Improve the DFI-based Network Traffic Classification Performance by Using QoS Metrics

Yongcheng Zhou and Anguo Zhang*

***Abstract*—Network traffic classification methods based on network flow characteristics and machine learning classifiers have received extensive attention in academia. However, in actual industrial applications, the current mainstream flow identification engines, especially commercial engines, still mainly adopt port-based and deep packet inspection (DPI)-based network traffic dentification methods, deep flow inspection (DFI) has not been officially promoted yet. In addition to the fact that causal reasoning of general machine learning classifiers is difficult to analyze, another big reason is that the machine learning classifiers in most of the current research results are difficult to work well in different Internet network situations after training on a training set. Through experiments, we found that the basic QoS parameters of the network, such as packet loss rate, transmission delay, throughput rate and network jitter, in addition to being able to describe the performance state of the current network, will further affect some DFI features of the network flow. In this paper, we do not try to come up with completely new traffic classification features or completely new classifiers, but rather try to make some small improvements on the existing DFI methods so that the DFI classifiers can work precisely and robustly under different network topologies and network QoS parameters. Experimental results in different network environments show that these additional QoS parameters can significantly improve the robustness of the existing DFI machine learning classifiers.**

***Index Terms*—Network traffic classification, network QoS metrics, deep flow inspection.**

I. Introduction

Due to the rapid development of Internet technology and the increasing requirements of network security, trend analysis and user experience promotion, network traffic identification has been an important and attractive issue in recent years. With network traffic classification technique, network servers/administrators can obtain the current network status, especially the critical applications, services and user behaviors such like daily usage, anomaly behaviors [1]. Therefore, quality of service (QoS) can be achieved through traffic classification, which is the one of the most concerned issues for content providers and network administrators.

Network traffic identification technology has gone through three stages of development, namely, port-based methods, deep packet inspection (DPI) based methods and deep flow inspection (DFI) based methods. Port-based technology is the earliest and simplest approach for researchers and engineers to identify the network flows, i.e. by directly detecting the ports from the packet header which taking into account the fact that most traditional applications use IANA to allocate standard ports. However, since about 2002, more and more applications adopted dynamic ports for network communication to avoid the problem of port occupation. It led to the failure of classic port-detection based method [2]. Under this situation, DPI-based methods were proposed. Just as the name indicates, DPI methods need to inspect the payload of every packet to identify network traffic. Although the DPI-based methods are still widely used in various applications, their drawbacks are also obvious, i.e. (1) they require plenty of continuous human efforts to build and update the feature library, which is highly costly; (2) these methods are not capable to detect the encrypted network traffic; and (3) these methods violate users' privacy due to the need of detecting the payload of the packets.

In recent years, Deep flow inspection (DFI) has been considered as a promising and effective method, and it seems to be capable to address the aforementioned problems. In the DFI-based network traffic identification tasks, various port independent and payload-independent packet or flow level statistical characteristics (features) are generally extracted for different classifier models in research and application works, such as packet number related, packet size related, and inter packet time related characteristics. To handle these large feature sets, machine learning algorithms, such as artificial neural networks (ANNs) [3]-[5], support vector machines (SVMs), decision tree [6], random forest (RF) [7] have been widely employed. Although there are many works demonstrate that the machine learning algorithms have achieved high accuracy in terms of network traffic classification, it still needs to be pointed out that in most works, the machine learning models are trained and tested in the same and relative stable network conditions. Under such conditions, for a certain type of network traffic flow, the impact of the network itself on the data packet flow is approximately the same. Thus, the difference of the flow behaviors is largely due to the essential attributes of different flow types. However, in practical applications, the performance of the communication network carrying the data itself may fluctuate greatly, or when we try to apply a machine learning model of unified training to different network nodes, because their network performance is different, even for the same kind of traffic flow, the behavior under different network parameters will behave differently. In this case, the classification accuracy of the machine learning model will be greatly reduced. In the test, we found

Manuscript received October 12, 2020; revised December 15, 2020.

Zhou Yongcheng is with Ruijie Networks Co., Ltd, Fuzhou 350002, China (e-mail: zhouyongcheng_024@163.com).

