BFSN: A Novel Method of Encrypted Traffic Classification Based on Bidirectional Flow Sequence Network

Xinxin Tong[1], Xiaobin Tan[2], Lingan Chen[2], Jian Yang[2], Quan Zheng[2]
1.*Department of Cyberspace Security, University of Science and Technology of China*
2.*Department of Automation, School of Information Science and Technology, University of Science and Technology of China*
AnHui Province Key Laboratory of Future Networks, University of Science and Technology of China, Hefei, China
Email: txx1204@mail.ustc.edu.cn, xbtan@ustc.edu.cn, cla@mail.ustc.mail.edu.cn, jianyang@ustc.edu.cn, qzheng@ustc.edu.cn

***Abstract*—With the rapid development of network technology and encryption technology, network security issues have received more and more attention, and network encryption traffic is increasing, which results in a huge challenge for network traffic classification. Combining machine learning algorithms with manual design has become a mainstream approach to solve this problem, However, it requires a large amount of human effort to extract and process features, which depend on professional experience heavily. In this paper, We discuss the essential reason why convolutional neural network(CNN) can deal with the problem of encrypted traffic classification and propose a novel classification framework the Bidirectional Flow Sequence Network(BFSN) based on long short-term memory (LSTM). Compared with the traditional traffic classification scheme, the BFSN is an end-to-end classification model that learns representative features from the raw traffic and classifies them. Moreover, We apply the length and direction information of the encrypted traffic to construct the bidirectional traffic sequence and then process it based on LSTM. Our Experiments gains the excellent accuracy about 91% based the ISCX VPN-NonVPN dataset.**

***Index Terms*—Encrypted Traffic Classification, Bidirectional Flow Sequence Network, Long Short-Term Memory**

I. Introduction

ALONG with the rapid development of Internet technology, the Internet has been playing an important role in public life and bringing great convenience to network users. Thereby, an accurate traffic classification becomes the basic task of crucial importance so as to more effectively enhance the level of network management, improve service quality and guarantee the network security [1]– [3]. Whereas, the network traffic is increasingly expanding with advent of various new network application. Thus, network traffic classification becomes more and more challenging owing to the proposed facts [4].

up to now, many researchers have been proposed various methods for traffic classification. However due to the wide use of cryptographic protocol, network traffic payload characteristics and its statistical characteristics have obvious change, so that it no longer has the characteristics of the fixed field, which results in a significant reduction in traffic visibility, thus the traditional network traffic identification methods, such as port-based classification [5], which only needs to obtain the first packet of traffic can realize traffic identification, deep packet inspection(DPI) [6], which conducts the traffic classification by regular expression for matching the packet payload data. However, due to the hiding of port information in the case of encryption and the occurrence of dynamic port, the DPI-based methods could only identify the traffic that its type was recoded in the expression library, but it is difficult to obtain the expression of the encrypted payload. Although due to the emergence of machine learning, encrypted traffic identification has a certain progress [7]– [11], whereas, its good performance is highly dependent on the excellent feature extraction and the classification algorithm selection, which are complicated and time-consuming.

Recently, deep learning has rapidly developed and has witnessed its great success in a variety of areas, such as speech recognition, natural language processing, etc. Meanwhile, deep learning methods have been widely used in the scenario of communication networks. Network traffic classification can be regarded as a common classification problem in the field of machine learning. Therefore, this paper proposes an encryption traffic identification method based on the discussion about deep learning algorithms, where two deep learning algorithms-Convolutional Neural Network(CNN) and Long Short Term Memory(LSTM)–are employed. The CNN is used to adopting automatically for flows feature extraction to deal with the task of traffic classification, which has emerged in many existing studies. However, in this paper, we mainly discuss and reveal the essential reason why CNN can classify the encrypted traffic, and then construct the bidirectional traffic sequence for each flow based on above discussion, besides, the different sequence with packet length and direction of flow are handled by the LSTM to improve traffic classification performance.

This paper's contributions can be summarized as follows:

- We propose a novel DL-based method which can establish a framework for classify multiple kinds of encrypted traffic, where the LSTM is employed. The presented method takes advantage of supervised learning, builds a more more efficient and accurate framework owing to the

Identify applicable funding agency here. If none, delete this.

dataset.

