Received March 31, 2021, accepted April 9, 2021, date of publication April 19, 2021, date of current version April 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3073967

VoIP Traffic Detection in Tunneled and Anonymous Networks Using Deep Learning

FAIZ UL ISLAM[1], GUANGJIE LIU[2], JIANGTAO ZHAI[2], AND WEIWEI LIU[1], (Member, IEEE)
[1]School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China
[2]School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

Corresponding author: Guangjie Liu (gjieliu@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant U1836104, Grant 61772281, Grant 61702235, Grant 61801073, Grant 61931004, and Grant 62072250.

ABSTRACT Network management is facing a great challenge to analyze and identify encrypted network traffic with specific applications and protocols. A significant number of network users applying different encryption techniques to network applications and services to hide the true nature of the network communication. These challenges attract the network community to improve network security and enhance network service quality. Network managers need novel techniques to cope with the failure and shortcomings of the port-based and payload-based classification methods of encrypted network traffic due to emergent security technologies. Mainly, the famous network hopping mechanisms used to make network traffic unknown and anonymous are VPN (virtual private network) and TOR (Onion Router). This paper presents a novel scheme to unveil encrypted network traffic and easily identify the tunneled and anonymous network traffic. The proposed identification scheme uses the highly desirable deep learning techniques to easily and efficiently identify the anonymous network traffic and extract the Voice over IP (VoIP) and Non VoIP ones within encrypted traffic flows. Finally, the captured traffic has been classified into four different categories, i-e., VPN VoIP, VPN Non-VoIP, TOR VoIP, and TOR Non-VoIP. The experimental results show that our identification engine is extremely robust to VPN and TOR network traffic.

INDEX TERMS Encrypted network traffic, onion router network, virtual private network, VoIP, anonymous network traffic, convolutional neural network.

I. INTRODUCTION

In today's world, the Internet is the fast growing technology industry and become the essential need to facilitate human being in widespread fields of life. Network traffic comprises Internet activities in the shape of data encapsulated in network packets. Network traffic needs accurate analysis methods such as identification and classification for associating network traffic flows to a specific application class according to network planning and network management. Monitoring encrypted network traffic becomes a challenging issue for many network tasks, including firewall enforcement, quality of service (QoS) implementations, traffic engineering and network security. Due to the exponential proliferation of numerous network applications, network traffic identification techniques need to keep pace with many real-world developments. Generally, the obfuscation tools [1], encryption protocols [2] and tunneling techniques [3] are used to masquerade network traffic control devices and provide secure user's online privacy-preserving.

Classifying encrypted network flows and anonymous communications by their application types is the fundamental process of many crucial network traffic flow monitoring, controlling tasks and forensic investigation of cybercrime. Generally, it mainly focuses on accurate identification and detailed classification. Thus, encrypted and anonymous network traffic lose their unique characteristics. Therefore, the traditional traffic classification (TC) techniques are based on Transmission Control Protocol TCP/ or User Datagram Protocol UDP port mapping assigned by the Internet Assigned Numbers Authority IANA [3] and deep packet inspection (DPI) classification approaches avoid detection due to the usage of non-standard port and encryption techniques [4]. However, these methods can attain high accuracy in the classification of non-encrypted network traffic. Due to the rapid growth of encrypted network traffic flows and emergent security

The associate editor coordinating the review of this manuscript and approving it for publication was Thomas Canhao Xu.

technologies such as encapsulation e.g., Virtual Private Network (VPN) and anonymization e.g., Onion Router (TOR) networks, the traditional classification techniques become obsolete. Hence, the feature-based TC techniques overcome the failure of these conventional classification methods. TOR is a publically available software tool to provide anonymization to the internet user identity. TOR network consists of three nodes (routers): entry node, middle relay, and exit node. The network data is processed with three layers of encryption. Therefore, the VPN and TOR network traffic is difficult to analyze.

In the last few years, the telecommunication arena received unprecedented growth in Voice over IP (VoIP) protocols for making phone calls between VoIP end-users due to low end-to-end delay and high bandwidth requirements. Its dramatic functionality over the traditional telephone network and cost-effectiveness radically revolutionized formal telephone communication. VoIP also provides cheaper communication forms for international calls, online meetings and education. Moreover, the VoIP market attracts people to switch from Public Switched Telephone Network (PSTN) because it offers various features and manages a single network that supports both voice and data. The majority of the VoIP services are encrypted to hide the true contents of the flows. However, VPN and TOR networks make it encapsulated and anonymous to ensure the security of the user's identity and provide end-to-end secure communication.

Tunneled and anonymous VoIP services are essential to be trace and classify them into different categories to either prioritize or restrict them for commercial purposes. Therefore, a competent identification engine is required to differentiate encrypted encapsulated traffic and anonymous traffic and further detect VoIP media traffic from Non-VoIP ones. However, the adoption of more and more anonymous networks for VoIP services makes the analysis of VoIP media traffic a challenging task.

The followings are the main research contributions of this paper:

(1) The main objective of this paper is to present the Flow Spatio-Temporal Features (FSTFs) for distinguishing VPN and TOR traffic. These FSTFs set of attributes are composed of packet length and timing components, which are more suitable for characterizing the VPN and TOR network traffic into VoIP and Non-VoIP ones.

(2) A prolific dataset is generated via FSTFs, which is mature enough to train the classifier based on deep neural network and accurately identify the VoIP traffic flow in VPN and TOR network traffic.

(3) The light-weight proposed identification method is validated via three state-of-art deep learning algorithms, including multi-layer perceptron (MLP), convolution neural network (CNN), and long-short term memory (LSTM). The neural network models are trained with a training set, validated with a validation set and finally tested with 20% unseen data of the total dataset. According to the consideration of the practical implementation efficiency demand, only these three deep learning techniques are employed and tested in this paper.

The structure of the remaining paper is interrelated sections. The background and related work are enlightened in Section II. The preliminaries are discussed in Section III. Section IV outlines the main design of the proposed scheme. Initially, this section described the data pre-processing, features selection, and datasets generation. Secondly, the experimental setup is explained. Thirdly, the architectures of the proposed deep learning models are discussed. Section V briefly evaluated the simulation results and discussed the predictive power of the proposed scheme at the end of this section. Section VI draws conclusions and forecasted future directions.

II. BACKGROUND AND RELATED WORK

In this section, we provide a brief overview of the existing methods of network TC. These are majorly categorized into the port-based classification approach, payload-based inspection technique, and statistical classification. The brief review of the recent work is given on plain network TC, encrypted network TC, tunneled and TOR network traffic identification. Finally, we also survey VoIP traffic detection methods.

A. EXISTING METHODS OF NETWORK TC

1) PORT-BASED CLASSIFICATION APPROACH

The oldest and easiest way of network TC is to classify network traffic based on well-known port numbers, which are visible in TCP/UDP headers of the IP packet as defined by IANA [5]. This type of classification is possible for the specific network application that scan default assigned port numbers. However, due to the utilization of non-standard ports and allocation of dynamic port numbers, usage of tunnels, and Network Address Port Translation (NAPT), it fails to classify the network traffic accurately [6]. The success rate of this classification approach is 70%, while the failure rate is 30-70% of the time [7].

2) PAYLOAD-BASED INSPECTION TECHNIQUE

Payload-based network TC approach identify traffic flows by inspecting payload to find distinctive application signatures. It shows high accuracy for unencrypted network traffic identification [8]. This approach have mainly two drawbacks; firstly, it is not applicable to encrypted network traffic and secondly, it needs the examination of the entire payload of the traffic flow. As it is computationally expensive and not capable of encrypted TC, therefore it is not recommended for tunneled and anonymous network TC.

3) STATISTICAL CLASSIFICATION

The network management got attention to need a novel method to overcome the limitations of the port-based and payload-based classification techniques. Statistical paradigm relies on hand-crafting the unique payload-independent network traffic flow patterns [9]. Despite the payload inspection, this approach connects some flow-based (e.g., bytes per second, packets per second, flow duration, inter packets idle time, inter-arrival time, etc.) and packet-based (e.g., packet length, standard deviation of packet size, packet directions, packet intervals etc.) attributes with the classification task [10]. These hand-designed attributes are further deployed to machine learning (ML) algorithms to classify the specific network traffic flow and characterize them according to the need of the network administrators. As the statistical approach avoids inspection of the packet payload, hence it is applicable to encrypted, encapsulated and anonymous network traffic flows. In general, two types of ML strategies are used for the classification task. The one is unsupervised learning approaches like PCA, DBSCAN, and k-means. The network traffic is classified with similar network traffic clusters generated by diverse protocols. Some of the unsupervised learning are contributed work in the arena of network TC are [11]–[13]. The other is supervised learning approaches like Support Vector Machine (SVM), Genetic Programming (GP), Multi-Objective Genetic Algorithm (MOGA), Naïve byes, and decision trees. A classification model is trained and tested with a set of pre-labeled data entries with handcrafted statistical features. This learning approach maps an input attribute to output class labels. A variety of supervised learning techniques such as SVM, GP, MOGA, Naïve Bayes, and decision trees are used for the identification goal such as classification of P2P traffic, VoIP services and encrypted TC with much higher accuracy [7], [14]–[19].

B. LITERATURE REVIEW ON NETWORK TC

Due to enormous growth in encrypted, tunneled and anonymous network traffic, the network administrators and traffic engineers need to replace the traditional ways of network traffic analysis. ML provides this opportunity to the network research community to overcome the challenges of port-based and payload-based classification respecting user's confidentiality. We paid attention to the research work done in the field of network TC based on statistical classification and ML techniques. Moreover, it will be the motivation for using statistical classification and deep learning approaches. To this end, Table 1 provide the comparative overview of the plain TC along with multiple key features discussed in this section. We have briefly discussed some of the recent work done in the field of network TC.

Yu *et al*. [20] proposed a network video TC based on statistical flow features. A Hierarchical K-Nearest Neighbor (KNN) classifier was developed to classify the network video traffic into six different applications, i.e., QQ, HTTP-download, AHD, ASD, Sopcast, and Xunlei. Lopez-Martin *et al*. [21] proposed a network TC scheme based to classify multiple network traffic services. Six features (inter-arrival time, TCP window size, direction of the packet, source port, destination port, and number of bytes per packet payload) were extracted from the packet headers and employed recurrent neural network (RNN), CNN, and the combination of CNN-RNN architecture for the classification purpose. The CNN-RNN shows the enhanced performance to classify 108 services among the three tested classifiers. Chen *et al*. [22] identified the network protocols (HTTP, FTP, TFTP, TLSV, and SSH) and network applications (Skype, Instagram, Facebook, YouTube and WeChat) using CNN. The authors converted the network flow packets into images and then fed into CNN for the identification of specific application type. Lotfollahi *et al*. [23] utilized the earliest 1500 byte of the flow packets and fed it into Stack Auto Encoder (SAE) and CNN to classify encrypted network traffic. The proposed "Deep Packet" scheme is able to differentiate major classes (e.g., P2P and FTP) and further categorize the network traffic applications like chat, FTP, email, torrent, Skype, etc. The Deep packet scheme achieved 0.94 recall in the classification of major traffic classes and 0.98 in the network traffic application identification. Klenilmar *et al*. [24] present Naïve Bayes video streaming TC. The proposed study shows that video streaming network traffic can be classified into Netflix streaming, YouTube streaming, and background traffic with a 98.88% accuracy level. Recently, Antonio *et al*. [25] focused on identifying IoT devices and behavior in a smart home. Five standard ML techniques, namely, RF, KNN, SVM, majority voting, and decision tree are applied to identify the IoT traffic based on packet-based features. The experimental results show that RF achieves up to 96% accuracy in IoT device identification.