Zhang Anguo is with College of Physics and information Engineering, Fuzhou University, Fuzhou 350108, China, and the Key Laboratory of Medical Instrumentation and Pharmaceutical Technology of Fujian Province, Fuzhou 350116, China (Corresponding author, e-mail: anguo.zhang@hotmail.com).

doi: 10.18178/jacn.2020.8.2.278

that some network QoS parameters can effectively describe the current network performance. Adding these QoS parameters to the DFI feature set can make the machine learning model better adapt to different network performance and different network nodes.

The rest of this paper is organized as follows: Section II reviews related works of network traffic classification. In section III, we will discuss the impacts of QoS parameters to network performance. The experiments details will be given in section IV. We will summarize the whole paper in section V, and give the conclusions.

II. Related Works

Feature selection is a very important task prior to building classifier models, and the quality of feature set has a huge impact on the classification accuracy. In general, irrelevant and redundant features will cause the "dimensional disaster" problem for classifiers (especially the machine learning algorithm based classifiers), which may have a great negative influence on the accuracy. So, in most DFI-based network traffic classification works, researchers and engineers always tending to find the smallest but still effective feature set for their own methods.

In [8]-[10], Zhang *et al.* proposed a series of semi-supervised methods to obtain effective network traffic classifiers and tackle the problem of unknown applications just with a small labeled training dataset. [11] adopted the training method and clustering algorithm to detect unknown applications and extend labeled flows from a few labeled and many unlabeled flows. It was demonstrated that all the above-mentioned works had achieved the state-of-the-art classification performance in terms of accuracy and robustness. And what should be pointed out is that in the flow statistical features used in [8]-[11] are all the same as Table I shows.

TABLE I: Network Flow Statistical Features of [8–11]

Category of features	Description of feature	Number of feature
Packets	Number of packets transferred in bi-direction	2
Bytes	Volume of bytes transferred in bi-direction	2
Packets Size	Min, Max, Mean and Standard deviation of packets size in the unidirection	8
Inter-Packet Time	Min, Max, Mean and Standard deviation of inter packet time in unidirection	8
	Total	**20**

[12] proposed an improved random forest model by setting the variable selection probability according to the importance of the corresponding variable. And their experimental results showed that their method performed better classification accuracy and took less time to build the classifier model. The features used in [12] are presented as Table II.

In [13], Fan *et al.* measured the worth of 248 features by evaluating the information gain with respect to the traffic categories, and each feature was extracted from the traffic flow behavior. Eventually, they selected 30 features as the most suitable subset for their machine learning classifiers. The most valuable 30 features are listed in Table III.

TABLE II: Network Flow Statistical Features of [12]

Category of features	Description of feature	Number of feature
Tuples	IP address and protocol version, TCP ports in bi-direction	5
Flags	TCP flags	3
Packets Related	Max, Mean and Variance of packets, and others	12
Bytes Related	Number of bytes, and others	2
TTL Related	Mean, Variance of TTL, and others	4
Inter-Packet Time	Max, Mean and Variance of inter packet time	3
	Total	**29**

TABLE III: Network Flow Statistical Features of [13]

Category of features	Description of feature	Number of feature
Tuples	IP address and protocol version, TCP ports in bi-direction	2
Packets Related	Num of packets	3
Bytes Related	Min, Max, Mean of segment size	15
Segment Size Related	Min, Mean of window advertisement size	5
Window Size	Max, Mean and Variance of inter packet time	3
Time Related	round-trip and packet inter-arrival time	2
	Total	**30**

Ref. [14] used only the first 20 packets exchanged in a flow lifetime, and six selected features were extracted from the packets' header, thus, a feature matrix of 20×6 dimension is generated to be the input of LSTM (Long Short-Term Memory Neural Network)and CNN (Convolutional Neural Network) combined deep learning classifier. It was also reported that the proposed model provided better detection results than alternative algorithms without any feature engineering. And the 6 selected features of each packet are presented in Table IV.

TABLE IV: Network Flow Statistical Features of [14]

Category of features	Description of feature	Number of feature
Tuples	TCP ports in bi-direction	2
Bytes Related	Num of payload bytes	1
Window Size	TCP window size	1
Time Related	inter-arrival time and direction (0/1)	2
	Total	**6**

For identifying the traffic types of VPNs (virtual private networks), Zeng *et al.* used 12 features which extracted from flow context, behavior of source-side hosts on flow and DNS (Domain Name System) behavior of source-side host on DNS respectively to construct the detection model [15].