- we discuss and reveal the essential reason why convolutional neural network can deal with the problem of encrypted traffic classification through our comprehensive experiments.
- Based on the previous discussion, we take into consideration the length and direction information of encrypted traffic and propose the BFSN. Besides, applying the LSTM to process the bidirectional traffic sequence and obtain the results of classification in that feature vectors associated with continuous packets in a given traffic exhibit correlated time behavior similar to that of the NLP.

The structure of this paper is as follows. Section II describes some of the related work of network traffic classification. Section III details the methodology of the proposed method based on LSTM. In section IV and V, we discuss several models based on CNN and describe the network traffic classification model BFSN we presented in this paper, respectively. Finally, section VI gives a conclusion of this this article.

II. Related Work

A. Traditional Traffic Classification

Port-based method is the most traditional traffic identification that relies on the port, which use the information in the TCP or UDP headers of the packets to exact the port number that is assumed to be associated with a particular application, and then compared them with the assigned IANA TCP or UDP port numbers for traffic classification. Of course, the extraction procedure and the identification are so simple and fast. whereas, the method is failed in the situation with the port obfuscation, port forwarding, protocol embedding and the dynamic port [5], which resulted in significant reduction of the identification accuracy. DPI analyzes the whole packets and use predefined patterns like the expressions as the signature for every protocol and application [6], but the patterns need to update whenever a new protocol released that is more difficult for the encrypted traffic and privacy issues are among the drawbacks of the method.

B. Machine Learning Traffic Classification

With the goal of classifying the encrypted traffic, many of the current researchers mainly adopt the machine learning approach, such as Decision Tree, Random Forest [8]– [9] and SVM [7], etc. For example, Auld et al. [11] proposed a Bayesian neural network, which is trained to classify the traffic with P2P protocol such as the bittorrent and kazaa and the accuracy was up to 99%. Ding et al. [8] first filtered the HTTPS traffic with a machine learning algorithm C4.5 decision tree, then classify the services with random forest. Liu et al. [10] preferred to packet-level statistical features which include maximum and mean of packets of the sessions and built a composite feature-based semi-supervised method for encrypted traffic identification. However, their efficiency extremely rely on manually selected features, rich prior-experiences and profession knowledge, where some features may come to own privacy. Besides, owing to taking statistic features into account, i.e., the features of flow and packet(e.g. flow bytes per second and packet size).

C. Deep learning Traffic Classification

Therefore, Deep learning has been implemented to the encrypted traffic in that it can automatically select underlying features from the raw traffic. What's more, there is a more stronger learning capability for methods based on deep learning, so that they make it easier to deal with heavy network tasks with multiple layers of neural networks. In addition, they require fewer computation resources but can achieve a higher accuracy than traditional methods. Most methods adopted convolution neural network(CNN) or long short-term memory(LSTM) or spare auto encoder(SAE) [12]–[18]. Wang et al. [12] proposed an end-to-end encrypted traffic classification method based on the 1dCNN for encrypted traffic service classification, but 1dCNN requires features have no relation to their location while we think the order of the packets and its location of bytes content should be taken into account. Zeng et al. [17] applied the combination of CNN and LSTM to directly deal with the first 784 bytes data of the flow. Although the results seems ideal, the data input into LSTM does not actually have time characteristics. Lotfollahi et al. [14] extracted the first 1500 bytes of only one packet and took them as the inputs of CNN. They train CNN and Stack Auto Encoder (SAE) to classify encrypted traffic. They both perform well in the classification problem, but they ignore the time sequence features hidden in a flow. Liu et al. [18] proposed the FS-Net framework and applied the bi-GRU to process the sequence feature of encrypted traffic.

In summary, traditional methods proposed above may not deal with the classification problem of encrypted network traffic. As for the machine learning methods, They considered the whole flow of statistical information, they depend on professional experience heavily, which require a large amount of human effort to extract and process features. Therefore, more and more researchers apply the deep learning that can automatically extract the features from the flow and process them efficiently to solve the classification of encrypted traffic. However, these hybrid methods above involve the payload or some header information of the traffic packet, in the era of traffic explosion, handling packets alone may be costly, privacy policies and laws prohibit accessing, which limits the use of payload features [19], More importantly, in some cases, although convolutional neural network can solve the problem of encrypted traffic identification to a certain extent, the essential reason of its identification has not been discussed in previous studies. In addition, the direction information of traffic is usually ignored, but it should be as one of the important characters of traffic identification. Thereby, we proposed the BFSN and apply LSTM that can effectively process the information based on the time to deal with the problem of classifying encrypted traffic.

