TOR Anonymous Traffic Fingerprint Extraction and Recognition Based on Neural Network

TaoPu*, JuanWang
*School of Cyberspace Security, Chengdu University of Information Technology

*891952873@qq.com

Abstract. The TOR anonymous communication system is an important means to protect network communication security and user privacy, but there are still criminals trying to destroy the confidentiality of the anonymous communication system through some special methods. Aiming at the abuse of the TOR anonymous communication system, this paper proposes a neural network-based anonymous traffic identification method, which uses a one-dimensional convolutional neural network for feature extraction, prediction and classification, and finally integration. In the experiment, nearly 100 websites were selected for the flow feature extraction and recognition based on one-dimensional convolutional neural network. The recognition accuracy rate is 87.5%, indicating that this method can effectively fingerprint TOR anonymous communication.

Keywords: Anonymous communication; TOR network traffic; Neural Networks; Feature extraction; Flow identification

1. Introduction
Nowadays, various countries in the world have regarded network information security as an important development direction of the current era. With the rapid development of the Internet, network communication security has become a hot spot, and criminals have embezzled users' private information to do illegal things. People have to start looking for more reliable communication tools, and anonymous communication systems such as TOR (The Onion Router) and SSH (Secure Shell) have emerged. Due to SSH copyright issues, people use TOR more widely than SSH on the market. The TOR anonymous network was proposed by scholar Roger [1] and others. It is a link-based, low-latency anonymous network. Compared with other anonymous network communications, the TOR anonymous network has stronger availability and more convenient deployment and features such as sex and safety. It can provide users with efficient, reliable, flexible and secure services through simple deployment.

This paper proposes a new type of anonymous traffic identification method [2-3], which is a traffic analysis technology based on neural network self-learning. It uses passive traffic analysis and comparison to identify the real address information of the communication terminal of the regulated party. Aiming at the abuse of the TOR anonymous communication system, this paper proposes a neural network-based anonymous traffic identification method, which uses a one-dimensional convolutional neural network for feature extraction, prediction and classification, and finally integration. In the experiment, nearly 100 websites were selected for the flow feature extraction and recognition based on one-dimensional convolutional neural network. The recognition accuracy rate is 87.5%, indicating that this method can effectively fingerprint TOR anonymous communication.

Keywords: Anonymous communication; TOR network traffic; Neural Networks; Feature extraction; Flow identification

1. Introduction
Nowadays, various countries in the world have regarded network information security as an important development direction of the current era. With the rapid development of the Internet, network communication security has become a hot spot, and criminals have embezzled users' private information to do illegal things. People have to start looking for more reliable communication tools, and anonymous communication systems such as TOR (The Onion Router) and SSH (Secure Shell) have emerged. Due to SSH copyright issues, people use TOR more widely than SSH on the market. The TOR anonymous network was proposed by scholar Roger [1] and others. It is a link-based, low-latency anonymous network. Compared with other anonymous network communications, the TOR anonymous network has stronger availability and more convenient deployment and features such as sex and safety. It can provide users with efficient, reliable, flexible and secure services through simple deployment.

This paper proposes a new type of anonymous traffic identification method [2-3], which is a traffic analysis technology based on neural network self-learning. It uses passive traffic analysis and comparison to identify the real address information of the communication terminal of the regulated party. Aiming at the abuse of the TOR anonymous communication system, this paper proposes a neural network-based anonymous traffic identification method, which uses a one-dimensional convolutional neural network for feature extraction, prediction and classification, and finally integration. In the experiment, nearly 100 websites were selected for the flow feature extraction and recognition based on one-dimensional convolutional neural network. The recognition accuracy rate is 87.5%, indicating that this method can effectively fingerprint TOR anonymous communication.

Keywords: Anonymous communication; TOR network traffic; Neural Networks; Feature extraction; Flow identification
2. Related technology

2.1. One-Dimensional Convolutional Neural Network [4]

Convolutional Neural Network (CNN) is one of the representative algorithms of deep learning. It includes convolution calculations and has the ability to characterize learning. It is mainly used to process neural networks with data structured data. One-dimensional convolutional neural network is one of the network structures in convolutional neural network, which is usually used for time series analysis of data. The core algorithm is:

Suppose there is a one-dimensional sequence data of length \( p \) that can be expressed as:

\[
x_{1:p} = x_1, x_2, \cdots, x_p
\]

(1)

