ENCRYPTED 5G OVER-THE-TOP VOICE TRAFFIC CLASSIFICATION USING DEEP LEARNING

Zhuang Qiao[1,2], Shunliang Zhang[1,2], Liuqun Zhai[1,2], Xiaohui Zhang[1,2]
[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, 100093, China, [2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, 100049, China

NOTE: Corresponding author: Shunliang Zhang, zhangshunliang@iie.ac.cn

Abstract – *With the commercialization of fifth-generation (5G), the rapid popularity of mobile Over-The-Top (OTT) voice applications brings huge impacts on the traditional telecommunication voice call services. Tunnel encryption technology such as Virtual Private Networks (VPNs) allow OTT users to escape the supervision of network operators easily, which may cause potential security risks to cyberspace. To monitor harmful OTT applications in the context of 5G, it is critical to identify encrypted OTT voice traffic. However, there is no comprehensive study on typical OTT voice traffic identification. This paper mainly focuses on analyzing OTT voice traffic in the 5G network specifically. We propose employing Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs) to identify encrypted 5G OTT voice traffic, study the identification performance of used deep learning methods in three different scenarios. To verify the performance of the proposed approach, we collect 28 types of typical OTT and non-OTT voice traffic from the experimental 5G network. Experimental results prove the effectiveness and robustness of the proposed approach in encrypted 5G OTT voice traffic identification.*

Keywords – 5G OTT, classification, deep learning, encrypted traffic, voice traffic

1. INTRODUCTION

In recent years, most Internet enterprises around the world have devoted themselves to providing OTT services, which refer to various services developed by these companies based on the Internet, including voice, video, and text message services. This paper mainly focuses on OTT voice applications and the various OTT voice traffic generated by them. These OTT voice applications typically use various encryption protocols for data transfer. 5G is characterized by low latency, high reliability, and large connections [1] and provides the possibility for high-quality OTT voice and video calls anytime and anywhere on most platforms, fundamentally changing the consumer experience. Facing the 5G era, OTT voice applications have a broad prospect, and their types, quantities and traffic are bound to grow dramatically. OTT voice applications commonly have features with data encryption, deployment flexibility and are not controlled by the operator. Under the condition of the OTT service provider which does not open data monitoring interfaces, the traffic is difficult to be effectively regulated, leading to OTT voice applications that may illegally utilized for fraud and malicious information diffusion. Due to the increasing awareness of network security and privacy protection, more and more smartphone users begin to access the Internet through VPN.

The rapid growth of encrypted mobile traffic brings many challenges to the security of the 5G wireless network [2]. In non-cooperative scenarios, OTT voice applications may be utilized to cause the following security risks:

- Spread false and harmful information. Criminals may use end-to-end encrypted OTT voice services provided by QQ, WeChat, Skype, WhatsApp, etc., to evade the supervision of traditional telecommunication network voice services. Thus spreading harmful information that endangers national stability and unity.
- Engagement in drug trafficking, money laundering and terrorist activities, where encrypted voice applications can provide customers with anonymous and free encrypted services.
- Network fraud; the use of artificial intelligence technology for voice synthesis, face change, with forged voice or video contact with the victim, causing severe deception.
- Attack the network by forging the traffic on the real network, posing threats to network security.

Javier et. al [3] investigated potential security risks, taking into consideration its functionality layers and the operational requirements in order to achieve a more complete and useful classification.
For network operators, a clear understanding of traffic in the network is indispensable to manage the network safely and effectively [4]. Understanding traffic management systems [5] should be taken as a basis for analyzing 5G uses cases and their requirements.
The monitoring and analysis of network traffic have always been hot topics to researchers. The rule-based classification methods support searching for specific strings or regular expressions in the packets. Papadogiannaki et al. [6] used rule-based language expressions to identify OTT application events, such as sending messages, audio, and video calls through encrypted traffic. But they used

only the metadata of the data packet and ignored the information provided by the payload. Meanwhile, detecting the content of data packets may involve privacy protection issues. Machine learning approaches do not require inspecting the content of data packets but identify encrypted traffic based on expert knowledge directly, which overcome the rule-based method's shortcomings of frequent updates of regular expressions and the risks of user privacy leaking. ISLAM [7] still used statistical features extracted artificially as the analysis object when identifying VPN and Tor VoIP traffic, thus losing the advantages of deep learning to automatically extract features. In addition to the above problems, one challenging problem of classifying network traffic is the imbalanced property of network data [8]. It is often difficult for researchers to obtain ground truth traffic, and the inherent characteristics of the Internet will also lead to an imbalanced distribution of data categories [9]. This paper is an extension of our previous work [10], which only considered OTT VPN voice traffic. This pa- per mainly focuses on the classification of OTT voice traffic under complex 5G network traffic. Since the service providers do not provide traffic monitoring interfaces, we build a 5G Non-Standalone (NSA) network to obtain 28 types of typical 5G OTT voice and non-OTT voice traffic. Without using features extracted artificially, we deploy lightweight LSTM, 1-Dimension CNN (1D-CNN) [11], and 2-Dimension CNN (2D-CNN) [12] networks to perform temporal and spatial analysis of 5G OTT VPN voice traffic, respectively. We input the original bytes stream information into the network directly, which guarantees the authenticity and integrity of the data to the greatest extent. It is convenient for network administrators to legally supervise service providers and maintain network security with these methods. The main contributions of our paper can be summarized as follows:

- We set up a 5G NSA experimental network, and collect 28 types of traffic including OTT voice traffic and non-OTT voice traffic in the network. We establish a dataset with OTT voice traffic and non-OTT voice traffic, among which 10 types are OTT voice traffic encrypted through VPN, six types are ordinary OTT voice and video traffic, and the remaining 12 types are non-OTT voice traffic.

- We verify that VPN voice traffic and non-VPN voice traffic are different by using the ISCX VPN-non VPN dataset. In the experimental results, binary classification accuracy exceeds 99.9% and the accuracy of eight classifications exceeds 97.1%.

- We propose LSTM, 1D-CNN [11], and 2D-CNN [12] as feasible strategies, and use Random Forest (RF) and Logistic Regression (LR) as the baseline models, to analyze the VPN OTT voice traffic. The experiment results reveal that LSTM classifies 5G OTT VPN voice traffic with an accuracy of up to 98.30%.

- We propose LSTM, 1D-CNN, and 2D-CNN to identify the OTT voice and non-OTT voice traffic. The experimental results of 2 classifications show that both the proposed deep learning methods can effectively distinguish OTT and non-OTT voice traffic, and the results of 28 classifications show that the classification performance of lightweight LSTM is better than that of 1D-CNN and 2D-CNN.

In general, compared with previous work, we consider a more complex traffic scenario and collect more types of network traffic. The remainder of this paper is organized as follows. Section 2 is related work. Section 3 mainly introduces the proposed network framework and basic knowledge. Section 4 is the experimental evaluation, and the conclusion is in Section 5.

2. RELATED WORK

The reason for the continuous growth of encrypted traffic is the increasing awareness of privacy protection among users and the increasing demand for encrypted data transmission. The diversity of encryption algorithms and encryption methods brings great trouble to the identification of network traffic. At present, traditional classification methods encounter questions in the face of encrypted traffic, and the identification of encrypted traffic has become a headache for operators and service providers. Traditional traffic classification methods have been difficult to meet the needs of current networks' management requirements. Artificial intelligence technology, which has excellent performance in image, language, text processing and other aspects, seems to be able to solve this problem. From the perspective of Artificial Intelligence (AI), network traffic classification methods mainly fall into the following categories: rule-based, traditional machine learning-based and deep learning-based methods. AI technology will play a fundamental role in effectively managing the dynamic flow of information for future applications[3].