III. METHODOLOGY

In the real network, many hidden security problems of plaintext traffic make network service providers start to apply various network encryption protocols, which bring some challenges to traffic identification. This chapter briefly introduces the interaction process of network encryption protocols and the deep learning method.

A. Encryption Protocol

The interaction process of encryption protocol generally includes two processes: first, the establishment of a secure connection, that is, the encryption algorithm supported by the communication parties through negotiation.The second phase is mainly used for the first phase of the secret key encryption to transmit content, implementation of normal communication [10]. such as SSH secure shell protocol, the communication between the two parties establishes encrypted channel to keep the transmission of data from eavesdropping, and uses the secret key exchange algorithm to ensure the safety of the secret key itself HTTPS encryption protocol. The main are to create an encrypted security channel to guarantee safe information transmission and realize the server certification that is the HTTP protocol with SSL protocol or a combination of TLS protocol can be verified.

B. LSTM

As a special RNN, LSTM has the feature that the output of the neuron can directly act on itself at the next moment. In this paper, LSTM is selected mainly because it adds the following key components in the neuron, so as to eliminate the influence caused by the disappearance of the gradient.

(1)forget gate: The first step in the LSTM network is to determine the information to be discarded from the cell state C_{t-1}, which is provided a parameter value between 0 and 1 by the forget gate layer, 0 indicates complete discard of information, and 1 indicates complete reservation of information. Besides, The forget gate not only reads the input layer x_t, but considers the information of the previous output layer h_{t-1} to obtain the value and operate to the cell state C_{t-1} through the activation function.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

(2)input gate: The process of input gate is similarity to those operations above of forget gate, which is mainly to filter the information of input layer x_t.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

(3)input node: How much information should be selected to add the cell state is the main task of the step, which considers the information of the previous output layer h_{t-1} and the input layer x_t to get a value between 0 and 1 through the $tanh$ function, the purpose of the $tanh$ is to map content between -1 and 1, so as to generate the candidate value of input layer.

$$y_t = \tanh(W_y \cdot [h_{t-1}, x_t] + b_y) \tag{3}$$

(4)update cell state: The step is to update the original cell state C_{t-1} by the corresponding calculation of the results obtained above.

$$C_t = f_t * C_{t-1} + i_t * y_t \tag{4}$$

(5)output gate: It is necessary to determine the state of the cell, so that we can obtain the final output value. Of course, First, Considering the input and the previous output layer to generate a filter value to determine which parts of the cell state should to act on the output. Then, the cell state needs to be mapped through the $tanh$ function, and the result obtained after multiplying the filter value.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * tanh(C_t) \tag{6}$$

C. Dataset

In this paper, the public dataset published by the University of New Brunswick is selected in the evaluation of our proposed model: ISCX VPN-nonVPN traffic dataset [20]. The dataset contains 25GB raw traffic in the pcap format, which includes 14 network application classes. We relabel the raw traffic into 6 classes from the VPN encrypted traffic to be used in our experiments, according to the work [12]. The classes of the datasets applied to our model are listed in Table I in details. The dataset is split into the flow level according to the five tuples: source IP, destination IP, source port, destination port, protocol. We select randomly split the dataset into 90% training dataset and 10% testing dataset, respectively.

TABLE I
DATASET TYPE TABLE

Type	*the number of flows*	*the number of packets*
Vpn_Chat	4029	86117
Vpn_Email	298	21558
Vpn_File	1020	376703
$Vpn\_Streaming$	349	528299
Vpn_P2P	477	422074
Vpn_VoIP	1618	723569

IV. DISCUSSION ABOUT MODELS BASED ON CNN

Although convolutional neural network can solve the problem of encrypted traffic identification to a certain extent, the essential reason of its identification has not been discussed in previous studies. Therefore, we proposed different model graphs and discuss the reason with comprehensive experiments based on the CNN. Fig. 1 illustrates the workflow of our encrypted traffic classification.