In today's life, the usage of mobile messaging applications increased abruptly due to multiple tasks, such as texting, stream video chat, sharing photos, voice notes, location sharing, ticket booking, paying utility bills, and shopping etc. Therefore, communication services providers and network managers need to properly monitor and priorities the huge amount of encrypted network traffic generated daily. Some of the research work done in the field of mobile TC are reported in the past [26]–[28].

C. TUNNELED AND ANONYMOUS NETWORK TC

Besides tunneling network traffic, anonymous tools have been employed to preserve user's privacy in various factors, i.e., hiding the nature of the communication between the end-users, hiding the source and destination, or sometimes the user's identity too. During past years, many researchers were attracted to anonymous network traffic identification. We summarized the relevant works in Table 2, where columns represents the key aspects of each research paper. Some of the recent works are summarized here.

For instance, Gil *et al*. [29] employed time-related features (inter-arrival time, duration of flow, flow bytes per second, etc.) to characterize encrypted and VPN tunneled network traffic into different categories e.g., streaming, browsing, File Transfer, VoIP, etc. C4.5 and KNN were used as a classifier and achieved accuracy above 80%. Shahbar *et al*. [30] discussed multilayer-encrypted anonymity networks. Packet momentum is employed to successfully identify multilayer-encrypted anonymity network flows using a small quantity of packets and attributes. Bagui *et al*. [39] used time-related

TABLE 1. Summarized comparative review of the plain network TC.

Data type	Input data	Traffic object	Dataset used	Classification model	Traffic type	References
Raw network traffic traces (payload and packet headers)	The size and the direction of the first few packets of the TCP connection	TCP	◑	Clustering techniques (GMM, k-means, HMM)	Applications (NNTP, POP3, SMTP, SSH, HTTPS, POP3S, HTTP, FTP, edonkey, kazaa)	Bernaille *et al.* [31]
Headers of IP packet	TCP connection (packet size, direction, and arrival time)	TCP	●	KNN, HMM	HTTP, HTTPS, SMTP-IN, SMTP-OUT , FTP, SSH, Telnet, and AIM	Wright *et al.*[32]
Packet headers	TCP flows (246 attributes)	BF	●	Bayes estimator and a Bayesian neural network	Attack, bulk, interactive, database, mail, P2P, services, WWW, games, multimedia	Auld *et al.* [33]
Raw traffic traces	TCP traffic flow	BF	●	SVM	Application protocols (HTTP, SMTP, POP3, FTP, MSN, HTTPS, edonkey, IMAPS, gnutella, etc.)	Este *et al.* [14]
Raw traffic traces	TCP/UDP traffic flows (first five packets)	BF	○	Port number, Deep-packet inspection, C4.5, Naïve Bayes	Web, Mail, Bulk, Attack, Chat, P2P, Database, Multimedia, VoIP, Services, Interactive, Grid, Games	Li *et al.* [34]
Raw traffic traces (L7)	TCP/UDP traffic flows (38 flow attributes: total forward packets, max backward inter arrival time, mean active, max idle, etc.)	BF	●	K-Means, EM, DBSCAN, and MOGA	SSH, MSN, HTTP, FTP, DNS, RMCP, ORACLE SQL*NET, NPP, POP3, NETBIOS, IMAP, LDAP, NCP, RTSP, IMAPS and POP3S.	Bacquet *et al.* [35]
Raw traffic traces (HTTPS, SSH, TOR, Update, FTP, Oscar, HTTP, POP3, SMTP, SNMP, route/u, Telnet)	Signature-based methods and statistical analysis methods	UF	◑	Naïve Bayes	HTTPS, ICQ, TOR and other protocols	Sun *et al.* [36]
Raw traffic traces	Statistical flow based attributes and packet header based attributes	BF	●	C4.5, GP, and AdaBoost	Skype and SSH	Alshammari *et al.* [7]
Network layer and transport layer packet headers (L7)	TCP/UDP flows (statistical attributes of the first few application interaction rounds)	BF	○	Naïve Bayes, Bayesian network, PART, C4.5, zeroR and oneR.	59 protocols (bittorrent, DNS, FTP, Skypeout, SSH, Telnet, ppstream, SSL, etc.)	Huang *et al.* [37]
Raw traffic traces	9 flow statistical attributes (client-to-server number of packets, client-to-server average packet bytes, client-to-server minimum inter-packet time, server-to-client maximum packet bytes etc.)	UF	◑	RF, correlation-based classification, one-class SVM, and semi-supervised clustering	FTP, HTTP, IMAP, POP3, RAZOR, SSH, SSL	Zhang *et al.* [38]
Raw network traffic traces	4 statistical flow attributes (Ratio of downstream bytes to upstream bytes, Information entropy of	UF	○	KNN	Video Traffics (AHD, ASD, QQ, HTTP-Download, Sopcast, and Xunlei)	Dong *et al.* [20]

TABLE 1. ***(Continued.)*** **Summarized comparative review of the plain network TC.**

	packet size downstream, the number of downstream sub-flows, Average packet inter-arrival time downstream)					
Network flows (first 20 packets exchanged in a flow lifetime)	6 statistical flow attributes (inter-arrival time, TCP window size, direction of the packet, source port, destination port, and number of bytes per packet payload)	BF	●	RNN, CNN, and the combination of CNN-RNN	HTTP, DNS, SSL_OTHER, Google, TCP_EMPTY, Telnet, TCP_DATA, QUIC, SMTP, YouTube, UDP_DATA, Apple, Office365, Microsoft, NTP	Lopez-Martin *et al.* [21]
Raw network traffic traces (initial 10 network flow packets)	6-channel images	BF	○	CNN	Network protocols (HTTP, FTP, TFTP, SSH, TLSV) and network applications (Skype, Instagram, Facebook, YouTube And WeChat)	Chen *et al.* [22]
Raw network traffic traces at data-link layer (pcap files)	1500 bytes vector	UF	●	SAE and CNN	Network application identification (Aim Chat, Facebook, Gmail, Hangouts, ICQ, Netflix, SCP, SFTP, Skype, Torrent, Tor, YouTube, etc. and traffic characterization (Chat, File Transfer, VoIP, Torrent, etc.)	Lotfollahi *et al.* [23]
Raw network traffic traces (120 second capture, composed of 3 YouTube videos, 2 Netflix videos, and 2 file downloads)	Network traffic flow packets with the set of 14 and 11 statistical variables	PKT	○	Naïve Bayes	Video streaming network traffic (Netflix streaming, YouTube streaming, and background traffic)	Dias *et al.* [24]
Raw traffic traces (pcap files contains the number of bytes transmitted over a one-second window)	Packet length statistics (the statistical mean, the standard deviation and the number of bytes transmitted over a one-second window)	PKT	●	RF, KNN, SVM, majority voting, and decision tree	IOT devices and behavior in a smart home (three smart plugs, six security cameras and one device of each of the following types: personal assistant, smoke alarm, blood pressure meter, sleep sensor, light bulb, printer etc.)	Pinheiro *et al.* [25]

Columns and acronyms meaning is reported hereafter (TABLE I, TABLE II, and TABLE III).

— when not mentioned

Dataset used: ● public dataset ○ private dataset ◐ hybrid dataset

Traffic object: TCP connection (TCP), unidirectional flow (UF), bidirectional flow (BF), packet (PKT)

Input data: Network traffic attributes, Raw network traffic bytes, images, network traffic packets

Data type: Raw network traffic traces (PCAP files), N_{th} layer of OSI or ISO model, network applications traces

Classification model: Adaboost, Bayes estimator, Bayesian neural network, Big Data-enabled hierarchical (BDeH) framework (RF), BiGRU, BayesNet, C4.5, CNN, DBSCAN, Deep-packet inspection, decision tree, EM, GBT, GCA, GMM, GP, HMM, jRip, j48, k-means, KNN, logistic regression, LSTM, MLP, majority voting, Naïve Bayes, OneR, PART, Port number, RF, REPTree, RNN, SVM, , SAE, semi-supervised clustering, and zeroR etc.

features to distinguish tunneled VPN traffic and normal encrypted traffic. Six ML algorithms, namely SVM, Naïve Bayes, RF, logistic regression, KNN and Gradient Boosting Tree (GBT) are compared. The experimental results show that the ensemble methods GBT and RF models outperform the other classifiers in terms of low overfitting and higher accuracy. Wang *et al.* [40] presented a one-dimensional CNN-based end-to-end encrypted TC framework to distinguish VPN traffic from Non-VPN ones. The proposed method is validated with a publically available ISCX dataset [29] contains seven encrypted network traffic types and seven protocol encapsulated network traffic. Lashkari *et al.* [41] detected TOR network traffic based on time-related features via WEKA [42]. RF, C4.5, Zero R, and KNN were employed to characterize them into application types: VoIP, file transfer, mail, audio streaming, video streaming, browsing, chat and P2P. Deng *et al.* [43] identified TOR anonymous traffic via gravitational clustering algorithm (GCA). Statistical features were employed to four clustering algorithms, including k-means, GCA, DBSCAN, and Expectation-Maximization (EM). The standard evaluation criteria explicitly show that GCA is a better choice for TOR traffic recognition. Similarly, Shahbar *et al.* [44] studied the mechanism of I2P network in terms of anonymizing a user's activities and the effect of bandwidth shared by the user's traffic on the I2P network. Huang *et al.* [45] proposed a CNN based multitask learning model to classify VPN network traffic recognition, Trojan classification, and malware detection. The experiments are validated on the public ISCX VPN traffic dataset and CTU-13 malware dataset. Pescape *et al.* [46] attempted flow-based anonymous traffic flows classification. The anonymous traffic obtained is I2P, TOR, and JonDonym. Five ML techniques (Multinomial Naïve Bayes, Naïve Bayes, Bayesian Networks, RF, and C4.5) are used to differentiate the three encrypted traffic flows. Collectively. C4.5 and RF outperform other algorithms. Kim *et al.* [47] proposed a method to classify TOR/Non-TOR traffic. The authors used UNB-CIC [41] dataset to validate the proposed experiments. The authors parsed the first 54 bytes of TCP/IP raw packet header, which are fed to the 1D-CNN model as an input for the classification task. Two sets of experiments were done. In scenario 1, the anonymous traffic was classified as TOR and Non-TOR. In scenario 2, the anonymous traffic was characterized into eight classes, i.e., audio streaming, chat, P2P, video streaming, file transfer, VoIP, and browsing, and e-mail. Yao *et al.* [48] classified regular encrypted and VPN network traffic flows using hierarchical attention network and LSTM. The models used the preprocessed data in the shape of M*N-dimensional matrix, where M represents the number of packets in a network traffic flow and N represents the count of bytes in a packet. Zeng *et al.* [49] proposed a deep learning model termed "deep-full-range (DFR)" to accomplish encrypted network traffic classification and intrusion detection. In order to feed the deep learning models, the authors generated idx files from raw traffic traces. The proposed DFR model is validated with two publically available datasets, ISCX 2012 IDS dataset [50] and ISCX VPN-nonVPN traffic dataset [29]. Montieri *et al.* [51] investigated anonymity tools for network TC purpose via a hierarchical approach. The proposed hierarchial approach consists of four ML based classifiers, i.e. Bayesian Network, C4.5, Naïve Bayes, and RF. The ultimate goal of this approach is to classify I2P, JonDonym, and Tor network traffic into Anonymity networks adopted (L1), network traffic types (L2), and specific application type (L3). Bovenzi *et al.* [52] addresses the anonymous network TC generated by Tor, I2P, and JonDonym. The authors presented a Big Data-enabled Hierarchical framework (BDeH) to implement double parallelism i.e. model and data parallelism, provided by the combination of Hierarchical TC and big data technologies. Recently, Aceto *et al.* [53] proposed a novel multimodal deep learning based encrypted multitask TC (termed DISTILLER). The proposed DISTILLER classifier provides a solution of encrypted TC. The proposed classifier is validated with the public dataset ISCX VPN-nonVPN [29].