2.1 Rule-based methods for traffic classification

The traffic classification methods based on port and Deep Packet Inspection (DPI) belong to the rule-based method, which is characterized by matching classification according to fixed rules determined manually. The port-based approach identifies traffic by associating ports in the TCP/UDP header with TCP/UDP port numbers assigned by Internet Assigned Numbers Authority (IANA). However, the random port strategies, port camouflage technologies, tunneling technologies, and network address translation protocols make the port-based method impossible to identify encrypted traffic. Papadogiannaki et al. [6] focused on the identification of OTT application events using patterns of packet size sequences. They developed a DPI engine that matched and reported events

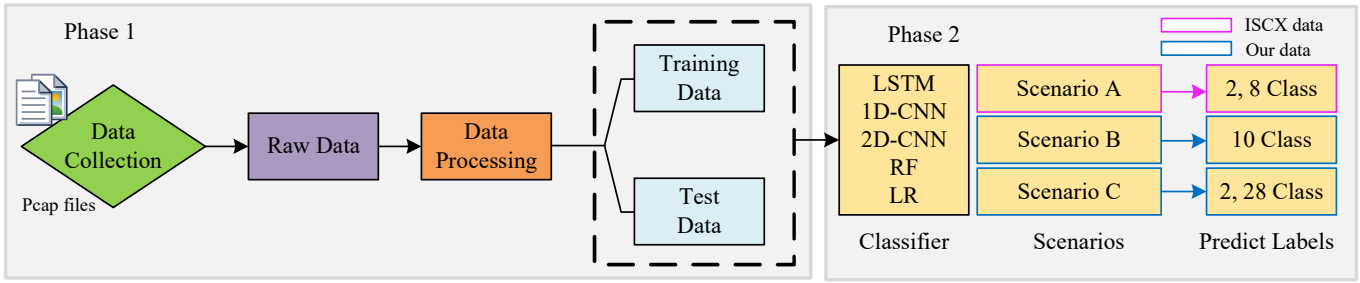

Fig. 1 – The framework of the proposed method

in encrypted network traffic effectively by using an automatic consumption packet size to match a rule set with a packet train pattern on traditional substring and port number patterns.

2.2 Machine learning for encrypted traffic classification

The machine learning-based approach classifies encrypted traffic based on domain expert knowledge, which relies heavily on features extracted manually. For a long time, machine learning has been the main technology of encrypted traffic classification in academia. To enhance network traffic supervision, Yao et al. [13] proposed a new traffic classification model based on Gaussian mixture models and hidden Markov models, MGHMM, which used a K-means clustering algorithm, hidden Markov model, and mixed Gaussians to reduce the details of the flow gradually. The experiments showed a good result on the performance of the proposed scheme. Taylor et al. [14] proposed a fingerprint scheme for Android applications with statistical characteristics of packet length, called AppScanner. The system had the function of automatic fingerprint identification and real-time identification.

2.3 Deep learning for encrypted traffic classification

Deep learning selects features through training automatically, which makes it an ideal traffic classification method. Compared with traditional machine learning methods, deep learning has a relatively high learning ability and can learn high complex patterns. Nowadays, Internet users prefer to communicate through encrypted channels. Due to the great growth of using tunnel and anonymous networks, network management requires new technologies to monitor and analyze network traffic. Islam et al. [7] analyzed the tunnel (VPN) and anonymous (Tor) network traffic based on deep learning technologies (Multi-layer perceptron, 1D-CNN, and LSTM). First, the captured raw traffic is preprocessed, and a 15-second FLT is used to generate a dataset based on FSTFs. The selected FSTFs are then used to distinguish between VPN VoIP traffic and Tor VoIP traffic. Wang et al. [11] proposed an end-to-end encryption traffic classification method based on a one-dimensional convolutional neural network. The method integrates feature extraction, feature selection, and classifier into a unified end-to-end framework, aiming to learn the nonlinear relationship between the original input and the expected output automatically. D'Angelo et al. [15] proposed a deep neural network structure based on Autoencoders (AEs) and studied the following combinations, CNN, LSTM, CNN-LSTM, ConvLSTM, and Stack-CNN LSTM. CNN is used for spatial feature extraction, and LSTM is used for time feature extraction. The introduction of CNN and LSTM layers in a Sparse Autoencoder (SAE) can significantly improve the classification performance. Aceto et al. [16] proposed a new multi- mode framework for encrypting traffic, MIMETIC, which can take full advantage of the heterogeneity of traffic data, overcome the performance limitations of existing traffic classification schemes based on single-mode deep learning, and support challenging mobile scenarios. Vu et al. [17] proposed a deep network for unsupervised learning called Auxiliary Classifier Generative Adversarial Network to generate synthesized data samples for balancing between the minor and the major classes. Inspired by the above deep learning approaches, our study also uses raw bytes stream data to classify encrypted OTT voice traffic in an end-to-end way. The influence of sample sizes and the deep learning methods on classification performance are taken into consideration.

3. ENCRYPTED OTT VOICE TRAFFIC CLASSIFICATION FRAMEWORK

3.1 The proposed framework

Fig. 1 outlines our encrypted traffic identification framework, which includes two phases. Phase 1 consists of data collection and data processing. For some reason, we are unable to capture the unencrypted VPN traffic of some OTT voice applications, for example, WhatsApp. In order to explore the impact of VPN on voice application traffic, we choose to conduct exploratory experiments with data from a public dataset, although this traffic is not 5G traffic. We firstly choose the VPN and non-VPN voice traffic from the ISCX VPN-non VPN dataset. Then we build an experimental 5G NSA network, detailed in Fig. 2, to acquire OTT voice traffic and other types of traffic. The data processing tool

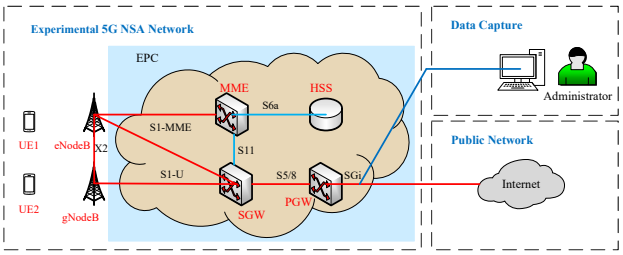

Fig. 2 – 5G NSA network system

USTC-TK2016 developed by Wang et al. [11] is used to process this traffic. The process is divided into four stages. They are traffic split, traffic clean, image generation, and IDX conversion. It should be noted that the IDX3 file processed by this tool contains the original bytes stream information of the data, not the characteristics of the streams or packets. Phase 2 is the classification of encrypted OTT voice traffic; there are mainly three different scenarios. We use a pink box to represent the ISCX VPN-non VPN voice data, which is seen as scenario A in this study. Using 1D-CNN, 2D-CNN and random forest to classify the VPN and non-VPN voice. We use the blue box to represent our own data. Scenario B is 10 classifications of 5G OTT VPN voice traffic, using logistic regression and random forest as the baseline mod- els, and using LSTM, 1D-CNN, and 2D-CNN to identify this traffic. Scenario C is two classifications and 28 classifi- cations of 5G OTT and non-OTT voice traffic, LSTM, 1D- CNN and 2D-CNN are used to identify this traffic. LSTM extracts the temporal characteristics of the traffic automatically, while CNN extracts the spatial characteristics and then they give predictions based on their learning results respectively.