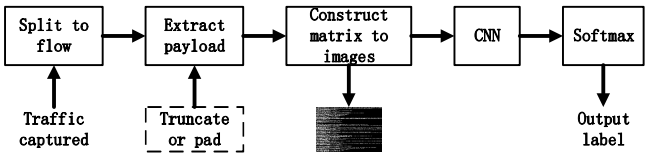

Fig. 1. The architecture of the CNN model

A. Graph of Flow

The packets that have the same five tuples belong to a flow, and we make use of the packet payload bytes as the inputs of neural networks. Different bytes lengths may appear in different packets, while the size of the inputs of CNN is necessarily constant. Thereby, in our approach, the number of the packets and the length of the payload byte sequence of any packet equals to N and M. Besides, the data packet header is wiped off in that it is stuffed by the some addresses information which contain useless features for traffic classification. Additionally, we will truncate if the number of packets exceeds N or the payload length exceeds M, and pad zero if less.

After the above operation, we obtained an $N*M$ two-dimensional matrix for each flow. As is known, one byte can be transformed into an integer in the range of [0,255], so a byte can be viewed as a pixel. Correspondingly, the matrix is constructed as an image with M-pixel width and N-pixel height. At last, we normalize the scale to [0,1] by dividing the bytes by 255.

B. Different Model Graphs of Flow

Firstly, we only consider the packets that contain payload, and extract the first 32 packets(containing payload) and the first 128 payload bytes to construct the graph of Type-1. As we know, this graph model appeared the previous research [12], but it ignores the actual sequence relationship of packets, that is, the existence of empty packets. Hence, the second kind of graph Type-2 arises, in this model, we extract the first 32 packets and the first 128 payload bytes to obtain the graph. Compared to the first, it takes less time to extract the feature information from the traffic, and can be faster to identify the categories of encrypted traffic, which can achieve the online classification of encrypted traffic. Moreover, We try to increase the number of packets to 128 to improve the accuracy, i.e., the Type-3.

Through the above model, we can explore whether the empty packets in the flow provide useful features for encrypted traffic identification, and lay a foundation for the future use of the direction as the identification features.

In order to explore whether the payload of the encrypted traffic has a real effect on the traffic identification, We treated the three model diagrams above, and overlaid the payload content with 255 after extracting the payload, that is, all the diagrams were made up of 255 or 0. Therefore, for the first type which is a black and white picture. However, because different packets convey bytes with different lengths, they have different segmentation profiles. For the second and third types model graph, there is another feature that the black and white lines appear at intervals(due to the existence of empty packets).

C. Detailed model architecture based on CNN

The convolutional neural network consists of two convolutional layers ($Con-1$ and $Con-2$) and a fully connect layer (Fc). For the design of convolutional layer, $Con-1$ firstly filters the 32(or 128)*128 graph with 32 (128)kernels, whose size is 5*5 with 1 stride. Relu [21] is employed as the activation function. After a 2*2 max-pooling layer, we take the output to $Con-2$, whose structure is same with the $Con-1$, and then the output of $Con-2$ is sent to the Fc layer. Finally, we can obtain a flow vector with 6 dimensions. the design of CNN is indicated in Table II (where N equals to 32).

TABLE II
DESIGN TABLE OF CNN

Layer	*Name*	*Input*	*Filter*	*Output*
$Cno-1$	$convolutional$	$32*128$	$5*5$	$32*128*32$
	$maxpooling$	$32*128*32$	$2*2$	$16*64*32$
$Cno-2$	$convolutional$	$16*64*32$	$5*5$	$16*64*128$
	$maxpooling$	$16*64*128$	$2*2$	$8*32*128$
Fc	$Dense$	$8*32*128$	$--$	6

D. Experiments and Analysis

In general, accuracy is used to evaluate the overall performance of a classifier.

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{7}$$

where TP represents the number of items correctly classified as P, TN is the number of items correctly classified as belonging to Not-P, FN is the number of items incorrectly classified as belonging to Not-P, and FP is the number of items incorrectly classified as P.