D. VOIP TC

In this section, we reviewed some of the previous research studies on VoIP traffic identification and classification. Due to significant growth in encrypted, tunneled, and anonymous network traffic utilization, the traditional classification techniques, i.e., port-based or payload-based, are obsolete now. Statistical classification and ML based classification approaches are hot areas to be studied for VoIP traffic analysis. The previous studies are listed in Table 3 with key features of the proposed methodologies. Some of the recent reported research done in the field of VoIP traffic analysis are discussed here.

Alshammari *et al.* [57] investigated encrypted VoIP traffic generated by Skype, GTalk, and Primus softphone. Initially, the NetMate toolset [58] is used to process the captured network traffic and generate traffic flows with flow-based features. For classification purposes, the authors tested three famous classifiers (GP, C4.5, and Adaboost) to accomplish the classification task based on extracted flow-based features. Qin *et al.* [59] proposed an identification scheme based on PSD instead of handcrafted attributes. The developed model tested the initial few IP packets to derive the PSD and employ it for the classification of P2P and VoIP applications. Recently, Mazhar *et al.* [60] studied encrypted and tunneled network traffic and proposed a statistical analysis-based method to detect VoIP traffic flows. After the call initiation, the 6 seconds captured traffic is utilized to detect real-time VoIP media calls with the FPR up to 0.00015% and TPR up to 97.54%.

According to the previous studies done in the field of network TC and identification, very few works focused on normal network traffic analyses, mixed encrypted network TC, and tunneled network traffic detection. Most of these works based on traditional ML algorithms. With the pace of time, the network management needs to monitor and control the illegal usage of encryption tools such as VPN and TOR networks. The cybersecurity engineers need novel techniques and solutions to properly classify and identify the encrypted, encapsulated, and anonymous network traffic flows worldwide. Due to these considerations, this paper specifically focused on VPN and TOR network traffic identification based on deep learning. In addition to the identification task, the VoIP traffic flows are detected both in tunneled and anonymous network flows.

TABLE 2. Tunneled and anonymous network TC.

Data type	Input data	Traffic object	Dataset used	Classification model	Traffic type	References
Raw tunneled network traffic traces	49 flow attributes	BF and UF	○	Naïve Bayes	IPsec tunneling and Point-to-Point Tunneling Protocol (PPTP) (HTTP, FTP, SMTP, SSH)	Okada *et al.* [54]
PPTP and IPsec encrypted network traffic flows (Packet headers)	29 network traffic attributes	UF	○	Classifying of encryption type via C4.5, and application identification via SVM	IPsec, BULK, INTERACTIVE, P2P, STREAMING, WEB	Kumano *et al.* [55]
Raw network traffic traces (14 traffic categories: VOIP, VPN-VOIP, P2P, VPN-P2P, etc.)	Time-related features (inter-arrival time, duration of flow, flow bytes per second, etc.)	BF	●	C4.5 and KNN	Encrypted and VPN tunneled network traffic (e.g., streaming, browsing, File Transfer, VoIP, etc.)	Draper-Gil *et al.* [29]
Raw network traffic traces (The first 3-packets of the multilayer-encryption anonymity networks)	11 packet momentum attributes (maximum packet size, packet sequence, second maximum packet size, packet momentum, etc.)	PKT	●	C4.5, Random Forest Naive Bayes, Bayesian Network	22 classes of anonymous network traffic and the obfuscated traffic (HTTPS, IMAPS, SSH, I2P, Tor, Flashproxy, JonDonym, etc.)	Shahbar *et al.* [30]
Raw network traffic flows (pcap files)	24 Time-related features	UF	●	SVM, Naïve Bayes, RF, logistic regression, KNN and GBT	Tunneled VPN traffic and normal encrypted traffic (chat, file transfer, e-mail, P2P, media streaming and VoIP)	Bagui *et al.* [39]
Raw network traffic flows (pcap files)	idx files (784*1)	UF	●	1D-CNN	Tunneled VPN traffic and normal encrypted traffic (chat, file transfer, e-mail, P2P, media streaming and VoIP)	Wang *et al.* [40]
Raw network traffic flows (pcap files)	23 Time-related attributes	BF	●	RF, C4.5, Zero R, and KNN	TOR network traffic (VoIP, file transfer, mail, audio streaming, video streaming, browsing, chat and P2P)	Lashkari *et al.* [41]
Raw network traffic flows (TOR, DNS, Http, SSH, SSL)	26 statistical attributes	BF	○	k-means, EM, GCA, and DBSCAN	TOR anonymous traffic	Deng *et al.* [43]
Raw network traffic flows (pcap files)	23 flow-related and time-related traffic attributes	BF	●	BayesNet, OneR, jRip, REPTree, and j48	TOR network traffic (chat, audio-streaming, video-streaming, mail browsing, P2P, VoIP and file transfer)	Cuzzocrea *et al.* [56]
I2P network traffic traces (browsing, downloading, and chat)	91 statistical attributes (flow direction, duration, and the frequencies related to the packets in a flow)	UF	○	C4.5	jIRCii, I2Psnark, Eepsites, Exploratory and Participating Tunnels	Shahbar *et al.* [44]
Raw network traffic traces (VPN and malware raw traffic)	idx format (32*32 bytes.)	BF	●	CNN	VPN (audio streaming, chat, P2P, video streaming, browsing, e-mail, etc.) and malware traffic (Neris, Emotet, Kazy, Geodo, etc.)	Huang *et al.* [45]
Raw network traffic traces (I2P, TOR, and JonDonym)	81 statistical attributes (Flow direction (A/B), Inter-Arrival Time (IAT), TCP header related attributes, etc.)	BF	●	C4.5, Naïve Bayes, RF, and Bayesian Networks	I2P, TOR, and JonDonym	Pescape *et al.* [46]
Raw network traffic flows (pcap files)	The first 54 bytes of TCP/IP raw packet header (108*1)	BF	●	1D-CNN	TOR/Non-TOR traffic (audio streaming, chat, P2P, video streaming, file transfer, VoIP, and browsing, and e-mail)	Kim *et al.* [47]
Raw network traffic traces (pcap files)	Traffic flow matrix (traffic flow*the number of bytes in a packet)	BF	●	LSTM and hierarchical attention network	VPN and Non VPN network traffic flows (Chat, e-mail, file transfer, P2P, streaming, VoIP, and web browsing)	Yao *et al.* [48]

TABLE 2. ***(Continued.)*** **Tunneled and anonymous network TC.**

Raw network traffic traces (pcap files)	idx files (30 bytes * 30 bytes)	BF	●	1D-CNN, LSTM, SAE	Regular encrypted traffic (Chat, e-mail, file transfer, P2P, streaming, VoIP, and web browsing) Intrusion detection (Brute force SSH, DDoS, HttpDoS, Infiltrating transfer, and Normal)	Zeng *et al.* [49]
Raw network traffic (I2P, JonDonym, and Tor)	TC_set (traffic flows: 81 per-flow statistical features) EarlyTC_set (the first K packets of each flow)	UF and BF	●	Hierarchical approach (C4.5, Bayesian Network, Naïve Bayes, and RF)	Anonymity networks adopted (L1), network traffic types (L2), and specific application type (L3)	Montieri *et al.*[51]
Raw network traffic traces (Tor, I2P, and JonDonym)	74 flow statistical attributes	UF	●	Big Data-enabled hierarchical (BDeH) framework (RF)	Tor, I2P, and JonDonym	Bovenzi *et al.* [52]
Raw network traffic traces	Transport layer payload (first 784 bytes) and the first 32 header packets	BF	●	1D-CNN & BiGRU 1D-CNN, 2D-CNN MLP, 2D-CNN+LSTM	VPN/Non-VPN traffic (VoIP, file transfer, P2P, streaming, chat, e-mail)	Aceto *et al.* [53]

III. PRELIMINARIES

The following section presents a detailed description of the deep learning models used in the proposed scheme and discussed the assessing predictive ability metrics used for evaluation of the experimental results. For reader's convenience, Table. 9 summarizes the acronyms used in the manuscript, which are listed at the end of the paper.

A. DEEP LEARNING MODELS

This section briefly explained the deep learning models used for proposed identification and detection purpose.

1) MLP

MLP is a kind of feed-forward neural network model consists of multiple layers, including an input layer, multiple hidden layers, and an output layer [65]. The dimension of these layers varies model-to-model according to the nature of the problem. The neurons of each layer are fully connected to the neurons of the subsequent layer. Mathematically, MLP can be succinctly expressed as:

$$f(n) = y(b^{(i+1)} + w^{(i+1)}(s(b^{(i)} + w^{(i)}n))) \quad 1 \leq i \leq L \quad (1)$$

where w indicates weight matrices of the hidden layer and output layer, b is bias vector, s and y represents activation functions of the layers. The weight matrices and bias vectors are randomly selected in the initial phase, then updating during the training session for optimization.

2) 1D-CNN

The convolutional neural network (CNN) is a widely-used model of deep learning; initially, it is preferred for image recognition problems. Later on, the researchers applied it in various fields and achieved state-of-art accuracies such as object detection, image classification, and network TC. CNN has a strong ability to automatically extract the critical features via chaining convolutional layers. Each layer is comprised of a set of filters (or kernels) that are convolved with the input units to extract spatial features of the certain input region. Another important feature of CNN architecture is the pooling layer. It is located in between successive convolutional layers, aiming downsampling to reduce complexity and parameters and also reduce the overfitting [28]. The output layer, commonly called the softmax layer, contains the activation function to accomplish the classification task. The output layer outputs N-dimensional (N is the number of the output classes) probability distribution vector [0, 1]. Each real value represents an output class score. The architecture of the CNN could be 1D or 2D or 3D, depends upon the nature of the specific problem.

3) LSTM

The RNN is an ineffective technique for long sequence modeling due to gradient disappearance. To overcome the shortcoming of standard RNN, Hochreiter *et al.* [66] introduced the developed form of RNN called long-short term memory (LSTM) model, which is able to model long-term dependencies. LSTM works on the sequential form of data and has been used in a variety of fields such as speech recognition, handwriting recognition, natural language processing tasks such as machine translation, speech recognition and constituency parsing, and language modeling. LSTM contains complex memory units instead of neurons in general neural networks. LSTM units have the ability to store the information for longer time periods in the shape of a state vector. The memory units contain several gates such as input gate, forget gate, output gate to control the information passing along a sequence.

B. ASSESSING PREDICTIVE ABILITY

In order to examine the effectiveness of the proposed identification scheme, we employed a number of parameters to evaluate the performance assessment metrics [67]. 1) True Positive (TP): When the network traffic flow is classified correctly as F. 2) False Positive (FP): When the network traffic flow is classified incorrectly as F. 3) True Negative (TN): When the network traffic flow correctly classified as Not-F. 4) False Negative (FN): When the network traffic flow is incorrectly classified as Not-F. The identification scheme is evaluated with four performance assessment metrics to determine the predictive power of the proposed scheme,

TABLE 3. Summary of VOIP TC methods.