3.2 5G NSA network

According to the definition of a 5G network by 3GPP, Non-Standalone (NSA) and Standalone (SA) are two standard options classified according to different 5G network deployment architectures. The NSA network is a heterogeneous network formed by adding 5G base stations to the 4G core network. SA uses a 5G core network and 5G new air technology, which is a more advanced communication network.
Due to the current 5G technology standards are not unified, the existing open source 5G network functions are not stable and relatively single. In terms of academic research, 5G OAI is a widely used open source project at present. Therefore, this paper selects a 5G OAI open source project to build the laboratory 5G network environment. Until we first finish this research, 5G OAI only has an NSA network, so we choose to build an experimental 5G NSA network.
Therefore, when building the experimental network, we use the 4G core network as the 5G NSA core network, gNodeB, and eNodeB as the 5G and 4G base stations, respectively. We deploy two HP-Z6-G4 workstations with the Ubuntu 18.04 system, one of which is equipped with Evolved Packet Core (EPC) and eNodeB as the core net- work and the other is equipped with gNodeB. The work- station uses Intel Xeon(R) Silver 4210 CPU with 2.20GHz, 40 cores, and 16GB RAM. Besides, two Ettus-USRP B210 are used as RF-front end of eNodeB and gNodeB respectively to transmit data. The experimental 5G NSA network system is shown in Fig 2. In this paper, as an administrator, we use Wireshark to capture experimental data from the SGi interface. The following are the components of EPC:

- Mobility Management Entity (MME): responsible for management and control.
- Service Gateway (SGW): in charge of handling business processes.
- Packet Data Network Gateway (PGW): responsible for interfacing with the Internet.
- Home Subscriber Server (HSS): System database, responsible for storing key user information.
- SGi: It is the user plane interface of a Packet Data Network (PDN). It realizes functions such as protocol encapsulation/decapsulation and address conversion.

3.3 Data collection

In this study, there are three different scenarios, which require different types of traffic.
In scenario A, we select eight types of VPN voice and non-VPN voice traffic in the ISCX VPN-non VPN dataset [18] to make proof experiments, to explore whether VPN has an impact on voice traffic. Details about the dataset can be seen in Table 1. The size of the dataset used is approximately 1.04GB.
Due to the imbalance between the various voice applications in the above ISCX dataset, using only a single dataset hinders the research. So in scenario B, to explore the difference between different 5G OTT VPN voice traffic in detail, we use two 5G mobile phones connected to the 5G base station through VPN. Then, we capture users' traffic from the SGi interface, which is shown in Fig. 2, as the operator's administrator. We collect the OTT voice traffic of 10 OTT applications in total, and the detailed information is shown in Table 2. The reason why we choose these 10 types of typical OTT voice applications is that they are all very popular OTT voice applications in the world at present, and it is representative to choose them as the research objects. The cumulative collection time of each OTT VPN voice traffic ranges from 130 minutes to 523 minutes, and the total data packet size is about 5.15GB.
The limitation of the above two scenarios is that all the traffic are OTT voice traffic, and the classification performance under complex traffic is not taken into consideration. So in scenario C, we collect 28 types of traffic including OTT voice traffic and non-OTT voice traffic in our experimental 5G NSA network, among which 16 types are OTT voice traffic, and the remaining 12 types are non-OTT

voice traffic. The information of 5G OTT and non-OTT voice traffic data is shown in Table 3, which includes not only various OTT voice traffic, but also various non-OTT voice traffic such as live stream, video stream, shopping, and short video.

Table 1 – Scenario A, details of ISCX VPN-non VPN voice traffic

Trace Type	Name	Size	F_A	S_A
non-VPN Voice	Facebook	225MB	6132	5833
	Hangouts	268MB	5501	5202
	Skype	113MB	3715	3281
	Voipbuster	111MB	1554	846
VPN Voice	Facebook_VPN	5.72MB	278	139
	Hangouts_VPN	149MB	549	278
	Skype_VPN	93.9MB	993	515
	Voipbuster_VPN	90.8MB	94	50
	Sum	1.04GB	18816	16144

Table 2 – Scenario B, details of 5G OTT VPN voice traffic

Trace Type	Name	Size	F_A	S_A
VPN Voice	ICQ	725MB	12396	6274
	KakaoTalk	453MB	13300	6802
	Line	596MB	13844	7089
	QQ	432MB	14243	7167
	Skype	661MB	13258	6828
	Telegram	289MB	13607	6879
	ToTok	138MB	13295	6678
	Viber	296MB	12815	6506
	WeChat	1.23GB	13877	6971
	WhatsApp	419MB	12850	6517
	Sum	5.15GB	133485	67711

3.4 Traffic representation and visualization

The two most common choices for traffic representation are sessions and flows. A session is a flow unit divided based on the five-tuple (source IP, source port, destination IP, destination port, and transport layer protocol). The difference between a flow and a session is that a flow only contains traffic in one direction, i.e., the source IP and destination IP ports are not interchangeable. At the same time, the bytes in each data packet can be divided into multiple protocol layers. To solve the problem of representation of traffic, Wang et al. [11] used layer 7 in the ISO/OSI model, layer 4 (L7) in the TCP/IP model, or all protocol layers (ALL) to represent traffic. They proposed four possible representation types: ”Session + L7”, ”Session + ALL”, ”Flow + L7” and ”Flow + ALL”. Inspired by their work, we use ”Session + ALL” and ”Flow + ALL” to represent our traffic in this paper. In the following parts of the paper, we abbreviate ”Session + ALL” as ”S_A” and ”Flow + ALL” as ”F_A”. We only use the first 784 bytes of each session or flow, the excess part will be trimmed, and the part less than 784 bytes will be filled with zeros. After data packets are split, data extraction, decimal conversion and gray image generation are carried out. The visualization results represented by ”S_A” are shown in Fig. 3, and the size of each gray image is 28*28 = 784 bytes. Each pixel with a gray value has an integer between 0 and 255. It can be seen from Fig. 3(a) that ISCX VPN and non-VPN voice traffic in scenario A are obviously different from each other. Therefore, we can say that VPN has an impact on voice traffic. For scenario B, as is shown in Fig. 3(b), the visualization results of different OTT VPN voice traffic are very similar, because the traffic through VPN may use the same encryption technology and have the same five tuples. Fig. 3(c) and Fig. 3(d) show that the nature of OTT voice traffic changes a lot through VPN. Because we can see from the gray image that VPN changes the content of the QQ voice traffic. Fig. 3(e) shows the traffic visualization results of scenario C, as we can see that OTT voice are also very similar, and OTT non-OTTtraffic are different in some way. This may imply that we can easily distinguish between OTT and non-OTT voice traffic using deep learning.

After processing the ISCX VPN-non VPN voice, OTT VPN voice traffic and OTT non-OTT voice traffic, the results obtained are shown in Table 1, Table 2 and Table 3 respectively.

From Table 1, we can see that ISCX VPN-non VPN voice traffic is very unbalanced, taking the representation of ”F_A” as an example, ranging from 94 to 6132 for flows. Table 2 shows the details of the processed 5G OTT VPN voice data. Because the number of samples can be controlled effectively in the laboratory network, a similar number of samples are collected for each type of OTT voice applications. Each type of OTT VPN voice has balanced samples. Table 3 shows the details of processed 5G OTT and non- OTT voice traffic data. Each type of traffic is very imbalanced too, ranging from 154 to 3000 for sessions.

3.5 Deep learning models

Deep learning models can learn high-level features from the input raw data automatically and they overcome the problems of feature design in traditional machine learning. Many recent studies [15] [16] [17] prove the effectiveness of deep learning methods for classifying encrypted traffic.