The results of experiments is showed in Table III. According to the Type-1 and Type-2, the accuracy of Type-2 is similar to that of Type-1, but Type-2 takes into account the empty packets in the flow, so that it can identify the type of the encrypted traffic in less time. Although Type-3 increases the number of packets to 128, it does not improve the accuracy in that many flows have fewer packets than this value. Additionally, we found that the accuracy was also similar to that before padding payload content with 255 for each type. Therefore, we can conclude that the input really processed by the convolutional neural network is not the encrypted payload content, but the contour features constructed by different packet lengths.

TABLE III
RESULTS OF EXPERIMENTS

Type	*Size*	*Empty packets*	*Pad 255*	*Accuracy*
$Type-1$	$32*128$	No	No	71%
			Yes	70%
$Type-2$	$32*128$	Yes	No	71%
			Yes	71%
$Type-3$	$128*128$	Yes	No	70%
			Yes	71%

V. SCHEME OF BFSN

To solve the problem of classifying the encrypted traffic, some researchers made use of the payload based on CNN [12] [16], whereas, it is limited the use of the payload features due to the privacy policies and laws prohibit accessing. More importantly, according to the results on encrypted traffic identification in the previous discussion of models based on

CNN, the encrypted payload data of flow does not play a role in the identification, but the contour information formed by the load length of different flows has some effect in the process. Therefore, we consider length sequence information of packets as one of the features,

In addition, Owing to the difference of the interactions information of different types of traffic and the results on encrypted traffic identification in the previous discussion on the empty packets, we believe that the directional information of packet should be as the feature.

A. Design of BFSN

As is showed in Fig. 2, we proposed the BFSN based on the analysis above, i.e., we first split the encrypted traffic according to the information of five tuples, the temporal and directional correlation features that contained in the back and forth packets belonging to a flow between the source node and destination node, also attach important information for traffic classification. Then extracting the length sequence information of the first N packets and directional information for each packet, where the directional feature is set to -1 or 1. Besides, in order to verify the validity of directional information of the packets, we make the experiment FS-Net [18] that only considers the sequence information of the first N packets.

Due to recurrent neural networks have shown a considerable power for exploiting sequential tasks [22]. In this paper, considering the effectiveness of LSTM in processing the time series, we apply LSTM to process the extracted feature sequence. We forward the 2-dimension (i.e., the length and direction information) packet feature vectors to the LSTM cells with the time sequence T=N.

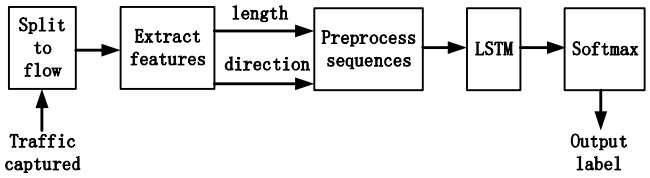

Fig. 2. The architecture of the BFSN model

In our experiments, the design of LSTM is showed in Table IV, and the N equals to 32. Besides, the sequence information is sent to the embedding layer that learns from word embedding in natural language processing [23], and take the 2-layer hidden layer, where the number of the hidden units is set to 64, in LSTM network. Furthermore, we take dropout [24]with 0.5 ratio to in the LSTM to avoid over-fitting problem and the Adam optimizer [25] with learning rate 0.001 is applied. Finally, the outputs of the last hidden layer are fed into a 6-way softmax which produces a distribution over the 6 class labels.

B. Experiments and Analysis

First, we make use of the accuracy to evaluate the overall performance of model, and the results is showed in Table V.

TABLE IV
DESIGN TABLE OF LSTM

Layer	*Input*	*Output*	*dropout*
Embedding	$2 * 32$	$128 * 32$	$--$
LSTM	$128 * 32$	$128 * 32$	0.5
Softmax	$128 * 32$	6	$--$

TABLE V
RESULTS OF EXPERIMENTS

Type	*Length sequence*	*Directional sequence*	*Accuracy*
BFSN	*Yes*	*Yes*	91%
FS-Net	*Yes*	*No*	81%