Data type	Input data	Traffic object	Dataset used	Classification model	Traffic type	References
Raw network traffic traces	7 packet based attributes (the time between first and last packet, average packet size, average packet/sec, the number of packets etc.)	PKT	○	i- Average Packets/sec are between 20 and 40 ii- Average packet size in bytes is between 100 and 200	VoIP services (MSN, Skype, Google Talk, and YAHOO) and Non-VoIP traffic (file downloading, MSN text chat, file transfer, video streaming, and gaming)	Fauzia *et al.* [61]
Raw network traffic traces	PSD and port association	—	○	Supervised clustering	P2P, gaming, FTP, streaming, HTTP	Lin *et al.* [62]
Raw network traffic traces	i- The host behavior estimation (IP addresses and port numbers) ii- Flow statistical attributes (packet size and the inter-packet time)	BF	○	i- Function of entropy (packet size modeling) ii- Selfadaptive estimation (inter-packet time)	VoIP applications (Tom-Skype, Express Talk, QQ voice etc.) Non-VoIP applications (web browser, multimedia pplications, online gaming, etc.)	Li *et al.* [63]
Raw network traffic traces	Statistical and signature-based attributes	BF	◑	Naïve Bayes, Linear Discriminant Analysis, KNN, SVM	Skype (signalling traffic, calls, and file transfer)	Adami *et al.* [64]
Raw network traffic traces	22 flow-based features	BF	◑	GP, C4.5, and Adaboost	Skype, GTalk, Primus softphone, and non-VoIP	Alshammari *et al.* [57]
Real time traffic capture	Mixed network traffic flows	BF	○	PSD of the initial few packets in the entire flow	P2P (PPTV, Qvod, PT, Thunder) and VoIP (Skype, MSNtalk, Gtalk, QQTalk)	Qin *et al.* [59]
Raw network traffic traces (VoIP traffic flows: SKYPE, GTALK, Yahoo, QQ messenger, Zfone, and Blink etc.) (Non-VoIP traffic flows: torrent, software updates, online live TVs, FTP, YouTube etc.)	6 statistical parameters (Packets rate in packets/s, Mean value of all packets sizes in bytes, Standard Deviation value of all packets sizes in bytes etc.)	BF	◑	Threshold	Skype, GTalk, Yahoo, MSN, Asterisk, Zfone, X-Liet, Eyebeam clients etc.	Mazhar et al. [60]
Network traffic flows (VPN and TOR network traffic flows)	60 statistical features (see Table V)	BF	●	MLP, 1D-CNN, LSTM	VPN VoIP, VPN Non-VoIP, TOR VoIP, and TOR Non-VoIP	Proposed method

including 1) Precision (*Pr*). 2) Recall (*Rc*). 3) F-measure (*F-m*). 4) Accuracy (*acc*).

These metrics are defined as [67];

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

$$Rc = \frac{TP}{(TP + FN)} \tag{3}$$

$$F - m = 2 \times \frac{Rc \times Pr}{Rc + Pr} \quad 0 \leq F - m \leq 1 \tag{4}$$

$$acc = \frac{(TP + TN)}{(TP + TN + FN + FP)} \times 100\% \tag{5}$$

To visualize the correctly and incorrectly network traffic classes, we used the confusion matrix. Each row in the confusion matrix represents the actual class label, while each column represents the predicted class label. The diagonal of the confusion matrix indicates correctly classified class labels. Furthermore, we have drawn the accuracy and loss of the training phase and validation phase.

IV. WORK DESCRIPTION

The ultimate goal of the proposed work aims to detect VoIP traffic flows in tunneled (VPN) and anonymous (TOR) networks. The general flow diagram of the VoIP detector is

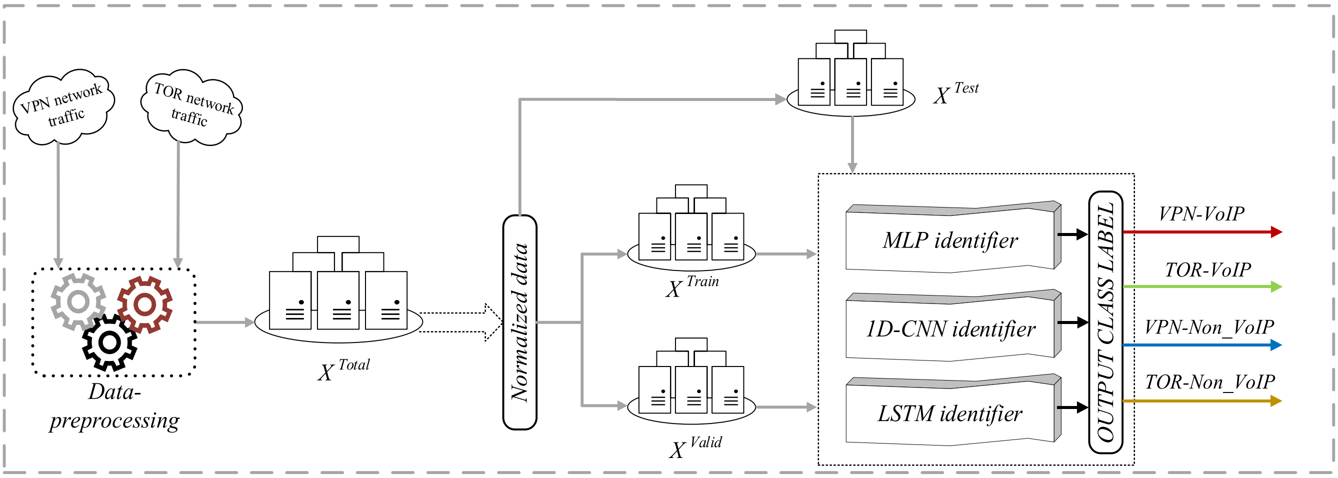

FIGURE 1. Overview of deep learning based tunneled and anonymous network traffic identifier.

TABLE 4. List of captured VPN and TOR traffic traces.

Traffic Type	*Content Type*	
VPN VoIP traces	Hangouts, Skype, Facebook, and VoIPbuster voice calls	
TOR VoIP traces	Hangouts, Skype, and Facebook voice calls	
VPN Non-VoIP Traces	Chat	Skype, IAM, ICQ, Facebook, and Hangouts
	P2P	Bittorent and µTorrent
	Streaming	Vimeo, Netflix, YouTube, and Spotify
	File Transfer	FTPS, SFTP, and Skype
	Email	SMTP/S, IMAP/SSL, and POP3/SSL
TOR Non-VoIP Traces	Chat	Skype, IAM, ICQ, Facebook, and Hangouts
	P2P	Bittorent (kali linux distribution and Vuze application)
	Streaming	Vimeo, and YouTube
	File Transfer	FTPS, SFTP, and Skype
	Email	SMTP/S, IMAP/SSL, and POP3/SSL

given in Figure. 1. Firstly, this section presents a detailed description of data pre-processing. It includes the dataset introduction, conversion of raw traffic into FSTFs based flows. Then followed by network traffic flows labeling based on the FSTFs set. Secondly, here the system parameters and the experimental environment has been explained. Finally, the experiments were done for the VoIP traffic detection in VPN and TOR network flows.

A. DATA PRE-PROCESSING

1) DATASETS DESCRIPTION

One of the main tasks of this proposed work is to generate a mixed dataset containing both VPN and TOR traffic. Therefore, we merged two publically available datasets published by the Canadian Institute of Cybersecurity (CIC), the University of New Brunswick, the ISCX VPN-NON_VPN dataset [29], and ISCX TOR-NON_TOR dataset [41]. Both the datasets consist of seven different types of traffic traces in pcap format. VPN-NON_VPN dataset consists of seven types (email, file transfer, VoIP, P2P, web browsing, chat, and streaming of encrypted and tunneled traffic traces). As we focused on VoIP detection in tunneled traffic, therefore we separated VPN traffic traces (about 2.3 GB) and grouped them into VPN VoIP and VPN Non-VoIP classes. Similarly, we extracted only TOR network traffic traces (about 8.8 GB) and grouped them into TOR VoIP and TOR Non-VoIP classes. Finally, we get an enriched dataset containing VPN and TOR network traffic flows. The details of the generated sub-dataset are given in Table 4.

2) CONVERSION OF RAW TRAFFIC INTO NETWORK TRAFFIC FLOWS

The VPN and TOR traffic traces listed in Table 4 is further processed to generate traffic flows based on five parameters, i.e., Protocol (TCP/UDP), Src port, Dst port, Src IP, and Dst IP with 76 FSTFs set. The CICFlowmeter (an open-source java-based tool) is used as a bidirectional flow generator, which aggregated the pcap traces into network traffic flows (CSV files) [68]. The flow latency period (FLP) is controlled by the flow generator as the earlier studies mentioned that smaller FLP depict significant results [67]. Therefore, we selected 15 sec FLP in the proposed VoIP detector. The traffic flows obtained as an output of the CICFlowmeter are labeled according to the application type mentioned in Table 4. The network traffic flows generated by chat, P2P, streaming, file transfer, and email applications are grouped into Non-VoIP class, while the flows computed by Hangouts, Skype, Facebook, and VoIPbuster voice calls are labeled as VoIP class. We eliminated the duplicate flows to ensure a refined dataset for further identification task.

3) FLOW SPATIO-TEMPORAL FEATURES SELECTION

The CICFlowmeter discussed in the previous section generates network traffic flows with a handsome amount of FSTFs. In the past, many researchers deployed the Netmate tool kit [58] for the flow generation. Netmate tool kit can generate network traffic flows with a maximum number of 22 flow features. In contrast, CICFlowmeter produces a wide variety

TABLE 5. Flow Spatio-Temporal features set.

	Abbreviation	*Description*
1	Flduration	Flow duration (microseconds)
2	Tot Fwd Pkts, Tot Bwd Pkts	Total packets in forward/backward directions
3	Totlen Fwd Pkts, TotLen Bwd Pkts	Total size of packets in Forward/ backward directions
4	Fwd pkt Len (Min, Max, Mean, Std)	Minimum, maximum, mean, and standard deviation size of packets in forward direction
5	Bwd pkt Len (Min, Max, Mean, Std)	Minimum, maximum, mean, and standard deviation size of packets in backward direction
6	Flow IAT (Min, Max, Mean, Std)	Minimum, maximum, mean, and standard deviation time between two packets sent in the flow
7	Fwd (IAT Min, Max, Mean, Std, Total)	Minimum, maximum, mean, standard deviation, and total inter-arrival time of the packet in the forward direction
8	Bwd (IAT Min, Max, Mean, Std, Total)	Minimum, maximum, mean, standard deviation, and total inter-arrival time of the packet in the backward direction
9	Bwd PSH Flags	Number of times the PSH flag was set in packets travelling in the backward direction (0 for UDP)
10	Fwd Header Len, Bwd Header Len	Total Bytes used for headers in the Forward/backward directions
11	Fwd Pkts/s, Bwd Pkts/s	Number of Forward and backward packets per second
12	Pkt Len (Min, Max, Mean, Std, and Var)	Minimum, maximum, mean, standard deviation, and Variance length of a packet
13	FLAG CNT (FIN, SYN, RST, PSH, and ACK)	Number of packets with FIN, SYN, RST, PSH, and ACK
14	Down/up Ratio	Download and upload ratio
15	Pkt Size Avg	Average size of packet
16	Seg Size Avg (Fwd and Bwd)	Average size observed in forward and backward direction
17	Subflow pkts (Fwd and Bwd)	Average number of packets in a sub flow in the forward and backward direction
18	Subflow Byts (Fwd and Bwd)	Average number of bytes in a sub flow in the forward and backward direction
19	Init Bwd Win Bytes	The total number of bytes sent in the initial window in the backward direction
20	Fwd Act Data Pkts	Count of the packets in forward direction with at least 1 byte of TCP data payload
21	Active (Min, Max, Mean, and Std)	Minimum, Maximum, Mean, and Standard deviation flow active before becoming idle
22	Idle (Min, Max, Mean, and Std)	Minimum, Maximum, Mean, and Standard deviation flow idle before becoming active

of 76 FSTFs set, which is more than enough to accurately classify encrypted network traffic. In the proposed identification task, we truncated the features with zero or undefined values because it will be ineffective in the next steps. The final effective FSTFs set contains 60 attributes, which are finally used in the proposed identification scheme listed in Table 5.