3.5.1 CNN

A Convolutional Neural Network (CNN) is a kind of network that contains convolutional computation with a certain depth of network structure; it is one of many deep learning networks. CNN can be divided into two processes, forward propagation and back propagation. Forward propagation includes five parts: input layer, convolution layer, activation layer,

Table 3 – Scenario C, details of 5G OTT and non-OTT voice traffic

Num.	Name	S_A	Category	Trace Type	Num.	Name	S_A	Category	Trace Type
0	Bilibili	327	video stream	non-OTT voice	14	Sohu_News	932	news	non-OTT voice
1	Dingding	1159	voice meeting	OTT voice	15	Soul	1723	voice call	OTT voice
2	Douyin	154	short video	non-OTT voice	16	Taobao	331	shopping	non-OTT voice
3	Douyu	3000	live stream	non-OTT voice	17	Telegram_VPN	3000	voice call	OTT voice
4	Huya	3000	live stream	non-OTT voice	18	Tencent_Meeting	1272	voice meeting	OTT voice
5	ICQ_VPN	3000	voice call	OTT voice	19	Tencent_Video	616	video stream	non-OTT voice
6	iQiYi_Video	944	video stream	non-OTT voice	20	ToTok VPN	3000	voice call	OTT voice
7	JD	3000	shopping	non-OTT voice	21	Viber_VPN	3000	voice call	OTT voice
8	KakaoTalk_VPN	3000	voice call	OTT voice	22	WeChat_Video	584	voice call	OTT voice
9	Kuaishou	1553	short video	non-OTT voice	23	WeChat_VPN	3000	voice call	OTT voice
10	Line_VPN	3000	voice call	OTT voice	24	WhatsApp_VPN	3000	voice call	OTT voice
11	QQ_Voice	828	voice call	OTT voice	25	Youku	624	video stream	non-OTT voice
12	QQ_VPN	3000	voice call	OTT voice	26	YouTube	1385	video stream	non-OTT voice
13	Skype_VPN	3000	voice call	OTT voice	27	Zoom	1052	voice meeting	OTT voice
						Sum	52484		

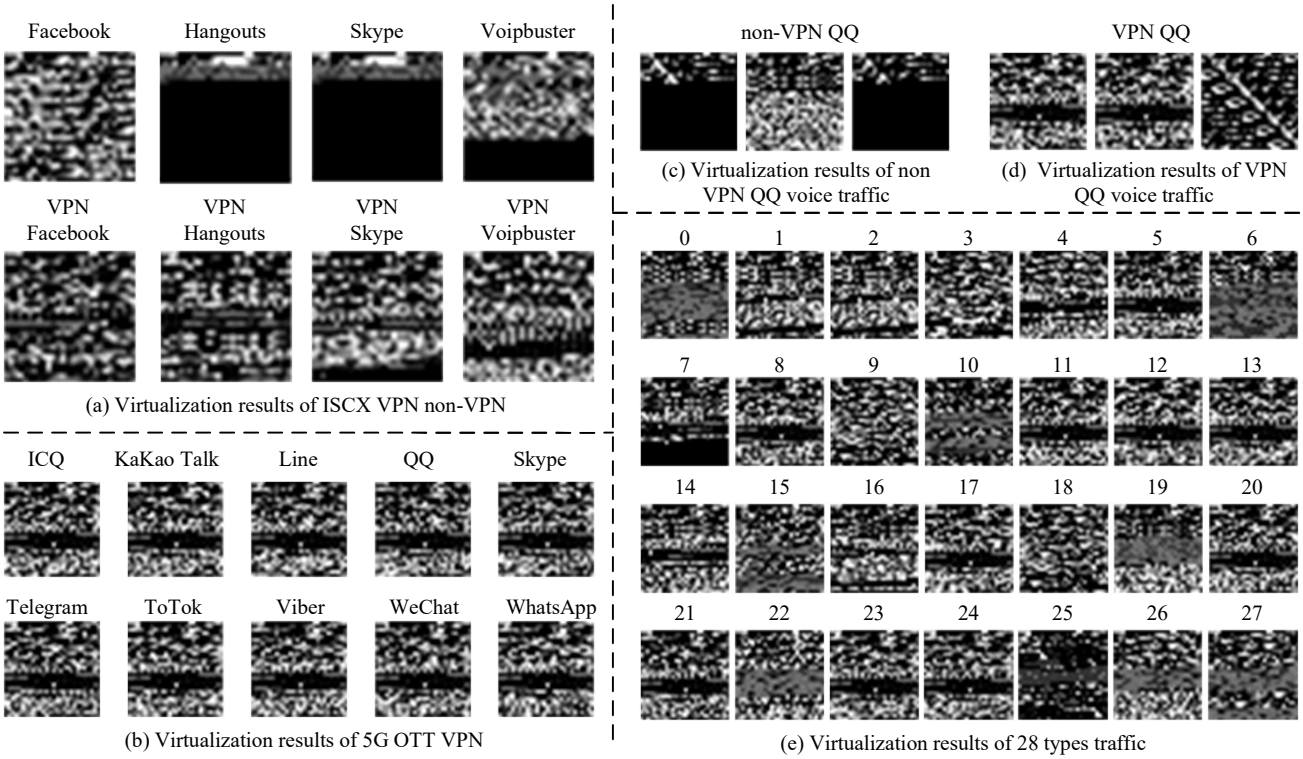

Fig. 3 – Visualization results of processed datasets. Number 0-27 respectively represent 28 types of traffic in Table 3.

pooling layer and full connection layer. The input layer is responsible for the input of data, and the format of the input data may be different when different network models are used. The convolutional layer is an important part in CNN. The main function of the convolutional layer is to extract features. The shallow convolutional layer is used to extract basic features of the image, while the deep convolutional layer is used to extract higher-order features of the image. The convolution operation is a linear translation invariance operation. If the weights of the data are different, the feature information extracted from the original data will be different. The function of the full connection layer is to reduce the dimension of the features output by the convolution layer and the pooling layer to further reduce the computational complexity. Meanwhile, the local features learned by the deep network are integrated to obtain the global features. The Back Propagation (BP) algorithm [19] usually uses the gradient descent method to find the optimal solution of the model. The gradient descent methods mainly include batch gradient descent, stochastic gradient descent and mini-batch gradient descent.

Fig. 4 is a typical LeNet5 network. This model was proposed by Yann LeCun [20]. It was first used for handwritten digit recognition and is now widely used in computer science. Similarly, the model is also effective in identifying encrypted traffic. We choose this model as our CNN classifier. The specific parameters are shown in Fig. 4. Changing the relevant parameters can make the 1D-CNN model become a 2D-CNN model.

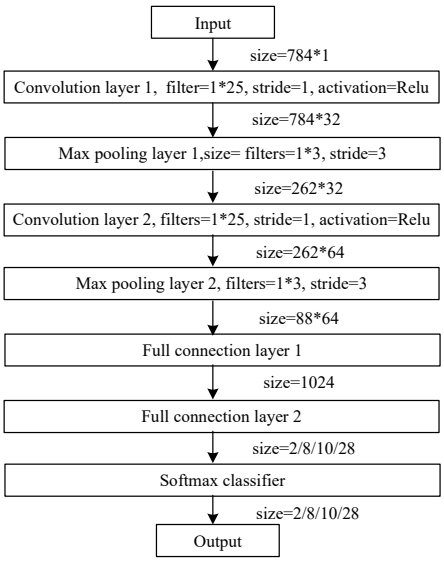

Fig. 4 – A typical LeNet5(1D-CNN) model used in [11]

3.5.2 LSTM

LSTM was proposed by Sepp Hochreiter and Jürgen Schmidhuber [21] in 1997. The LSTM unit is generally composed of a cell, an input gate, an output gate and a forget gate. The cell can remember the value at any time interval; three gates control the input and output information of the cell.

LSTM relies on the cell state throughout the hidden layer to achieve information transfer between hidden units, with only a small amount of linear intervention and changes. LSTM introduces a "gate" mechanism to add or delete cell state information. The "gate" mechanism consists of a $Sigmoid$ activation function layer and a vector dot product operation. The output of the $Sigmoid$ layer controls the ratio of the transfer information.

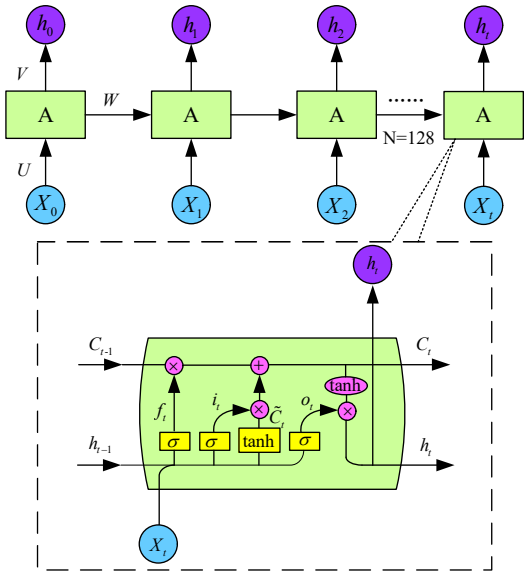

Fig. 5 – Structure of LSTM model

Remark: x_t is the input at time t , h_t is the output at time t, U,V and W are the connection weights, b is the bias, σ is the activation function and is usually $Tanh$ or $Sigmoid$.

