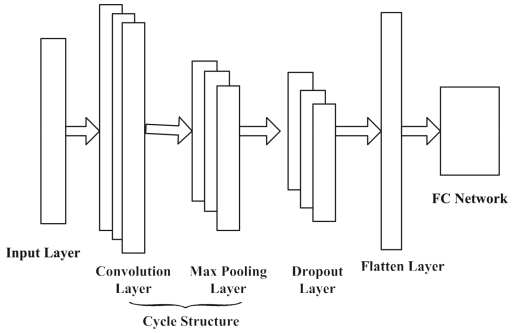

Fig. 2. CNN model

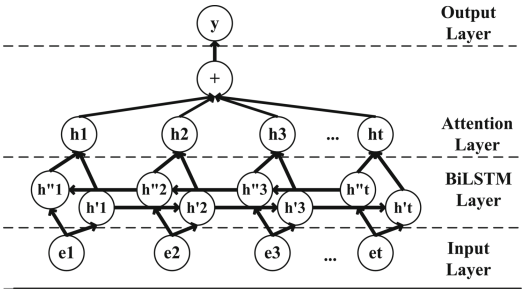

Fig. 3. BiLSTM-attention model

The input data needs to be normalized in format. The first 8 packets of each data stream are intercepted. The first 100 bytes of each packet are taken. The length of less than 100 bytes is supplemented with 0x00 at the end.

4 Experimental

4.1 Experimental Environment

In this paper, the experimental environment is windows 10 system, the CPU is i7-6700, the main frequency is 3.7 GHz, the memory is 8 g, and the environment is Python 3.6. Use keras to build the model.

4.2 Data Sources

In this experiment, the flow in CICIDS2017 [9] data set is used as the normal flow; Malware traffic analysis [10] and stratosphere IPS [11] are merged as malicious traffic. The content distribution of the experimental data set is shown in Table 1.

Table 1. Distribution of malicious data samples

Traffic category	Quantity	Proportion	Traffic category	Quantity	Proportion
Dridex	20429	36.14%	Neris	218	0.39%
Vawtrak	19260	34.07%	Tofsee	232	0.41%
Miuref	6771	11.97%	Shifu	322	0.57%
Razy	1141	2.02%	Htbot	631	1.13%
Emotet	53	0.09%	Zeus	2032	3.59%
Reposfxg	84	0.15%	Normal	5352	9.47%

4.3 Evaluating Indicator

In order to evaluate the performance of the detection model proposed in this paper, accuracy, precision, recall and F1 value are selected as the evaluation indexes of the model. These calculation formulas are shown in (1)–(4):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\Pr ecision = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Re}call = \frac{TP}{TP + FN} \tag{3}$$

$$F_1 = \frac{2TP}{2TP + FN + FP} \tag{4}$$

Among them, TP means to correctly identify the encrypted traffic belonging to a certain classification as the classification, FP means to identify the encrypted traffic not belonging to a certain classification as the classification, TN means to identify the encrypted traffic not belonging to a certain classification as non classification, FN means to identify the encrypted traffic belonging to a certain classification as non classification.

4.4 Experimental Results and Analysis

In order to evaluate the rationality of CNN-BiLSTM-Attention model design, this paper conducts multi classification experiments with CNN, LSTM and other basic models to verify the generalization ability of the model. For the training model, early stopping strategy is used to dynamically control the number of training iterations. The data set is randomly divided into training set and test set, accounting for 80% and 20% respectively. The experiment was conducted by using categorical_ Cross entropy function is used as the loss function and Adam as the optimizer. The evaluation indexes of each model on the open data set are shown in Table 2. Figure 4 shows the 12 dimensional confusion matrix output by the four models after classification on the test set. The abscissa represents the

Table 2. Evaluation indexes of each mode

Model	Accuracy	Precision	Recall	F1
LSTM	97.40	75.92	72.67	73.33
BiLSTM	98.34	78.91	79.33	78.83
CNN	99.41	88.50	85.83	85.91
CNN-BiLSTM-A	99.65	94.67	89.66	91.08