4) DATASET GENERATION

Several pre-processing steps, such as the conversion of raw traffic into network traffic flows (CSV files), data truncation, and flows labeling generates a compact set of data series. The details of the tunneled and anonymous VoIP and Non_VoIP traffic flows instances are tabulated in Table 6. The dimension of the X$^{\text{Total}}$ dataset is (28438, 60), where 28438 is the total number of instances and 60 FSTFs set. After the network traffic flow generation, the X$^{\text{Total}}$ dataset is normalized to get a homogeneous dataset using Equation 6.

TABLE 6. Details of FSTFs based dataset.

VPN VoIP, TOR VoIP, VPN Non-VoIP, and TOR Non-VoIP instances				
	X^{Total}	X^{Train}	X^{Valid}	X^{Test}
No of instances	(28438, 60)	(18200, 60)	(4550, 60)	(5688, 60)

TABLE 7. Hyperparameters settings of MLP, 1D-CNN, and LSTM.

Hyperparameters	*MLP*	*1D-CNN*	*LSTM*
Batch Size	10	256	100
No. of Epochs	100	100	100
Dropout	0.01	0.1	0.1
L2 regularizer	/	0.0001	0.0001
Optimizer	adadelta	adamax	adam
Activation	softmax	softmax	softmax
No. of hidden layers	3(32, 16, 8)	3(64, 64, 64)	3(64, 32, 32)

$$X_{n(normalized)} = \frac{X_n - X_{min}}{X_{max} - X_{min}}, \quad n = 1, 2, 3, 4, \ldots.., N \quad (6)$$

where n is the number of instances in the dataset, X_{max} and X_{min} are the maximum and minimum values of the dataset entries. The computed normalized dataset is divided into training (X$^{\text{Train}}$), validation (X$^{\text{Valid}}$), and testing (X$^{\text{Test}}$) datasets. Initially, the X$^{\text{Total}}$ dataset splits into 80% of training data and 20% of testing data. For the validation purpose, we further split the training data by cross-fold validation into a ratio of 80:20. Furthermore, the proposed models are evaluated with an unseen X$^{\text{Test}}$ dataset to check the VoIP traffic predictive power in VPN and TOR networks. The details of the final X$^{\text{Total}}$, X$^{\text{Train}}$, X$^{\text{Valid}}$ and X$^{\text{Test}}$ datasets are given in Table 6.

B. EXPERIMENTAL ENVIRONMENT

The raw traffic (pcap files) are processed through the CICFlowmeter to generate network traffic bidirectional flows with FSTFs. To execute the experiments, the Python Keras platform is employed to build the deep learning framework for identification and detection purposes. We used the machine equipped with windows 64-bit OS with the system specifications of 2.4 GHz Intel Core-i3 CPU and 6 GB of Random Access Memory (RAM) to run the python codes. Python is used to evaluate the performance indexes and confusion matrix discussed in Section III (B) to assess the predictive ability of the identification engine. The training parameters configured in the VoIP detector for MLP, 1D-CNN, and LSTM are discussed in each section.

C. ARCHITECTURES OF THE PROPOSED DEEP LEARNING MODELS

Our goal is to characterize tunneled and anonymous network traffic into TOR VoIP/Non-VoIP and VPN VoIP/Non-VoIP network traffic. Figure. 1 presents the overall framework of the proposed work to accomplish the identification and finally detect the VoIP traffic. Different experiments are executed to

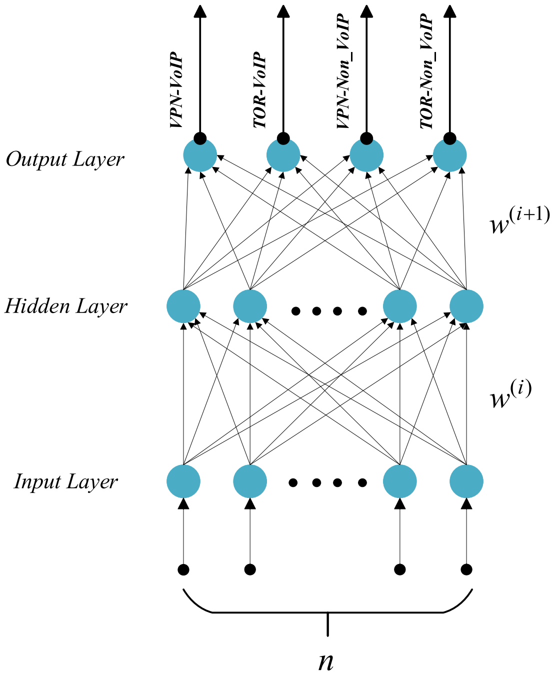

FIGURE 2. General architecture of an MLP with single hidden layer.

select the optimal values for the three deep learning methods. The selected hyperparameters are listed in Table 7.

The proposed deep learning architectures are briefly explained in this section.

1) MLP BASED IDENTIFIER

The architecture of the MLP model employed is shown in Figure 2. The proposed VoIP detector involves an input layer, three dense layers, and the output layer. All the three dense layers are composed of 32, 16, 8 neurons, respectively. The model is feed with a 1D-feature vector that contains 60 FSTFs listed in Table 5. The first dense layer has 32 neurons, which are fully connected to the input layer in the shape of 1×60. The rectified linear unit (ReLU) activation function is used throughout the three dense layers except for the output layer. Every dense layer is followed by a dropout regularizer to reduce overfitting and achieve a better generalization. The second and third dense layers are composed of 16 and 8 neurons, respectively. Finally, the output layer is composed of 4 neurons, where a softmax classifier is applied for final result labels. The model is trained with a categorical cross-entropy loss function. It is worth mentioning that the adadelta function is used as an optimizer. The optimizer is used to update the weights of the model. The training parameters are set as $\{BS{=}10, E_n{=}100\}$, where BS is the batch size, and E_n is the number of Epochs. The proposed MLP model exhibits satisfactory performance.

2) 1D-CNN BASED IDENTIFIER

1D-CNN shows better performance in 1D sequential data. We developed the tunneled and anonymous network traffic identifier using 1D-CNN. After the classification of these encrypted network traffic, the proposed model further detect VoIP traffic flows. The complete model structure is explained in this Section, and the model architecture is illustrated in Figure. 3. The proposed identification model consists of three convolutional layers, followed by max-pooling layers, dropout layers, a fully-connected dense layer a final output layer at the end of the model. Initially, the normalized dataset is fed to 1D-CNN based identifier. The first convolution layer processes the 1D-FSTFs vector as an input value. The first convolutional layer consists of 64 filters and the kernel size is 3. The result of this layer is inputted to the ReLU activation function followed by a 1D max-pooling layer with a pool size of 2 and stride 1. The 1D max-pooling layer is used to reduce the model parameter while keeping useful information. The ReLU activation function is kept the same all over the successive convolutional layers. The pooled layer is followed by the dropout layer to reduce the risk of overfitting. The output values of the dropout layer are processed by the identical second and third convolution layer composed of 64 filters. Each layer consists of a dropout layer for regularization purposes. L2 regularizer is employed in each convolutional layer to reduce overfitting and produce better identification results. The output of the third convolution layer is fed to the fully connected dense layer, which maps multiple abstracted features to a 1D tensor. The final layer used the softmax activation function to identify the type of the network traffic class. The training parameters are set as $\{BS{=}256, E_n{=}100\}$. During the identification process, the loss function is defined

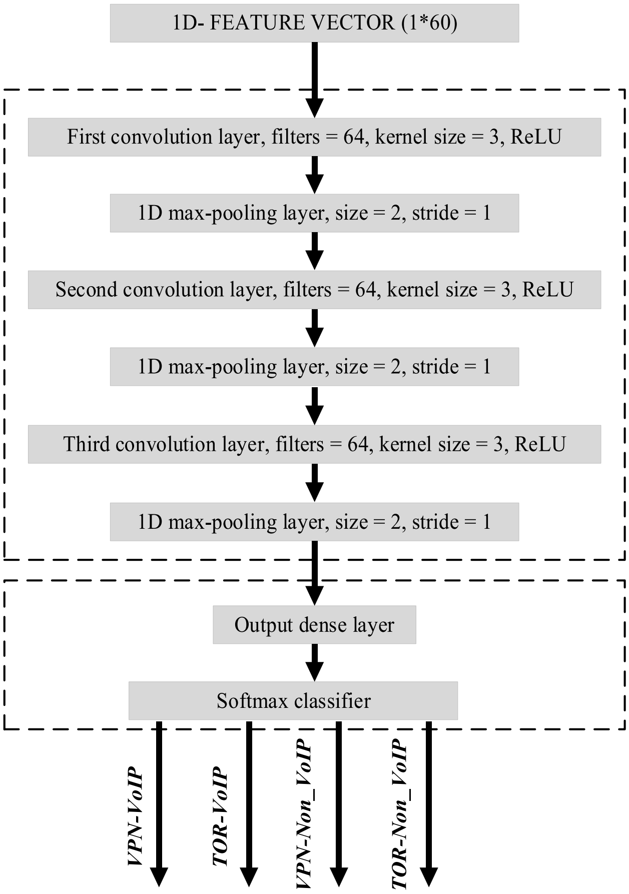
1D- FEATURE VECTOR (1*60)
First convolution layer, filters = 64, kernel size = 3, ReLU
1D max-pooling layer, size = 2, stride = 1
Second convolution layer, filters = 64, kernel size = 3, ReLU
1D max-pooling layer, size = 2, stride = 1
Third convolution layer, filters = 64, kernel size = 3, ReLU
1D max-pooling layer, size = 2, stride = 1
Output dense layer
Softmax classifier
VPN-VoIP
TOR-VoIP
VPN-Non_VoIP
TOR-Non_VoIP

FIGURE 3. The framework diagram of 1D-CNN based network identifier.

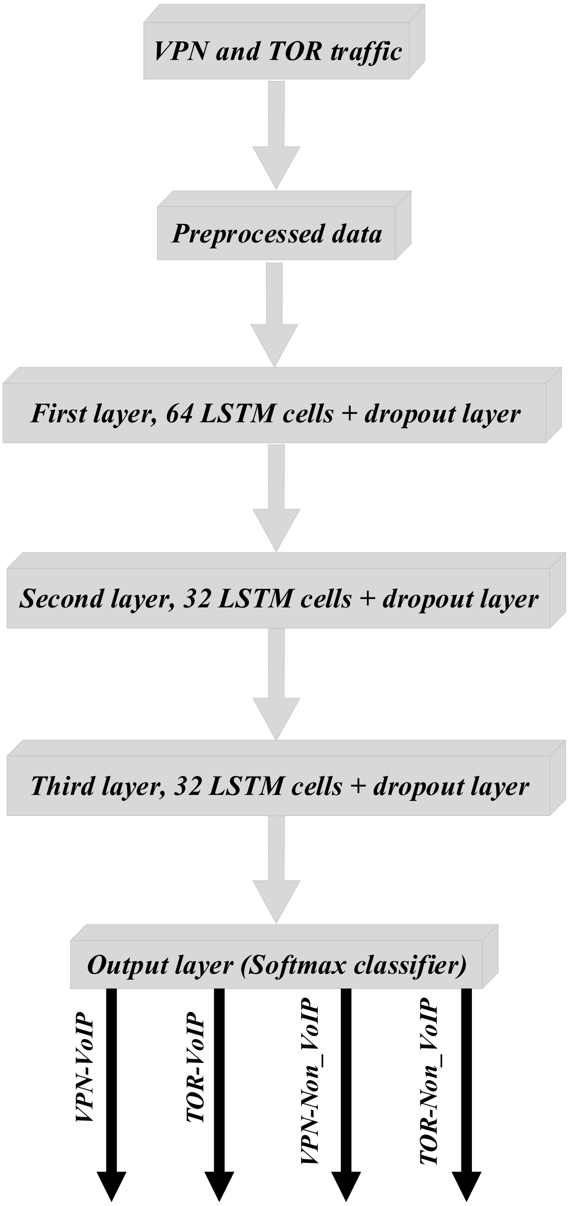
VPN and TOR traffic
Preprocessed data
First layer, 64 LSTM cells + dropout layer
Second layer, 32 LSTM cells + dropout layer
Third layer, 32 LSTM cells + dropout layer
Output layer (Softmax classifier)
VPN-VoIP
TOR-VoIP
VPN-Non_VoIP
TOR-Non_VoIP

FIGURE 4. The architecture of the proposed LSTM identifier.

by categorical cross-entropy and the adamax optimizer is selected for enhanced results.