Forget gate: LSTM controls the degree of forgetting of cell state information through a forget gate and outputs the current state of forgetting weight, which depends on $h_{t-1}$and x_t.

$$f_t = \sigma\left(U_f x_t + W_f h_{t-1} + b_f\right) \tag{1}$$

Input gate: LSTM controls the input receiving the degree of the cell state through the input gate and output the weights of the current input information. Input gate depends on h_{t-1} and x_t.

$$i_t = \sigma\left(U_i x_t + W_i h_{t-1} + b_i\right) \tag{2}$$

State update:

$$C_t = \tanh\left(U_C x_t + W_C h_{t-1} + b_C\right) \tag{3}$$

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

The "gate" mechanism adds or deletes cell state information, thus achieving long-term memory.

Output gate: LSTM controls the recognition degree of cell state output through the output gate. The output gate depends on h_{t-1} and x_t.

$$o_t = \sigma\left(U_o x_t + W_o h_{t-1} + b_o\right) \tag{5}$$

$$h_t = o_t \cdot \tanh(C_t) \quad (6)$$

To reduce the training cost as much as possible, we deploy a lightweight LSTM model, of which consists one hidden layer with 128 LSTM units and a $Softmax$ output layer.

4. EXPERIMENTAL EVALUATION

4.1 Evaluation metrics

The evaluation metrics used in this paper include accuracy, precision, recall, and F1-score. The metrics are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

In (7) - (10), TP stands for true positive, i.e., the samples of identifying X as X. FP is false positive, i.e., the samples of identifying not X as X. TN is true negative, i.e., the samples of identifying not X as not X. FN is false negative, i.e., the samples of identifying X as not X.

4.2 Experiment design

Our paper designs three application scenarios, namely, scenario A, scenario B and scenario C.

Scenario A is based on the ISCX VPN-non VPN dataset, from which we chose eight types of VPN and non-VPN voice traffic as research objects. Although this scenario is not unfolded in the 5G context, it provides a theoretical basis for our subsequent research. Scenario A consists of Exp. I and Exp. II.

Exp. I: Based on the binary classification problems of the ISCX VPN-non VPN voice dataset, the main purpose of this experiment is to verify whether VPN voice traffic and non-VPN voice traffic have different internal performances.

Exp. II: After obtaining the result of Exp. I, we explore eight classifications on ISCX VPN-non VPN voice traffic further. The main purpose of Exp. II is to verify whether there are differences in VPN voice traffic generated by different applications.

Scenario B is based on our own dataset, which consists of 10 types of OTT voice traffic encrypted through VPN. We select 10 typical OTT voice applications in the world as the research objects. Scenario B consists of Exp. III and Exp. IV.

Exp. III: The CNN models in Exp. I and Exp. II are extended to the OTT VPN voice classification model directly, and then the LSTM model is introduced. In Exp. III, we conduct experiments on both "S_A" and "F_A" traffic representation methods. There are three sub-experiments under each representation method. Each OTT application uses all the traffic samples (Sessions-All, Flows-All), 6000 samples of data (Sessions-6K, Flows-6K), and 3000 samples of data (Sessions-3K, Flows-3K) to explore the impact of sample sizes on traffic recognition performance. Exp. IV: We choose Sessions-3K as the research object to explore the classification performance of three deep learning algorithms (LSTM, 1D-CNN, 2D-CNN) in the case of a small number of samples (44.31% of the total sessions).

Scenario C is based on our own dataset too, it includes both OTT voice traffic and non-OTT voice traffic, among which 16 types are OTT voice traffic, and the remaining 12 types are non-OTT voice traffic. Scenario C consists of Exp. V and Exp. VI. For each category of traffic, a maximum 3000 samples are selected as experimental data. For traffic with less than 3000 samples, all data of the category are selected for the experiment. Exp. V: We choose at most of 3000 samples for each type of traffic as the research object to explore the classification performance of LSTM, 1D-CNN and 2D-CNN in the context of binary classification of OTT and non-OTT voice traffic. Exp. VI: As in Exp. V, we also select 3000 samples as the study object, where the task is to identify different OTT voices through the complex traffic environment. LSTM, 1D-CNN and 2D-CNN have been used as the deep learning methods. The above six experiments are based on the open-source library TensorFlow, the learning rate is 0.001, the training epoch is 100, 20% of the data is randomly selected as the test set, and the remaining 80% is used as the training set.

4.3 Experimental results

All the averages in the table are micro-averages.

4.3.1 Experimental results of scenario A

Firstly, we conduct experiments with the public dataset ISCX VPN-non VPN voice traffic as the research objects. Table 4 shows that the classification test accuracy of Exp. I and Exp. II are quite high. Taking 2D-CNN as an example, the lowest test accuracy of eight classifications also reaches 97.08% (for "F_A").

Table 4 – Scenario A, test accuracy of EXP. I and EXP. II

Task	2 Class (Exp. I)		8 Class (Exp. II)	
Sample Type	F_A	S_A	F_A	S_A
1D-CNN	0.9997	1.0000	0.9819	0.9848
2D-CNN	0.9995	0.9991	0.9708	0.9793
RF	0.9981	0.9985	0.9421	0.9780

Exp. I proves that VPN voice traffic and non-VPN voice traffic are different in some way. This result means that

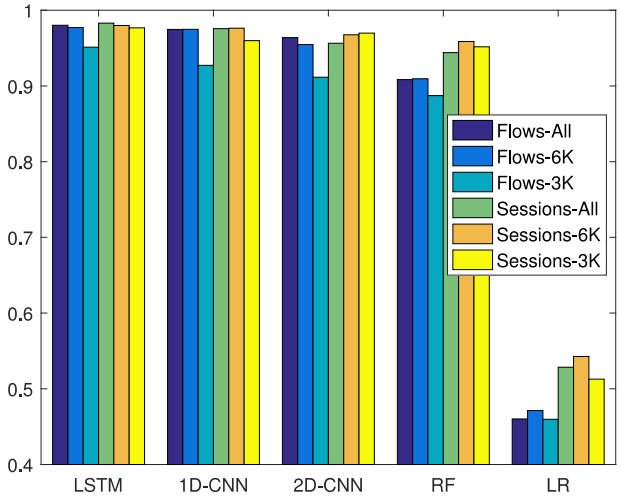

Fig. 6 – Scenario B, Exp. III, test accuracy of using all samples, 6k samples and 3k samples of each method

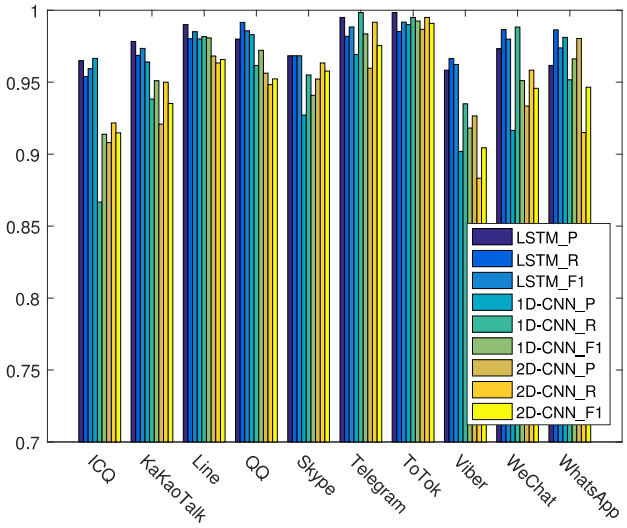

Fig. 7 – Scenario B, precision, recall and F1-score of LSTM, 1D-CNN and 2D-CNN in Exp. IV

VPN does have impact on traffic. Exp. II proves that VPN voice traffic generated by different applications is different. These are interesting results because we can start the classification research of 5G OTT voice traffic based on these results.