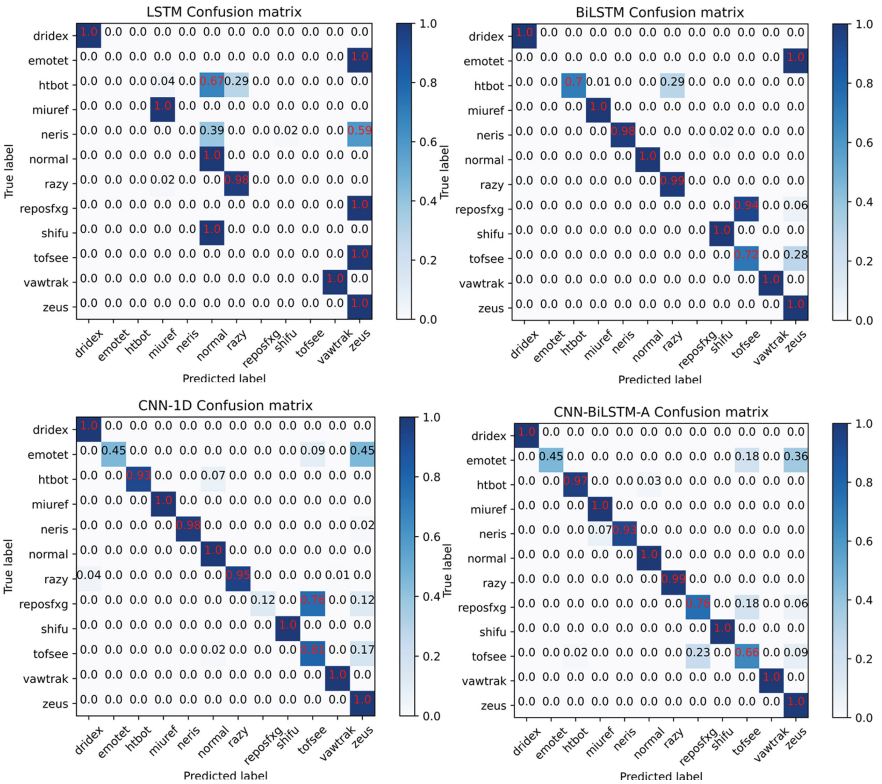

Fig. 4. Confusion matrix of prediction results of each model test set

prediction label, the ordinate represents the real label, and the value on the main diagonal of the confusion matrix is the correct proportion of each category predicted.

It can be seen from Table 2 that the accuracy of the model based on spatial features is higher than that based on temporal features when using single dimension features, which indicates that the spatial features used in this paper can better reflect the characteristics of malicious encrypted traffic compared with temporal features. In terms of temporal characteristics, the recognition effect of BiLSTM is better than that of LSTM.

After several training iterations, the performance of CNN-BiLSTM-Attention is improved compared with that of single feature. In addition, it can be seen from Table 1 that the malicious encrypted traffic data samples obtained in this paper are unbalanced. According to the confusion matrix in Fig. 4, the base model mistakenly discriminates small sample data into large sample data during training, which affects the overall accuracy of the model. In the case of sample imbalance, the F1 value of the proposed model reaches 91.08%, which is 5%–20% higher than other base models. This shows that the proposed model can also get better recognition rate in the case of unbalanced samples.

5 Conclusion

In order to improve the recognition effect of malicious encrypted traffic without decryption, this paper proposes CNN-BiLSTM-Attention model based on temporal and spatial characteristics of malicious encrypted traffic. The model makes full use of the hierarchical structure and temporal dynamic characteristics of traffic. BiLSTM-Attention model is introduced in the packet layer to extract the temporal features of session traffic, and CNN model is used to extract the spatial features of session traffic, and the two features are combined as the input of recognition. The experimental results show that the method has obvious improvement in F1 value, recall rate and so on, and improves the recognition rate of small sample data to a certain extent, and effectively strengthens the recognition effect of malicious encrypted traffic.

Acknowledgments. This project was supported in part by the Open Research Fund of Key Laboratory of Data Security, Xinjiang Normal University, under Grant XJNUSY102017B04 and University Scientific Research Project, Xinjiang Autonomous Region under Grant XJEDU2017S032.

References

1. Cisco. Encrypted Traffic Analytics White Paper [EB/OL], 31 December 2018. https://www.cisco.com/c/dam/en/us/solutions/collateral/enterprise-networks/enterprise-network-security/nb-09-encryted-traf-anlytcs-wp-cte-en.pdf
2. Chen, L., Gao, S., Liu, B., et al.: Research status and development trends on network encrypted traffic identification. Netinfo Secur. **19**(3), 19–25 (2019)
3. Rezaei, S., Liu, X.: Deep learning for encrypted traffic classification: an overview. IEEE Commun. Mag. **57**(5), 76–81 (2019)
4. Meng, P., Zhou, G.P., Meng, J.: Fast identification of encrypted traffic via large-scale sparse screening. In: Proceedings of International Conference on Advanced Cloud & Big Data, pp. 273–278. IEEE Press, Washington D.C. (2017)
5. Okada, Y., Ata, S., Nakamura, N., et al.: Comparisons of machine learning algorithms for application identification of encrypted traffic. In: Proceedings of International Conference on Machine Learning and Applications and Workshops, pp. 358–361. IEEE Press, Washington, D.C. (2011)
6. Callado, A.C., Kamienski, C.A., Szabo, G., et al.: A survey on internet traffic identification. IEEE Commun. Surv. Tutor. **11**(3), 52 (2009)

7. Wang, W., Zeng, X., Ye, X., et al.: Malware traffic classification using convolutional neural networks for representation learning. In: The 31st International Conference on Information Networking (ICOIN), pp. 712–717 (2017)
8. Cheng, H., Chen, L., Xie, L.: CNN-based encrypted C&C communication traffic identification method. Comput. Eng. **45**(8), 31–34, 41 (2019)
9. University of New Brunswick. ICIDS2017 [EB/OL] (2017). http://www.unb.ca/cic/datasets/ids-2017.html
10. BradDuncan. Malware-traffic-analysis [EB/OL] (2019). https://www.malware-traffic-analysis.net
11. Straosphere Lab. Malware Capture Facility Project [EB/OL] (2019). https://www.stratosphereips.org/datasets-malware