3) LSTM BASED IDENTIFIER

Figure. 4 illustrates the overview of the proposed three-layered LSTM based identifier for VoIP traffic detection in VPN and TOR network traffic. The first LSTM layer contains 64 LSTM cells, which takes the 1D-feature vector (1 × 60) as an input. L2 regularization and dropout layer are applied in every LSTM layer for better results. The output of the first dropout layer is fed to the second LSTM layer with 32 LSTM cells. Each layer is followed by the same dropout layer. The second and third layers are identical in structure. Finally, the output of the last dropout layer is connected to the output softmax layer, which produces the probabilistic results for the four network traffic classes. L2 regularizer and the dropout layers are added to avoid model overfitting. The adam optimizer and categorical cross-entropy loss function achieved better results during the identification task. The training parameters are set as $\{BS{=}100, E_n{=}100\}$.

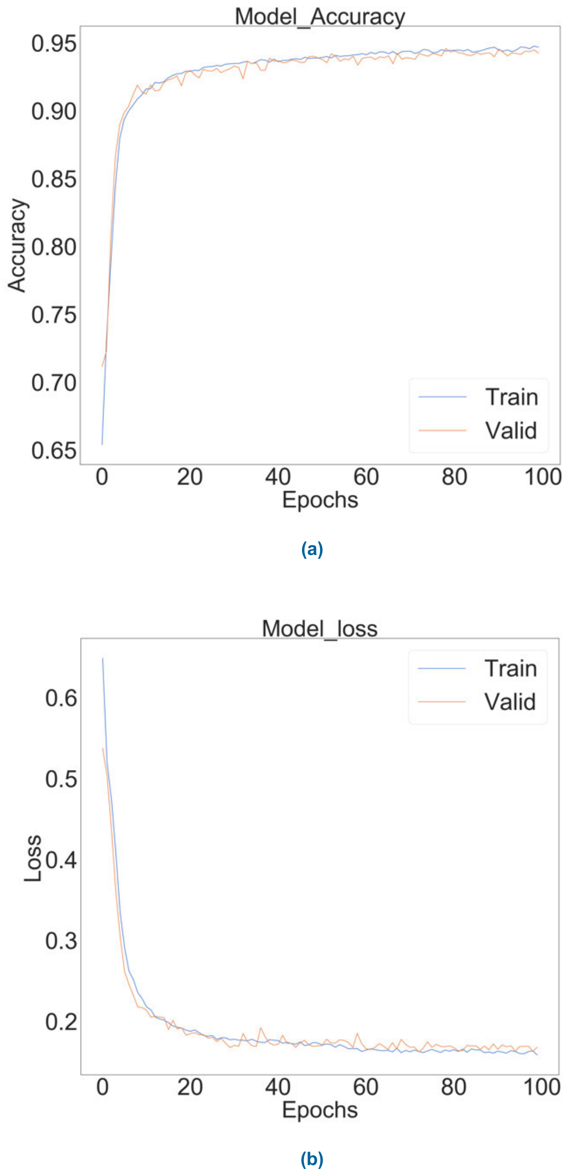

FIGURE 5. Accuracy and loss function illustration of the proposed MLP based identifier. (a) Accuracy of the training and validation phase. (b) Loss of the training and validation phase.

V. RESULTS AND DISCUSSION

This section gives a thorough analysis of deployed deep learning models to handle the multiclass identification and detection task. The experimental results reveal that the FSTFs are the effective set of attributes to successfully identify VPN and TOR network traffic and further characterize them into VoIP and Non-VoIP ones. The proposed DL models (MLP, 1D-CNN, and LSTM) outputs four different class outputs (VPN VoIP, TOR VoIP, VPN Non-VoIP, and TOR Non-VoIP). We will discuss the effectiveness of the above

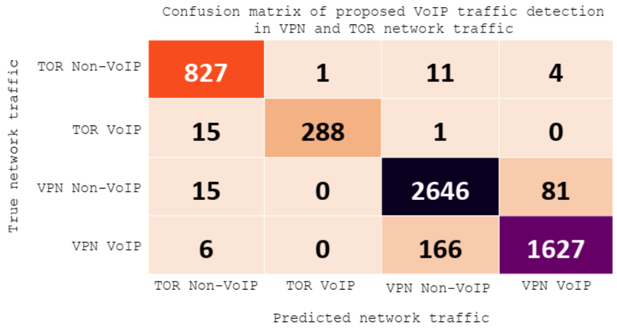

FIGURE 6. VoIP detection results of the MLP based network traffic identifier.

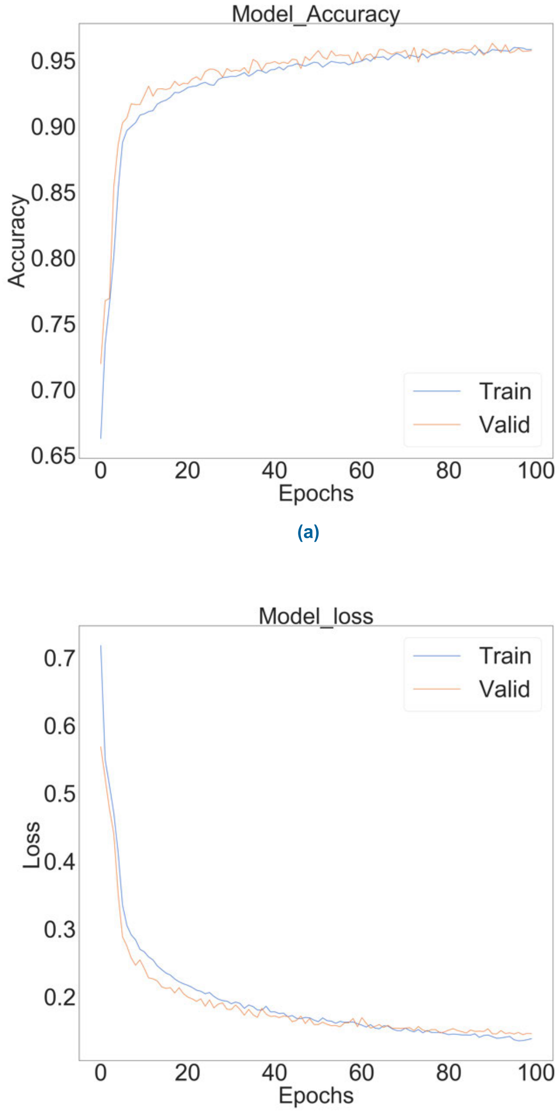

FIGURE 7. Accuracy and loss function illustration of the proposed 1D-CNN based identifier. (a) Accuracy of the training and validation phase. (b) Loss of the training and validation phase.

three deep learning models for the aforementioned problem. This section will evaluate the models in different aspects. Firstly, we will visualize the accuracy and loss of the training phase and the validation phase. Secondly, we will display the confusion matrix, which shows detailed identification results. Finally, we will list the macro-average predictive ability of the proposed system in terms of *Pr*, *Rc*, and *F-m* explained in Section III (B). During the training phase, a sufficient amount of data are fed into three DL-based (MLP, 1D-CNN, and LSTM) identification engines. We visualized the accuracy and loss of the training phase and validation phase (see Figure. 5). MLP based identifier achieved an accuracy of up to 94.7%. During the testing phase, X^{Test} dataset is employed for VoIP traffic detection among the mixed VPN and TOR network traffic. The test set contains unseen 5688 instances,

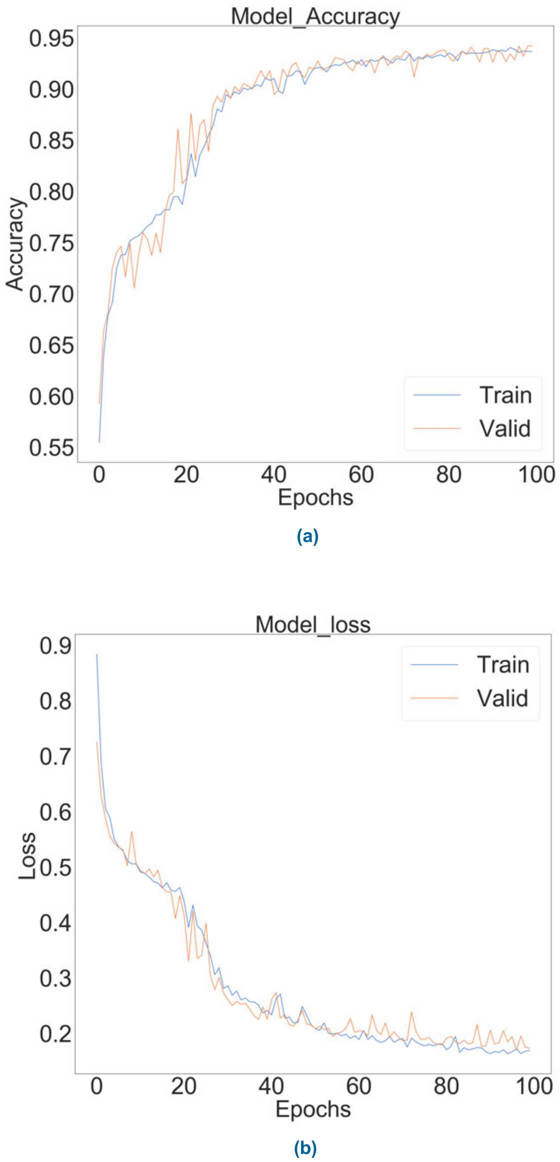

FIGURE 8. Accuracy and loss function illustration of the proposed LSTM based identifier. (a) Accuracy of the training and validation phase. (b) Loss of the training and validation phase.

TABLE 8. Macro-averaged predictive ability comparison of the proposed three deep learning based identification systems.

	Pr	*Rc*	*F-m*
MLP	0.96	0.95	0.954
1D-CNN	0.971	0.959	0.965
LSTM	0.952	0.947	0.949

TABLE 9. List of acronyms used in the manuscript.