4.3.2 Experimental results of scenario B

As for Exp. III, we do 30 experiments, and each experiment is repeated five times to get the best experimental results to explore the impact of sample sizes on classification performance. All the experimental results of Exp. III are shown in Table 5 in the appendix.

It can be seen from Fig. 6 that when the quantity of samples is reduced, the classification test accuracy of "F_A" traffic representation has a clear descending trend, and the performance of 2D-CNN decreases the most (about 5.24%). LSTM, 1D-CNN, and RF also show somewhat of a decline, and the performance of LR just reflects the limitations of traditional machine learning methods in the face of encrypted traffic. In the traffic representation of "S_A", test accuracy decreases as the number of samples decreases, but the maximum decrease is only 1.59% (1D-CNN). From this point of view, the number of samples does affect the performance of traffic classification. Therefore, the sample size for each application should also be taken into account when performing traffic classification.

The results of Exp. III show that "S_A" is a more effective way of identifying traffic than "F_A", so in Exp. IV, Sessions-3K is selected as a sample to examine the recognition performance of different deep learning models for each type of OTT VPN voice traffic. We use precision, recall, and F1-score to evaluate them, and the specific results obtained are shown in Table 6 in the appendix.

As shown in Fig. 7, the identification effect of LSTM in Exp. IV is the best. The average accuracy, average recall, and average F1-score are the highest among the three

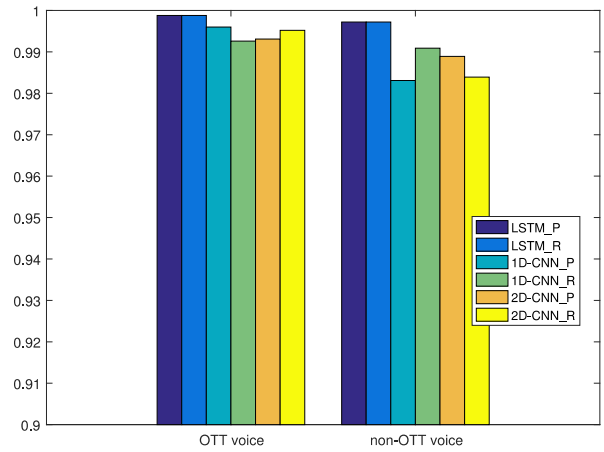

Fig. 8 – Scenario C, precision and recall in Exp. V

methods, reaching 97.68%. Taking the average F1-score as an example, the average F1-score of LSTM is 1.98% and 2.79% higher than that of 1D-CNN and 2D-CNN. In fact, network traffic can be regarded as time-varying data streams, which show their own natures in the distribution of time. Therefore, the lightweight LSTM model proposed in this study outperforms 1D-CNN and 2D-CNN in identifying OTT voice traffic.

4.3.3 Experimental results of scenario C

As seen in Fig. 8, the precision and recall of both OTT and non-OTT voice traffic in Exp. V exceed 98%, and the recognition performance of OTT voice traffic is slightly better than that of non-OTT voice traffic. The specific data is in Table 7 in the appendix. The reason for this is that OTT voice has a greater similarity in some aspects. However, non-OTT traffic is diverse and involves a wide range of areas, with large differences between each

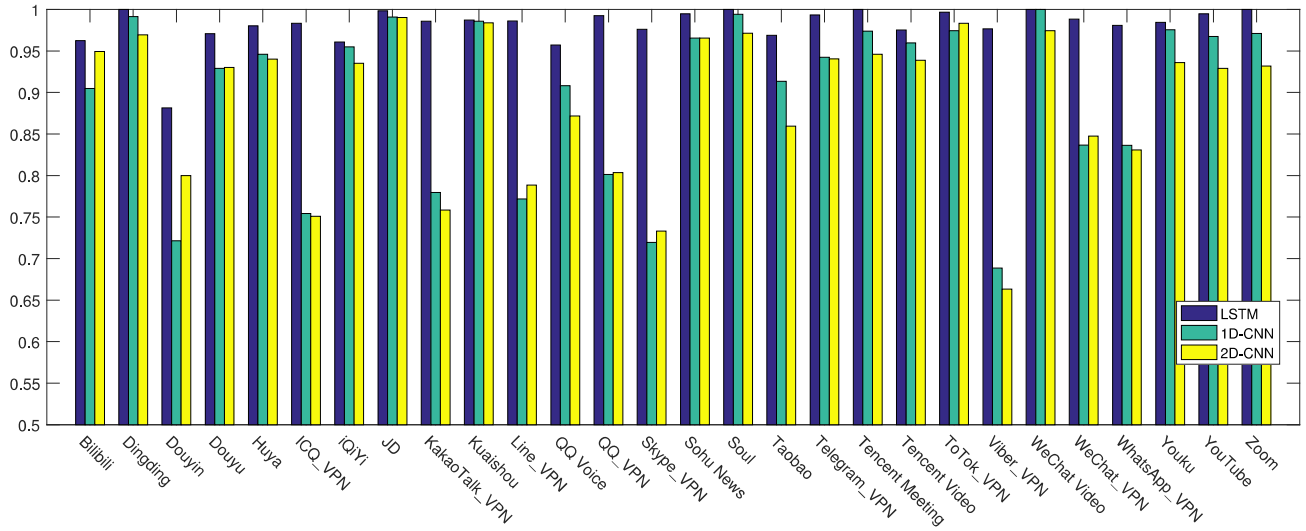

Fig. 9 – Scenario C, F1-score of 28 classifications using LSTM, 1D-CNN and 2D-CNN in Exp. VI

traffic. Overall, the identification performance of LSTM for two classifications is slightly better than that of 1D-CNN and 2D-CNN. Taking the average precision as an example, the average precision of LSTM is 99.80%, 0.84% and 0.7%higher than that of 1D-CNN and 2D-CNN respectively.

For 28 classifications in Exp. VI, from Fig. 9 we can clearly see that the classification performance of LSTM is significantly better than that of 1D-CNN and 2D-CNN. As can be seen from Table 8 in the appendix, taking the aver- age F1-score as an example, the F1-score of LSTM reached 98.58%, respectively 10.92% and 11.55% higher than 1D- CNN and 2D-CNN.

In scenario C, Wechat, QQ, Tencent Meeting and Tencent Video are all products of Tencent company, so they may adopt some similar traffic encryption methods, but the lightweight LSTM we use can still distinguish them well. However, 1D-CNN and 2D-CNN encounter bottlenecks in the face of complex traffic scenarios. As shown in Fig. 9, when facing OTT VPN voice traffic such as ICQ_VPN, their F1-scores are only about 75%, and the results are far off our expectations. Regarding the problem of counterbal- ance data, take Douyin as an example, its total data quan- tity is only 154. LSTM's F1-score can reach 88.14%, while 1D-CNN and 2D-CNN can only reach 72.13% and 80.00%. It can be said that no matter what kind of deep learning method is used, the fewer samples used, the worse the classification performance is.

LSTM learns the temporal characteristics of the data stream, while CNN learns the spatial characteristics of the data stream. These results show their nature in the distribution of space and time. The performance of our lightweight LSTM model is better than that of 1D-CNN [11] and 2D-CNN [12] in both of the three sets of scenarios.