Apart from the accuracy metric, there are two other common and widely-used metrics: precision and recall, where TP, FP and FN have been described in the section of previous experiments above. These metrics are formulated as follows:

$$precision = \frac{TP}{TP + FP} \tag{8}$$

$$recall = \frac{TP}{TP + FN} \tag{9}$$

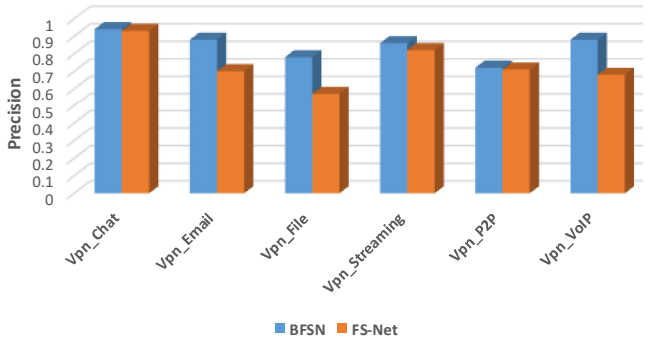

Fig. 3. The Precision comparison of BFSN and FS-Net

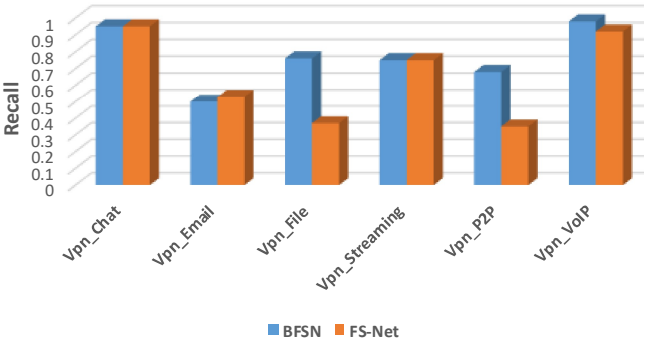

Fig. 4. The Recall comparison of BFSN and FS-Net

The overall comparison results between our model and FS-Net are exhibited in Fig. 3 and Fig. 4. A first glance of the figure shapes the performance advantages of our model in

contrast with FS-Net in different types of encrypted traffic. As Fig. 3 shows, the precision of all types in BFSN is better than that in FS-Net, and the performance of Vpn_File is more than 20% higher than that of FS-Net, in particular. Fig. 4 demonstrates that our model hits the higher values in recall for most of the types of traffic than FS-Net. Thereby, the fact beneath the results suggests the effectiveness of the flow direction sequence features in our model.

VI. Conclusion

In this paper, we construct different graphs of the payload that whether we consider empty packets and apply 255 to cover the encrypted payload, to explore the essential reason why convolutional neural network can deal with the problem of encrypted traffic classification. We come to the conclusion that the input really processed by the convolutional neural network is not the encrypted payload content, but the contour features constructed by different packet lengths. Therefore, we take into consideration the length and direction information of encrypted traffic based on the previous analysis and propose the BFSN, i.e, Creating length and direction feature sequences of consecutive packets in given flow. Besides, we apply the LSTM to process the bidirectional flow sequence in that features extracted associated with consecutive packets in a given traffic exhibit correlated time behavior. Experiment results have shown that our model outperforms the existing state-of-the-art model based on CNN or LSTM in terms of higher effectiveness.

Acknowledgment

This work is supported by the National Key R&D Program of China under grant No. 2018YFF01012200, the National Science Foundation of China under grant No. 61673360 and the CETC Joint Advanced Research Foundation under grant No. 6141B08080101.

References

[1] H. Shi, H. Li, D. Zhang, C. Cheng and X Cao, "An efficient feature generation approach based on deep learning and feature selection techniques for traffic classification," Computer Networks, pp. 81–98, 2018.

[2] V. F. Taylor, R. Spolaor, M. Conti and I. Martinovic, "AppScanner: Automatic Fingerprinting of Smartphone Apps from Encrypted Network Traffic," 2016 IEEE European Symposium on Security and Privacy, Saarbrucken, 2016, pp. 439–454.

[3] M. Conti, L. V. Mancini, R. Spolaor and N. V. Verde, "Analyzing Android Encrypted Network Traffic to Identify User Actions," in IEEE Transactions on Information Forensics and Security, pp. 114-125, Jan. 2016.

[4] F. Pacheco, E. Exposito, M. Gineste, C. Baudoin and J. Aguilar, "Towards the Deployment of Machine Learning Solutions in Network Traffic Classification: A Systematic Survey," in IEEE Communications Surveys Tutorials, pp. 1988–2014, Secondquarter 2019.