Acronyms	*Definition*
AHD	Asymmetric high definition video
ASD	Asymmetric standard definition videos
CAIDA	Center for Applied Internet Data Analysis
CNN	Convolution neural network
C4.5	An extension of Quinlan's earlier ID3 algorithm
DBSCAN	Density-based spatial clustering of applications with noise
DPI	Deep packet inspection
EM	Expectation-Maximization
eDonkey	A peer-to-peer file sharing application
FSTF	Flow Spatio-Temporal Features
FLP	Flow latency period
FPR	False Positive Rate
FTP	File Transfer Protocol
FP	False Positive
FN	False Negative
GMM	Gaussian Mixture Model
GP	Genetic Programming
Gtalk	Google talk
Gnutella	A kind of peer-to-peer network
GBT	Gradient Boosting Tree
GCA	Gravitational clustering algorithm
HMM	Hidden Markov Model
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IANA	Internet Assigned Numbers Authority
ICQ	A cross-platform messenger and VoIP client
I2P	Invisible Internet Project (a decentralized anonymous network)
KNN	k-nearest neighbor
LSTM	Long-short term memory
LBNL	Lawrence Berkeley National Laboratory
L7	Application layer
ML	Machine learning
MLP	Multi-layer perceptron
MOGA	Multi-Objective Genetic Algorithm
MSN	Microsoft Network
NAPT	Network Address Port Translation
PCA	Principal Component Analysis
P2P	Peer-to-Peer
POP3	Post Office Protocol 3
PPTP	Point-to-Point Tunneling Protocol
PSD	Packet size distribution
QQ	Interactive video communication class
RNN	Recurrent neural network
RF	Random Forest
ReLU	Rectified linear unit
SAE	Stack Auto Encoder
SVM	Support vector machine
SSH	Secure Shell
SSL	Secure sockets layer
SMTP	Simple Mail Transfer Protocol
Sopcast	Network live TV
TC	Traffic classification
TOR	Onion router (a free and open-source software for enabling anonymous communication)
TCP	Transmission Control Protocol
TFTP	Trivial File Transfer Protocol
TLS	Transport Layer Security
TPR	True positive rate
TP	True Positive
TN	True Negative
UDP	User Datagram Protocol
UNIBS	University of Brescia, Italy
UNB-CIC	The University of New Brunswick, Canadian Institute of Cybersecurity
VPN	Virtual private network
VoIP	Voice over Internet Protocol
Xunlei	Peer-to-peer video data sharing

which is 20% of the total dataset. Figure. 6 depicts the detailed identification results based on the MLP identifier. The diagonal of the confusion matrix represents correctly identified network traffic flows.

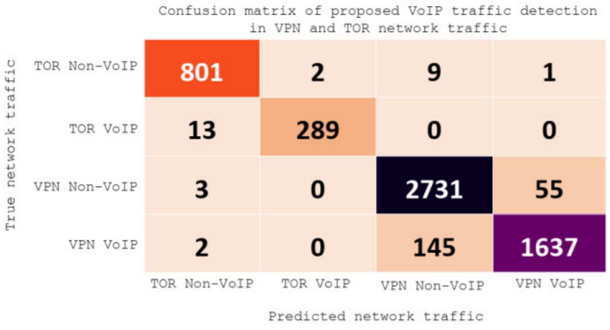

FIGURE 9. VoIP detection results of the 1D-CNN based network traffic identified.

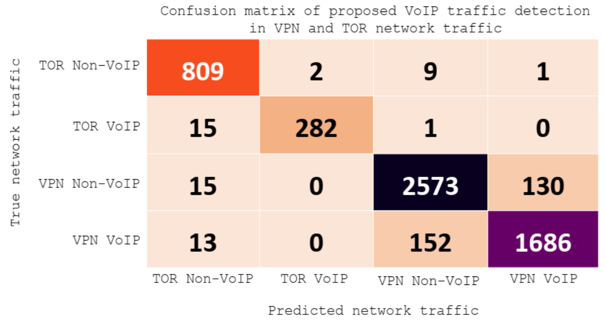

FIGURE 10. VoIP detection results of the LSTM based network traffic identified.

For better illustration, we computed the macro-averaged *Pr*, *Rc*, and *F-m* of the MLP based VoIP traffic detector and summarized in Table 8. Similarly, we computed the evaluation results (see Table 8) for 1D-CNN and LSTM based network traffic identifiers. The comparison Table of macro-averaged values of *Pr*, *Rc*, and*F-m* confirms the overall satisfactory performance of the deep learning models.

Figure. 7 and 8 illustrate the accuracy and loss of the 1D-CNN and LSTM based identifiers, respectively. 1D-CNN and LSTM achieved 96% and 94% accuracy, respectively. All the three deep learning models exhibits propitious convergence (see Figure. 5, Figure. 7, and Figure. 8). Figure. 9 and Figure. 10 visualized the confusion matrix to illustrate the overall performance of the proposed 1D-CNN and LSTM schemes. Overall the 1D-CNN is the better choice for the aforementioned problem. In summary, from the experimental results discussed above, we can conclude that all three state-of-art methods exhibit better identification and detection results.

VI. CONCLUSION

In today's life, network user's preferred to communicate through encrypted channels. Due to the enormous growth in tunneled and anonymous network usage, network management needs novel techniques to monitor it and analyze the network traffic. In this paper, we analyzed tunneled (VPN) and anonymous (TOR) network traffic based on deep learning techniques (MLP, 1D-CNN, and LSTM). Initially, we pre-processed the captured raw traffic and generated the dataset based on FSTFs with 15 sec of FLT. Furthermore, we employed the selected FSTFs to distinguish VPN network traffic from TOR network traffic. After the classification of

VPN and TOR network traffic, we characterized them into four different categories, i-e., VPN VoIP, VPN Non-VoIP, TOR VoIP, and TOR Non-VoIP. The experimental results show an accuracy level of (94.7%, 96%, and 94%) for MLP, 1D-CNN, and LSTM, respectively. The evaluation indexes indicate that the proposed scheme has a strong ability to distinguish VoIP traffic from Non-VoIP ones within the VPN and TOR networks. It is worthy to point out that all the study of this paper is based on the assumption that there are only one kind of flows, i.e. VoIP or Non-VoIP in the VPN or TOR traffic. It can be imagined that when multiple flows are mixed in one traffic, the accuracy will definitely decrease. The severity of the impact from other mixing flows and robustness of the identification under the mixed situation should be analyzed and studied in the future work. Extensive experiments are expected to be performed on bidirectional LSTMs and bidirectional GRUs in the future work.

REFERENCES

[1] T. Zink and M. Waldvogel, "BitTorrent traffic obfuscation: A chase towards semantic traffic identification," in *Proc. IEEE 12th Int. Conf. Peer Peer Comput. (P P)*, Sep. 2012, pp. 126–137.

[2] C. McCarthy and A. N. Zincir-Heywood, "An investigation on identifying SSL traffic," in *Proc. IEEE Symp. Comput. Intell. Secur. Defense Appl. (CISDA)*, Apr. 2011, pp. 115–122.

[3] J. Khalife, A. Hajjar, and J. Diaz-Verdejo, "A multilevel taxonomy and requirements for an optimal traffic-classification model," *Int. J. Netw. Manage.*, vol. 24, no. 2, pp. 101–120, Mar. 2014.

[4] Z. Cao, G. Xiong, Y. Zhao, Z. Z. Li, and L. Guo, "A survey on encrypted traffic classification," in *Proc. Int. Conf. Appl. Techn. Inf. Secur.* Berlin, Germany: Springer, 2014, pp. 73–81.

[5] T. Joe, E. Lear, A. Mankin, M. Kojo, K. Ono, M. Stiemerling, L. Eggert, A. Melnikov, W. Eddy, A. Zimmermann, B. Trammell, and J. Iyengar. (2013). *Service Name and Transport Protocol Port Number Registry*. The Internet Assigned Numbers Authority (IANA). [Online]. Available: http://www.iana.org/assignments/port-numbers

[6] P. Wang, X. Chen, F. Ye, and Z. Sun, "A survey of techniques for mobile service encrypted traffic classification using deep learning," *IEEE Access*, vol. 7, pp. 54024–54033, 2019.

[7] R. Alshammari and A. N. Zincir-Heywood, "Can encrypted traffic be identified without port numbers, IP addresses and payload inspection?" *Comput. Netw.*, vol. 55, no. 6, pp. 1326–1350, Apr. 2011.

[8] N. Namdev, S. Agrawal, and S. Silkari, "Recent advancement in machine learning based Internet traffic classification," *Procedia Comput. Sci.*, vol. 60, pp. 784–791, Jan. 2015.

[9] M. Finsterbusch, C. Richter, E. Rocha, J.-A. Müller, and K. Hanssgen, "A survey of payload-based traffic classification approaches," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 1135–1156, 2nd Quart., 2014.

[10] K. S. Shim, J. H. Ham, B. D. Sija, and M. Sup Kim, "Application traffic classification using payload size sequence signature," *Int. J. Netw. Manage.*, vol. 27, no. 5, p. e1981, 2017.

[11] R. Keralapura, A. Nucci, and C.-N. Chuah, "A novel self-learning architecture for p2p traffic classification in high speed networks," *Comput. Netw.*, vol. 54, no. 7, pp. 1055–1068, May 2010.

[12] J. Zhang, Y. Xiang, W. Zhou, and Y. Wang, "Unsupervised traffic classification using flow statistical properties and IP packet payload," *J. Comput. Syst. Sci.*, vol. 79, no. 5, pp. 573–585, Aug. 2013.

[13] Y. Wang, Y. Xiang, J. Zhang, W. Zhou, G. Wei, and L. T. Yang, "Internet traffic classification using constrained clustering," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 11, pp. 2932–2943, Nov. 2014.

[14] A. Este, F. Gringoli, and L. Salgarelli, "Support vector machines for TCP traffic classification," *Comput. Netw.*, vol. 53, no. 14, pp. 2476–2490, Sep. 2009.

[15] A. Finamore, M. Mellia, M. Meo, and D. Rossi, "KISS: Stochastic packet inspection classifier for UDP traffic," *IEEE/ACM Trans. Netw.*, vol. 18, no. 5, pp. 1505–1515, Oct. 2010.

[16] Z. Liu, H. Mingbo, L. Song, and W. Xin, "Research of P2P traffic comprehensive identification method," in *Proc. Int. Conf. Netw. Comput. Inf. Secur.*, vol. 1, 2011, pp. 307–310.

[17] T. T. T. Nguyen, G. Armitage, P. Branch, and S. Zander, "Timely and continuous machine-learning-based classification for interactive IP traffic," *IEEE/ACM Trans. Netw.*, vol. 20, no. 6, pp. 1880–1894, Dec. 2012.

[18] W. Ye and K. Cho, "Hybrid P2P traffic classification with heuristic rules and machine learning," *Soft Comput.*, vol. 18, no. 9, pp. 1815–1827, Sep. 2014.

[19] L. Peng, B. Yang, and Y. Chen, "Effective packet number for early stage Internet traffic identification," *Neurocomputing*, vol. 156, pp. 252–267, May 2015.

[20] Y.-N. Dong, J.-J. Zhao, and J. Jin, "Novel feature selection and classification of Internet video traffic based on a hierarchical scheme," *Comput. Netw.*, vol. 119, pp. 102–111, Jun. 2017.

[21] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Network traffic classifier with convolutional and recurrent neural networks for Internet of Things," *IEEE Access*, vol. 5, pp. 18042–18050, 2017.

[22] Z. Chen, K. He, J. Li, and Y. Geng, "Seq2Img: A sequence-to-image based approach towards IP traffic classification using convolutional neural networks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 1271–1276.

[23] M. Lotfollahi, M. J. Siavoshani, R. S. H. Zade, and M. Saberian, "Deep packet: A novel approach for encrypted traffic classification using deep learning," *Soft Comput.*, vol. 24, no. 3, pp. 1999–2012, Feb. 2020.

[24] K. L. Dias, M. A. Pongelupe, W. M. Caminhas, and L. de Errico, "An innovative approach for real-time network traffic classification," *Comput. Netw.*, vol. 158, pp. 143–157, Jul. 2019.