5. CONCLUSION AND FUTURE WORK

This paper is extended from [10], mainly studies the classification of the encrypted 5G OTT and non-OTT voice traffic under complex 5G network context by using deep learning methods. This makes up for the problem of a single research scenario before.
We propose a two-phase framework to identify OTT voice traffic by employing lightweight LSTM, which extracts depth features directly from the input raw bytes stream information in an end-to-end way, thus avoiding the trouble caused by manual feature extraction.
In scenario A, we firstly verify the explicit difference between ordinary traffic and VPN traffic using a public dataset. Then we set up an experimental 5G NSA network and we establish a dataset with OTT voice traffic and non-OTT voice traffic, among which 10 types are OTT voice traffic encrypted through VPN, six types are ordinary OTT voice and video traffic, and the remaining 12 types are non-OTT voice traffic. In scenario B, we evaluate the performance of different sample sizes and the performance of different deep learning methods on 10 types of OTT VPN voice traffic classification. Experiment results show that LSTM is a promising approach for encrypted OTT voice traffic classification. Taking the aver- age F1-score as an example, the average F1-score of LSTM of 10 classifications of OTT VPN voice traffic is 1.98% and 2.79% higher than that of 1D-CNN and 2D-CNN, respectively. In scenario C, we complete the research of OTT and non-OTT voice traffic classification in complex traffic scenario. The average F1-score of LSTM of 28 classifications is 10.92% and 11.55% higher than that of 1D-CNN and 2D-CNN, respectively.
The experimental results show that LSTM is a better identification method than CNN. Meanwhile, these supervised learning methods used in the work cannot identify unknown encrypted traffic. Given that various unknown traffic will appear in the real

network, the unsupervised approaches need to be investigated in future work.

ACKNOWLEDGEMENT

This work is funded by the National Key Research Program, China, of which the project number is 2021YFB2910105.

REFERENCES

[1] Rongpeng Li, Zhifeng Zhao, Xuan Zhou, Guoru Ding, Yan Chen, Zhongyao Wang, and Honggang Zhang. "Intelligent 5G: When Cellular Networks Meet Artificial Intelligence". In: *IEEE Wireless Communications* 24.5 (2017), pp. 175–183. DOI: 10.1109/MWC.2017.1600304WC.

[2] Eftychia Datsika, Angelos Antonopoulos, Nizar Zorba, and Christos Verikoukis. "Software Defined Network Service Chaining for OTT Service Providers in 5G Networks". In: *IEEE Communications Magazine* 55.11 (2017), pp. 124–131. DOI: 10.1109/MCOM.2017.1700108.

[3] Javier López, Juan E. Rubio, and Cristina Alcaraz. "Digital Twins for Intelligent Authorization in the B5G-Enabled Smart Grid". In: *IEEE Wirel. Commun.* 28.2 (2021), pp. 48–55. DOI: 10.1109/MWC.001.2000336. URL: https://doi.org/10.1109/MWC.001.2000336.

[4] Giuseppe Aceto, Domenico Ciuonzo, Antonio Montieri, and Antonio Pescapè. "DISTILLER: Encrypted traffic classification via multimodal multitask deep learning". In: *J. Netw. Comput. Appl.* 183-184 (2021), p. 102985.

[5] Dirk Kutscher. "It's the network: Towards better security and transport performance in 5G". In: *IEEE Conference on Computer Communications Workshops, INFOCOM Workshops 2016, San Francisco, CA, USA, April 10-14, 2016*. IEEE, 2016, pp. 656–661. DOI: 10.1109/INFCOMW.2016.7562158. URL: https://doi.org/10.1109/INFCOMW.2016.7562158.

[6] Eva Papadogiannaki, Constantinos Halevidis, Periklis Akritidis, and Lazaros Koromilas. "OTTer: A Scalable High-Resolution Encrypted Traffic Identification Engine". In: *RAID*. Vol. 11050. Lecture Notes in Computer Science. Springer, 2018, pp. 315–334.

[7] Faiz Ul Islam, Guangjie Liu, Jiangtao Zhai, and Weiwei Liu. "VoIP Traffic Detection in Tunneled and Anonymous Networks Using Deep Learning". In: *IEEE Access* 9 (2021), pp. 59783–59799. DOI: 10.1109/ACCESS.2021.3073967.

[8] Ly Vu, Cong Thanh Bui, and Quang Uy Nguyen. "A Deep Learning Based Method for Handling Imbalanced Problem in Network Traffic Classification". In: *SoICT*. ACM, 2017, pp. 333–339.

[9] Santiago Egea Gómez, Luis Hernández-Callejo, Belén Carro Martínez, and Antonio J. Sánchez-Esguevillas. "Exploratory study on Class Imbalance and solutions for Network Traffic Classification". In: *Neurocomputing* 343 (2019), pp. 100–119.

[10] Zhuang Qiao, Liuqun Zhai, Shunliang Zhang, and Xiaohui Zhang. "Encrypted 5G Over- The- Top Voice Traffic Identification Based on Deep Learning". In: *2021 IEEE Symposium on Computers and Communications (ISCC)*. 2021, pp. 1–7. DOI: 10.1109/ISCC53001.2021.9631458.

[11] Wei Wang, Ming Zhu, Jinlin Wang, Xuewen Zeng, and Zhongzhen Yang. "End-to-end encrypted traffic classification with one-dimensional convolution neural networks". In: *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*. 2017, pp. 43–48. DOI: 10.1109/ISI.2017.8004872.

[12] Wei Wang, Ming Zhu, Xuewen Zeng, Xiaozhou Ye, and Yiqiang Sheng. "Malware traffic classification using convolutional neural network for representation learning". In: *2017 International Conference on Information Networking (ICOIN)*. 2017, pp. 712–717. DOI: 10.1109/ICOIN.2017.7899588.

[13] Zhongjiang Yao, Jingguo Ge, Yulei Wu, Xiaosheng Lin, Runkang He, and Yuxiang Ma. "Encrypted traffic classification based on Gaussian mixture models and Hidden Markov Models". In: *J. Netw. Comput. Appl.* 166 (2020), p. 102711.

[14] Vincent F. Taylor, Riccardo Spolaor, Mauro Conti, and Ivan Martinovic. "Robust Smartphone App Identification via Encrypted Network Traffic Analysis". In: *IEEE Transactions on Information Forensics and Security* 13.1 (2018), pp. 63–78. DOI: 10.1109/TIFS.2017.2737970.

[15] Gianni D'Angelo and Francesco Palmieri. "Network traffic classification using deep convolutional recurrent autoencoder neural networks for spatial-temporal features extraction". In: *J. Netw. Comput. Appl.* 173 (2021), p. 102890.

[16] Giuseppe Aceto, Domenico Ciuonzo, Antonio Montieri, and Antonio Pescapè. "MIMETIC: Mobile encrypted traffic classification using multimodal deep learning". In: *Computer Networks* 165 (2019), p. 106944. ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2019.106944.

[17] Ly Vu, Cong Thanh Bui, and Quang Uy Nguyen. "A Deep Learning Based Method for Handling Imbalanced Problem in Network Traffic Classification". In: *SoICT*. ACM, 2017, pp. 333–339.

[18] Gerard Draper-Gil, Arash Habibi Lashkari, Mohammad Saiful Islam Mamun, and Ali A. Ghorbani. "Characterization of Encrypted and VPN Traffic using Time-related Features". In: *ICISSP*. SciTePress, 2016, pp. 407–414.

[19] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning Representations by Back Propagating Errors". In: *Nature* 323.6088 (1986), pp. 533–536.

[20] Harris Drucker, Corinna Cortes, L. D. Jackel, Yann LeCun, and Vladimir Vapnik. "Boosting and Other Ensemble Methods". In: *Neural Computation* 6.6 (1994), pp. 1289–1301. DOI: 10 . 1162 / neco . 1994.6.6.1289.

[21] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Comput.* 9.8 (1997), pp. 1735–1780.

AUTHORS

Zhuang Qiao is currently pursuing a master's degree in communication and information systems at the University of Chinese Academy of Sciences, Beijing, China. He received his B.S. degree in communication engineering in 2018 from the University of Science and Technology Beijing, Beijing, China. His research interests include mobile communication network security and encrypted traffic identification.