Among them, \( x_i \) is a data element. One-dimensional convolution operation uses a convolution kernel \( W \in \mathbb{R}^{h\times 1} \) of length as a sliding window, to extract a new feature \( c_i \) from a region \( x_{i:j} = (x_i, x_{i+1}, \cdots, x_{i+j}) \) of length \( j \) on \( x_{1:p} \). The formula is:

\[
c_i = f(w \cdot x_{i:j} + b)
\]

(2)

Among them, \( b \in \mathbb{R}^{1\times 1} \) is the bias term and \( f \) is the activation function. Commonly used activation functions include Relu, sigmoid and tanh. The feature map \( c \) obtained by a convolution kernel sliding on \( x_{1:p} \) is:

\[
c = [c_1, c_2, \cdots, c_{n-h+1}]
\]

(3)

In order to extract more high-dimensional data features, multiple different convolution kernels are usually used to extract features on \( x_{1:p} \), and all feature maps are superimposed and input into the next layer of neural network.

This article uses the characteristics of data processing in one-dimensional convolutional neural networks. Network traffic is essentially a series of data. These data can be regarded as discrete values in time series, and then a one-dimensional convolutional neural network is used to extract relevant features from the data to form an identification fingerprint for TOR traffic to identify TOR traffic.

3. TOR anonymous network feature extraction and recognition design

The CNN method is used to extract TOR anonymous traffic fingerprints and identify them. The core idea is to first preprocess the original traffic data, then convert it into a numerical format acceptable to the neural network, and finally use CNN for autonomous feature learning and implement fingerprint extraction and prediction recognition.

Algorithm flow:
- Step 1: data preprocessing, extract the flow and convert it into a numerical format;
- Step 2: Divide the data set into two parts: test set and training set;
- Step 3: Input the training set into the CNN model in batches for training, and then use the test set to evaluate the performance of the trained model.
3.1. Data Preprocessing

This paper extracts the start IP, destination IP, start port, destination port, and protocol type of the data packets in the captured traffic, and filters out the TCP stream encrypted by the upper layer using TLS as the research data stream.

The traffic is converted into a numerical form, and normalized, and the processed data set is input to the CNN network for feature extraction and classification prediction. In this paper, the network traffic is converted into numerical values in bytes, and each network flow in the data set is converted into a numerical sequence as the input of the neural network. The network flow can be expressed as:

$$FLOW = \frac{1}{255}[a_1, a_2, \ldots, a_k], [a_2, a_3, \ldots, a_k], [a_{k-1}, a_k, \ldots, a_k], a \in [0, 255]$$

(4)

Among them, k refers to the number of sessions, and m is the byte size of each session.

3.2. Extraction Model Construction

Convolutional neural network includes input layer, convolution layer, pooling layer, fully connected layer and output layer. This research will use convolutional layer and pooling layer to extract multi-dimensional features. Then the feature map is passed to the fully connected layer, and finally the output layer uses the softmax activation function to predict the classification of the input data. The specific process is as follows:
Convert the data stream into numerical form as the neural network input.

First enter the convolutional layer of the neural network. Cells are distributed in the convolutional layer. The convolution kernel is used to establish the mapping relationship between the input and the feature. The convolution operation can make the data learn the same features of the data input in different cells, forming a series of high-dimensional feature sets;

Then enter the pooling layer. The pooling process combines the features learned by the convolutional layer and combines them into higher-level features to reduce the overall number of features;

Transfer the pooled features to the fully connected layer, and the fully connected layer at the end of the neural network will perform the feature classification operation [5].

4. Experiment and Analysis
This article uses Pyshark to capture and parse the data packets of the device. According to the start IP, destination IP, start port, destination port and protocol type in the data packet, the upper layer protocol uses TLS encryption protocol to filter out the TCP stream to generate a .pcap file, as the experimental data set.

Divide the data in the data set into 30 groups, and the experimental conditions are as follows (70% of the training set, 30% of the test set)

4.1. The Impact of Test Time Interval on Model Recognition Accuracy
The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

Fig. 2 CNN feature extraction classification process
|   | Btrain | Btest | time interval | accuracy (%) |
|---|--------|-------|--------------|--------------|
| (1) | 1      | 1     | 0            | 95.23        |
| (2) | 1      | 1     | 1            | 90.2         |
| (3) | 1      | 1     | 29           | 30           |
| (4) | 4      | 4     | 0            | 91.5         |
| (5) | 4      | 4     | 24           | 51.13        |
| (6) | 6      | 6     | 0            | 88.03        |
| (7) | 6      | 6     | 24           | 56.62        |

**Tab.1** The Effect of Time Interval on Accuracy

It can be seen from the data in the table that when the training set and test set are constant, the larger the time interval, the lower the accuracy. It shows that the recognition accuracy of website fingerprints is greatly affected by the time factor.