[25] A. J. Pinheiro, J. de M. Bezerra, C. A. P. Burgardt, and D. R. Campelo, "Identifying IoT devices and events based on packet length from encrypted traffic," *Comput. Commun.*, vol. 144, pp. 8–17, Aug. 2019.

[26] M. Conti, L. V. Mancini, R. Spolaor, and N. V. Verde, "Analyzing Android encrypted network traffic to identify user actions," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 114–125, Jan. 2016.

[27] Y. Fu, H. Xiong, X. Lu, J. Yang, and C. Chen, "Service usage classification with encrypted Internet traffic in mobile messaging apps," *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2851–2864, Nov. 2016.

[28] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "Multi-classification approaches for classifying mobile app traffic," *J. Netw. Comput. Appl.*, vol. 103, pp. 131–145, Feb. 2018.

[29] G. Draper-Gil, A. H. Lashkari, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of encrypted and VPN traffic using time-related features," in *Proc. 2nd Int. Conf. Inf. Syst. Secur. Privacy (ICISSP)*, 2016, pp. 407–414.

[30] K. Shahbar and A. N. Zincir-Heywood, "Packet momentum for identificationof anonymity networks," *J. Cyber Secur. Mobility*, vol. 6, no. 1, pp. 27–56, 2017.

[31] L. Bernaille, R. Teixeira, and K. Salamatian, "Early application identification," in *Proc. ACM CoNEXT Conf. (CoNEXT)*, 2006, pp. 1–12.

[32] C. V. Wright, F. Monrose, and G. M. Masson, "On inferring application protocol behaviors in encrypted network traffic," *J. Mach. Learn. Res.*, vol. 7, pp. 2745–2769, Dec. 2006.

[33] T. Auld, A. W. Moore, and S. F. Gull, "Bayesian neural networks for Internet traffic classification," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 223–239, Jan. 2007.

[34] W. Li, M. Canini, A. W. Moore, and R. Bolla, "Efficient application identification and the temporal and spatial stability of classification schema," *Comput. Netw.*, vol. 53, no. 6, pp. 790–809, Apr. 2009.

[35] C. Bacquet, K. Gumus, D. Tizer, A. N. Zincir-Heywood, and M. I. Heywood, "A comparison of unsupervised learning techniques for encrypted traffic identification," *J. Inf. Assurance Secur.*, vol. 5, no. 1, pp. 464–472, 2010.

[36] G.-L. Sun, Y. Xue, Y. Dong, D. Wang, and C. Li, "An novel hybrid method for effectively classifying encrypted traffic," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2010, pp. 1–5.

[37] N.-F. Huang, G.-Y. Jai, H.-C. Chao, Y.-J. Tzang, and H.-Y. Chang, "Application traffic classification at the early stage by characterizing application rounds," *Inf. Sci.*, vol. 232, pp. 130–142, May 2013.

[38] J. Zhang, X. Chen, Y. Xiang, W. Zhou, and J. Wu, "Robust network traffic classification," *IEEE/ACM Trans. Netw.*, vol. 23, no. 4, pp. 1257–1270, Aug. 2015.

[39] S. Bagui, X. Fang, E. Kalaimannan, S. C. Bagui, and J. Sheehan, "Comparison of machine-learning algorithms for classification of VPN network traffic flow using time-related features," *J. Cyber Secur. Technol.*, vol. 1, no. 2, pp. 108–126, Apr. 2017.

[40] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, "End-to-end encrypted traffic classification with one-dimensional convolution neural networks," in *Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI)*, Jul. 2017, pp. 43–48.

[41] A. H. Lashkari, G. Draper-Gil, M. S. I. Mamun, and A. A. Ghorbani, ''Characterization of tor traffic using time based features,'' in *Proc. Int. Conf. Inf. Syst. Secur. Privacy (ICISSp)*, 2017, pp. 253–262.

[42] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, ''The WEKA data mining software: An update,'' *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.

[43] Z. Deng, G. Qian, Z. Chen, and H. Su, ''Identifying tor anonymous traffic based on gravitational clustering analysis,'' in *Proc. 9th Int. Conf. Intell. Hum.-Mach. Syst. Cybern. (IHMSC)*, vol. 2, Aug. 2017, pp. 79–83.

[44] K. Shahbar and A. N. Zincir-Heywood, ''Effects of shared bandwidth on anonymity of the I2P network users,'' in *Proc. IEEE Secur. Privacy Workshops (SPW)*, May 2017, pp. 235–240.

[45] H. Huang, H. Deng, J. Chen, L. Han, and W. Wang, ''Automatic multi-task learning system for abnormal network traffic detection,'' *Int. J. Emerg. Technol. Learn.*, vol. 13, no. 4, p. 4, Mar. 2018.

[46] A. Montieri, D. Ciuonzo, G. Aceto, and A. Pescape, ''Anonymity services tor, I2P, JonDonym: Classifying in the dark (Web),'' *IEEE Trans. Dependable Secure Comput.*, vol. 17, no. 3, pp. 662–675, May 2020.

[47] M. Kim and A. Anpalagan, ''Tor traffic classification from raw packet header using convolutional neural network,'' in *Proc. 1st IEEE Int. Conf. Knowl. Innov. Invention (ICKII)*, Jul. 2018, pp. 187–190.

[48] H. Yao, C. Liu, P. Zhang, S. Wu, C. Jiang, and S. Yu, ''Identification of encrypted traffic through attention mechanism based long short term memory,'' *IEEE Trans. Big Data*, early access, Sep. 20, 2019, doi: 10.1109/TBDATA.2019.2940675.

[49] Y. Zeng, H. Gu, W. Wei, and Y. Guo, ''Deep-full-range: A deep learning based network encrypted traffic classification and intrusion detection framework,'' *IEEE Access*, vol. 7, pp. 45182–45190, 2019.

[50] A. Shiravi, H. Shiravi, M. Tavallaee, and A. A. Ghorbani, ''Toward developing a systematic approach to generate benchmark datasets for intrusion detection,'' *Comput. Secur.*, vol. 31, no. 3, pp. 357–374, May 2012.

[51] A. Montieri, D. Ciuonzo, G. Bovenzi, V. Persico, and A. Pescape, ''A dive into the dark Web: Hierarchical traffic classification of anonymity tools,'' *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 1043–1054, Jul. 2020.

[52] G. Bovenzi, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescape, ''A big data-enabled hierarchical framework for traffic classification,'' *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 4, pp. 2608–2619, Oct. 2020.

[53] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, ''DISTILLER: Encrypted traffic classification via multimodal multitask deep learning,'' *J. Netw. Comput. Appl.*, Jan. 2021, Art. no. 102985.

[54] Y. Okada, S. Ata, N. Nakamura, Y. Nakahira, and I. Oka, ''Application identification from encrypted traffic based on characteristic changes by encryption,'' in *Proc. IEEE Int. Workshop Tech. Committee Commun. Qual. Rel. (CQR)*, May 2011, pp. 1–6.

[55] Y. Kumano, S. Ata, N. Nakamura, Y. Nakahira, and I. Oka, ''Towards real-time processing for application identification of encrypted traffic,'' in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, 2014, pp. 136–140.

[56] A. Cuzzocrea, F. Martinelli, F. Mercaldo, and G. Vercelli, ''Tor traffic analysis and detection via machine learning techniques,'' in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 4474–4480.

[57] R. Alshammari and A. N. Zincir-Heywood, ''Identification of VoIP encrypted traffic using a machine learning approach,'' *J. King Saud Univ. Comput. Inf. Sci.*, vol. 27, no. 1, pp. 77–92, Jan. 2015.

[58] D. Arndt. (2011). *How to: Calculating Flow Statistics Using Netmate*. [Online]. Available: http://dan.arndt.ca/nims/calculating-flow-statistics-using-netmate/

[59] T. Qin, L. Wang, Z. Liu, and X. Guan, ''Robust application identification methods for P2P and VoIP traffic classification in backbone networks,'' *Knowl.-Based Syst.*, vol. 82, pp. 152–162, Jul. 2015.

[60] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, ''Exploiting encrypted and tunneled multimedia calls in high-speed big data environment,'' *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4959–4984, Feb. 2018.

[61] I. Fauzia and U. A. Khan, ''A generic technique for voice over Internet protocol (VoIP) traffic detection,'' *Int. J. Comput. Sci. Netw. Secur.*, vol. 8, no. 2, pp. 52–59, 2008.

[62] Y.-D. Lin, C.-N. Lu, Y.-C. Lai, W.-H. Peng, and P.-C. Lin, ''Application classification using packet size distribution and port association,'' *J. Netw. Comput. Appl.*, vol. 32, no. 5, pp. 1023–1030, Sep. 2009.

[63] B. Li, M. Ma, and Z. Jin, ''A VoIP traffic identification scheme based on host and flow behavior analysis,'' *J. Netw. Syst. Manage.*, vol. 19, no. 1, pp. 111–129, Mar. 2011.

[64] D. Adami, C. Callegari, S. Giordano, M. Pagano, and T. Pepe, ''Skype-Hunter: A real-time system for the detection and classification of Skype traffic,'' *Int. J. Commun. Syst.*, vol. 25, no. 3, pp. 386–403, Mar. 2012.

[65] T. Kavzoglu and P. M. Mather, ''The use of backpropagating artificial neural networks in land cover classification,'' *Int. J. Remote Sens.*, vol. 24, no. 23, pp. 4907–4938, Jan. 2003.

[66] S. Hochreiter and J. Schmidhuber, ''Long short-term memory,'' *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[67] F. Ul Islam, G. Liu, and W. Liu, ''Identifying VoIP traffic in VPN tunnel via flow spatio-temporal features,'' *Math. Biosci. Eng.*, vol. 17, no. 5, pp. 4747–4772, 2020.

[68] L. A. Habibi, G. Draper-Gil, M. S. Mamun, and A. A. Ghorbani. (2017). *CICFlowMeter: Network Traffic Flow Generator and Analyser*. [Online]. Available: https://www.unb.ca/cic/research/applications.html

FAIZ UL ISLAM received the B.S. degree in electronic engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2013, and the M.S. degree in control theory and control engineering from the Nanjing University of Science and Technology, China, in 2017, where he is currently pursuing the Ph.D. degree in control science and engineering. His research interests include encrypted network traffic analysis, VoIP traffic analysis, and chaos-based cryptosystem.

GUANGJIE LIU received the B.S. degree in information engineering and the Ph.D. degree in control science and engineering from the Nanjing University of Science and Technology, Nanjing, in 2002 and 2007, respectively. From 2016 to 2017, he was a Visiting Scholar with the Department of Computer Science, University of California at Davis, Davis, CA, USA. He is currently a Professor with the Nanjing University of Information Science and Technology. His research interests include networks and communication security.

JIANGTAO ZHAI received the B.S. degree in electrical and information engineering and the M.S. and Ph.D. degrees in control science and engineering from the Nanjing University of Science and Technology, Nanjing, in 2006, 2008, and 2013 respectively. From 2013 to 2019, he worked with the Jiangsu University of Science and Technology. He is currently an Associate Professor with the Nanjing University of Information Science and Technology. His research interests include multimedia communication and network security.

WEIWEI LIU (Member, IEEE) received the B.S. degree in automation and the Ph.D. degree in control science and engineering from the Nanjing University of Science and Technology, Nanjing, in 2010 and 2015, respectively. From 2014 to 2015, he was a Visiting Scholar with the Department of Computer Science, University of California at Davis, Davis, CA, USA. He is currently an Associate Professor with the School of Automation, Nanjing University of Science and Technology. His research interests include multimedia signal processing and network traffic analysis. He has published more than 30 articles in these areas, including the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and *Annals of Telecommunications*. He is an Active Reviewer of several journals, including *Digital Signal Processing and Security* and *Communication Networks*.