Shunliang Zhang is currently a senior engineer at the Institute of Information Engineering of the Chinese Academy of Sciences. He received his Ph.D. degree (2004) in computer science from Zhejiang University, Hangzhou, China. From 2005 to 2011, he worked as a research engineer on LTE-A standardization at Nokia Siemens Networks. From 2011 to 2016, he was with Ericsson as a technical expert on 5G packet core networks. He has filed more than 50 PCT patent applications on mobile communication systems and holds around 30 USA/EU granted patents. He has published over 20 academic papers. His research interests include mobile communication systems and security.

Liuqun Zhai received a B.Eng. degree in communication engineering from Hangzhou Dianzi University, Hangzhou, China in 2019. He is currently pursuing a master's degree in computer technology at the University of Chinese Academy of Sciences, Beijing, China. His research interests include wireless networking and traffic classification.

Xiaohui Zhang is an engineer at the Institute of Information Engineering, Chinese Academy of Sciences. She received her master's degree in Beijing University of Posts and Telecommunications, Beijing, China. She has been pursuing her doctor's degree at the University of Chinese Academy of Sciences since 2017, Beijing, China. Her research focus is on physical layer security, mobile communications, etc.

APPENDIX

Table 5 – Scenario B, test accuracy of EXP. III

Task	10 Classifications of OTT VPN voice traffic					
Samples	**Flows-All**	**Flows-6K**	**Flows-3K**	**Sessions-All**	**Sessions-6K**	**Sessions-3K**
LSTM	**0.9801**	**0.9771**	**0.9512**	**0.9830**	**0.9798**	**0.9768**
1D-CNN	0.9747	0.9748	0.9273	0.9757	0.9763	0.9598
2D-CNN	0.9639	0.9546	0.9115	0.9564	0.9675	0.9698
RF	0.9084	0.9095	0.8872	0.9440	0.9586	0.9517
LR	0.4601	0.4712	0.4597	0.5285	0.5427	0.5127

Table 6 – Scenario B, precision, recall and F1-score of LSTM, 1D-CNN and 2D-CNN in Exp. IV

Classifier	LSTM			1D-CNN			2D-CNN		
Metrics	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
ICQ	0.9650	0.9539	0.9594	0.9665	0.8667	0.9139	0.9080	0.9217	0.9148
KakaoTalk	0.9783	0.9686	0.9735	0.9640	0.9383	0.9510	0.9208	0.9500	0.9352
Line	0.9900	0.9802	0.9851	0.9800	0.9817	0.9808	0.9682	0.9633	0.9657
QQ	0.9800	0.9916	0.9858	0.9830	0.9617	0.9722	0.9563	0.9483	0.9523
Skype	0.9683	0.9683	0.9683	0.9272	0.9550	0.9409	0.9522	0.9633	0.9577
Telegram	0.9950	0.9819	0.9884	0.9693	0.9983	0.9836	0.9597	0.9917	0.9754
ToTok	0.9983	0.9852	0.9917	0.9900	0.9950	0.9925	0.9868	0.9950	0.9909
Viber	0.9583	0.9664	0.9623	0.9019	0.9350	0.9182	0.9266	0.8833	0.9044
WeChat	0.9733	0.9865	0.9799	0.9165	0.9883	0.9511	0.9334	0.9583	0.9457
WhatsApp	0.9617	0.9863	0.9738	0.9811	0.9517	0.9662	0.9804	0.9150	0.9466
Average	0.9768	0.9769	0.9768	0.9580	0.9572	0.9570	0.9492	0.9490	0.9489

Table 7 – Scenario C, precision and recall of OTT voice and non-OTT voice traffic in Exp. V

Classifier	LSTM		1D-CNN		2D-CNN	
Metrics	**Precision**	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**
OTT	0.9988	0.9988	0.9960	0.9926	0.9931	0.9952
non-OTT	0.9972	0.9972	0.9831	0.9909	0.9889	0.9839
Average	0.9980	0.9980	0.9896	0.9917	0.9910	0.9896

Table 8 – Scenario C, precision, recall and F1-score of 28 classifications using LSTM, 1D-CNN and 2D-CNN in Exp. VI

Classifier	LSTM			1D-CNN			2D-CNN		
Metrics	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Bilibili	0.9412	0.9846	0.9624	0.9344	0.8769	0.9048	0.9492	0.9492	0.9492
Dingding	1.0000	1.0000	1.0000	0.9957	0.9871	0.9913	0.9823	0.9569	0.9694
Douyin	0.9286	0.8387	0.8814	0.7333	0.7097	0.7213	0.9167	0.7097	0.8000
Douyu	0.9748	0.9667	0.9707	0.9199	0.9383	0.9290	0.9300	0.9300	0.9300
Huya	0.9675	0.9933	0.9803	0.9421	0.9500	0.9461	0.9271	0.9533	0.9400
ICQ_VPN	0.9882	0.9783	0.9832	0.7392	0.7700	0.7543	0.7296	0.7733	0.7508
iQiYi_Video	0.9531	0.9683	0.9606	0.9574	0.9524	0.9549	0.9184	0.9524	0.9351
JD	0.9983	0.9983	0.9983	0.9820	1.0000	0.9909	0.9804	1.0000	0.9901
KakaoTalk_VPN	0.9932	0.9783	0.9857	0.8474	0.7217	0.7795	0.8052	0.7167	0.7584
Kuaishou	0.9810	0.9936	0.9872	0.9748	0.9968	0.9857	0.9839	0.9839	0.9839
Line_VPN	0.9818	0.9900	0.9859	0.8574	0.7017	0.7718	0.8527	0.7333	0.7885
QQ_Voice	0.9750	0.9398	0.9571	0.8947	0.9217	0.9080	0.8639	0.8795	0.8716
QQ_VPN	0.9950	0.9900	0.9925	0.7896	0.8133	0.8013	0.7990	0.8083	0.8036
Skype_VPN	0.9688	0.9833	0.9760	0.6858	0.7567	0.7195	0.7123	0.7550	0.7330
Sohu_News	0.9894	1.0000	0.9947	0.9529	0.9785	0.9655	0.9529	0.9785	0.9655
Soul	1.0000	1.0000	1.0000	0.9914	0.9971	0.9942	0.9577	0.9855	0.9714
Taobao	1.0000	0.9394	0.9688	0.9508	0.8788	0.9134	0.9455	0.7879	0.8595
Telegram_VPN	0.9901	0.9967	0.9934	0.9074	0.9800	0.9423	0.9084	0.9750	0.9405
Tencent_Meeting	1.0000	1.0000	1.0000	0.9918	0.9567	0.9739	0.9595	0.9331	0.9461
Tencent_Video	0.9916	0.9593	0.9752	0.9520	0.9675	0.9597	0.9426	0.9350	0.9388
ToTok VPN	0.9983	0.9950	0.9967	0.9656	0.9833	0.9744	0.9849	0.9817	0.9833
Viber_VPN	0.9799	0.9733	0.9766	0.7117	0.6667	0.6885	0.6960	0.6333	0.6632
WeChat_Video	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9744	0.9744	0.9744
WeChat_VPN	0.9883	0.9883	0.9883	0.8205	0.8533	0.8366	0.8432	0.8517	0.8474
WhatsApp_VPN	0.9816	0.9800	0.9808	0.8106	0.8633	0.8362	0.7840	0.8833	0.8307
Youku	0.9690	1.0000	0.9843	1.0000	0.9520	0.9754	0.9360	0.9360	0.9360
YouTube	1.0000	0.9892	0.9946	0.9744	0.9603	0.9673	0.9375	0.9206	0.9290
Zoom	1.0000	1.0000	1.0000	0.9901	0.9524	0.9709	0.9550	0.9095	0.9317
Average	0.9859	0.9858	0.9858	0.8770	0.8763	0.8766	0.8701	0.8705	0.8703