[5] A. Dainotti, A. Pescape and K. C. Claffy, "Issues and future directions in traffic classification," in IEEE Network, pp. 35-40, January-February 2012.

[6] G. Aceto, A. Dainotti, W. de Donato and A. Pescape, "PortLoad: Taking the Best of Two Worlds in Traffic Classification,"2010 INFOCOM IEEE Conference on Computer Communications Workshops, San Diego, CA, 2010, pp. 1-5.

[7] A. Saber, B. Fergani and M. Abbas, "Encrypted Traffic Classification: Combining Over-and Under-Sampling through a PCA-SVM,"2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS), Tebessa, 2018, pp. 1-5.

[8] R. Ding and W. Li, "A hybrid method for service identification of SSL/TLS encrypted traffic,"2016 2nd IEEE International Conference on Computer and Communications, Chengdu, 2016, pp. 250-253.

[9] B. Yamansavascilar, M. A. Guvensan, A. G. Yavuz and M. E. Karsligil, "Application identification via network traffic classification,"2017 International Conference on Computing, Networking and Communications, Santa Clara, CA, 2017, pp. 843-848.

[10] H. Liu, Z. Wang, Y. Wang, "Semi-supervised Encrypted Traffic Classification Using Composite Features Set,"Journal of Networks, vol.7, pp. 1195–1200, 2012.

[11] T. Auld, A. W. Moore and S. F. Gull "Bayesian Neural Networks for Internet Traffic Classification," in IEEE Transactions on Neural Networks, vol. 18, pp. 223-239, Jan. 2007.

[12] W. Wang, M. Zhu, J. Wang, X. Zeng and Z. Yang, "End-to-end encrypted traffic classification with one-dimensional convolution neural networks," 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, 2017, pp. 43-48.

[13] G. Aceto, D. Ciuonzo, A. Montieri and A. Pescap, "Mobile Encrypted Traffic Classification Using Deep Learning," 2018 Network Traffic Measurement and Analysis Conference (TMA), Vienna, 2018, pp. 1-8.

[14] M. Lotfollahi, R. Shirali, M. J. Siavoshani and M. Saberian,"Deep Packet: A Novel Approach For Encrypted Traffic Classification Using Deep Learning,"arXiv preprint arXiv:1709.02656, 2017.

[15] A. Pradhan, S. Behera and R. Dash, "Hybrid RBFN Based Encrypted SSH Traffic Classification," Signal Processing and Integrated Networks 2018 5th International Conference on, pp. 264-269, 2018.

[16] Y. Zhang, S. Zhao, J. Zhang, X. Ma and F. Huang, "STNN: A Novel TLS/SSL Encrypted Traffic Classification System Based on Stereo Transform Neural Network," 2019 IEEE 25th International Conference on Parallel and Distributed Systems, Tianjin, China, 2019, pp. 907-910.

[17] Y. Zeng, Z. Qi, W. Chen, Y. Huang, X. Zheng and H. Qiu, "TEST: an End-to-End Network Traffic Examination and Identification Framework Based on Spatio-Temporal Features Extraction,"arXiv:1908.10271, 2019.

[18] C. Liu, L. He, G. Xiong, Z. Cao and Z. Li, "FS-Net: A Flow Sequence Network For Encrypted Traffic Classification," IEEE INFOCOM 2019 - IEEE Conference on Computer Communications, Paris, France, 2019, pp. 1171-1179.

[19] S. Rezaei and X. Liu, "Deep Learning for Encrypted Traffic Classification: An Overview," in IEEE Communications Magazine, pp. 76-81, May 2019.

[20] ISCX Vpn-nonVpn dataset, http://www.unb.ca/cic/datasets/vpn.html.

[21] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," International Conference on International Conference on Machine Learning. Omnipress, Madison, USA, 2010, pp. 807C814.

[22] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko et al., "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," in IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 677-691, April 2017.

[23] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality,"in Advances in neural information processing systems, 2013, pp. 3111-3119.

[24] G. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," CoRR, abs/1207.0580, 2012.

[25] Z. Chang, Y. Zhang and W. Chen, "Effective Adam-Optimized LSTM Neural Network for Electricity Price Forecasting," International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2018, pp. 245-248.