4.2. The Impact of Training Data Set Size on Recognition Accuracy

It can be seen from the line graph that the larger the training set, the lower the rate of decline of the accuracy curve, indicating that the size of the training data set will also affect the performance of the model.

4.3. The Impact of Model Computational Complexity on Recognition Accuracy

Computational complexity: Let Btrain=4, Btest=4, Δ t=0, select 7 sets of data, and use three algorithms: Naive Bayes, Random Forest, and One-dimensional Convolutional Neural Network to train and evaluate the accuracy of the model.

The Naive Bayes method only uses the two-tuple of the packet length and the data direction as features. The Random Forest method adds the total number of packets, the total size of the site, and the required time on the basis of the Naive Bayes method. The classification of large and small intervals improves the accuracy rate. However, Random Forest method and Naive Bayes method both select fixed features and have certain limitations. The automatic feature extraction method used by CNN essentially solves the problem of feature selection. Therefore, the accuracy rate is improved. In fingerprint research, the result whose real category is in the top K of the prediction vector is usually recorded as the Top-K accuracy rate. This paper introduces the K = [1,5] accuracy rate as the evaluation index.
### Tab.2 The Recognition Accuracy of Different Algorithms

| Method     | Accuracy (%) | Top-5 Accuracy (%) |
|------------|--------------|---------------------|
| Naïve Bayes| 58.7         | 76.7                |
| Random Forest | 72.18     | 84.1                |
| CNN        | 92.03        | 98                  |

### Fig.4 Accuracy under different K values

It can be seen from the experimental results that the accuracy of using the one-dimensional convolutional neural network algorithm is the highest. Compared with Naïve Baye, Random Forest has a slightly higher accuracy. The analysis algorithm shows that the Naïve Bayes method only uses the two-tuple of the packet length and the data direction as features. The Random Forest method adds the total number of packets, the total size of the site, and the required time on the basis of the Naïve Bayes method. And the data packets are classified according to the size interval, so that the accuracy rate is improved to a certain extent. However, the Random Forest method and the Naïve Bayes method both select fixed features and have certain limitations. The automatic feature extraction method used by CNN essentially solves the problem of feature selection, so the accuracy rate is higher.

### 5. Conclusion

Aiming at the difficulty of monitoring the anonymous communication system abused by criminals, this paper proposes a TOR anonymous traffic feature extraction and identification method based on one-dimensional convolutional neural network. This method integrates feature extraction and prediction classification, and uses the CNN model to achieve fingerprint recognition for TOR anonymous traffic. It has achieved an accuracy of 87.5% on the experimental data set. It can effectively identify TOR anonymous traffic and anonymous communications. The network realizes supervision and review. This article also demonstrates the advancement and versatility of one-dimensional neural network methods by comparing Naïve Bayes, Random Forest and other methods. The experimental results show that the influence of the number of data sets on the recognition accuracy cannot be ignored. Therefore, a large number of data sets are needed to train the model. This is a challenge for research, and it is not easy to obtain anonymous network experiment data. How to capture more anonymous network traffic and use unsupervised deep learning algorithms for feature extraction and recognition is a direction worthy of further research. This article only does a preliminary exploration, and will further optimize the model construction and network parameters in the future.
Acknowledgments
The work is supported by Sichuan Science and Technology Program, No.2019YFG0425, No.2019YFG0186, No.2019YFG0408 and the open Project of Network and Data security Key Laboratory of Sichuan Province (No. NDS2018-1).

References
[1] Wang T, Goldberg I 2013 Improved website fingerprinting on TOR in: Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society( Chicago: ACM Press)pp201-210
[2] Guang Z Security analysis and research of Tor anonymous communication system( 2011 the academic degree's dissertation of Shanghai Jiaotong University)
[3] Carhartt 2005 Cryptography and Network Security(Beijing:Tsinghua University Press)
[4] Xin L Research on Anonymous Communication Based on Tor Network (2011 the academic degree's dissertation of East China Normal University)
[5] Zhang S, Deng Z, Cheng D et al 2016 Efficient kNN Classification Algorithm for Big Data J.Neurocomputing pp143-148