There are also plenty of other works have been done on the DFI-based network traffic classification. Considering that identifying individual application is of high importance, [16] concerned to identify the popular end-user applications. The authors reduced the number of features to only 12, while still maintained high classification accuracy. Although it is

widely reported that DFI-based approach to network traffic identification can achieve state-of-the-art accuracy only using the flow behavior pattern, we find there exists a serious problem when applying the DFI-based approach to practical engineering, that is, the accuracy of traffic identification will fluctuate greatly under different network environments. Obviously, network parameters have an impact on the final accuracy.

TABLE V: NETWORK FLOW STATISTICAL FEATURES OF [15]

Category of features	Description of feature	Number of feature
Flow Context	The number of flows having flow correlation with the flow.	5
Host Behavior on Flow	The max, sum of Flow Burst length, etc.	5
Host Behavior on DNS	The number of sensitive, unassociated domain names requested by the source-side host of the flow.	2
	Total	**12**

A. Robustness to Network Conditions

Network conditions, such as congestion, fragmentation, delay, retransmissions, duplications and packet losses are inherently different among different communication networks. [17] considered that the Application Protocol Data Units (APDUs) are vulnerable by the network's side effects, as well as the fact that the zero-length packets are frequent (approximately, zero-length packets comprise 33% of the TCP packets in count, while only comprise 2-3% in volume of bytes), therefore, the extracted TCP flow attributes are only sampled from zero-length packets, i.e., the packets contain control bits, but do not contain any payload (e.g., SYN,ACK, etc.).

III. NETWORK METRIC PARAMETERS ANALYSIS

In this section, we conducted a series of experiments to analysis the impacts of four main network QoS parameters, namely, packet delay, packet loss rate, throughput and packet delay variation/jitter, on the traffic flow characteristics. We tried to verify whether the changes in these QoS parameters would affect network traffic behaviors, or even cause the behaviors fundamentally change.

A. Packet Delay

Packet delay is also called the end-to-end delay that refers to the time taken for packets to be transmitted across a network from source to destination. The exact time delay between two points A and B of an IP network can be measured by using the synchronized clock: A records a timestamp on a packet and then send it to B, B obtains the timestamp when receiving the packet and calculates the difference to current timestamp. Packet delay mainly includes transmission delay, propagation delay, process delay and queueing delay.

B. Packet Loss Rate

Packet loss is measured as a percentage of packets lost with respect to packets sent. It occurs when one or more packets of data travelling across a network but fail to reach their destination. Packet loss is either caused by errors in data transmission [18], typically across wireless networks or network congestion. Network congestion affects all types of networks. When content arrives for a sustained period at a given router or network segment at a rate greater than it is possible to send through, there is no option than to drop packets. If a single router or link is constraining the capacity of the complete travel path or of network travel in general, it is known as a bottleneck. In some cases, packets are intentionally dropped by routing routines. Packet loss can also be caused by a packet drop attack. The Transmission Control Protocol (TCP) detects packet loss and performs retransmissions to ensure reliable messaging. Packet loss in a TCP connection is also used to avoid congestion and thus produces an intentionally reduced throughput for the connection. In streaming media and online game applications, packet loss can affect a user's quality of experience (QoE).

C. Throughput

In general terms, throughput is the maximum rate of production or the maximum rate at which something can be processed. When used in the context of communication networks, throughput or network throughput is the rate of successful message delivery over a communication channel. The data these messages belong to may be delivered over a physical or logical link, or it can pass through a certain network node. Throughput is usually measured in bits per second (bit/s or bps), and sometimes in data packets per second (p/s or pps) or data packets per time slot. The system throughput or aggregate throughput is the sum of the data rates that are delivered to all terminals in a network, which is essentially synonymous to digital bandwidth consumption; it can be analyzed mathematically by applying the queueing theory, where the load in packets per time unit is denoted as the arrival rate (l), and the throughput, where the drop in packets per time unit, is denoted as the departure rate (m) [19].

D. Packet Delay Variation/Jitter

In computer networking, packet delay variation (PDV) is the difference in end-to-end-way delay between selected packets in a flow with any lost packets being ignored [20]. The effect is sometimes referred to as jitter, although the definition is an imprecise fit. Internet protocol services performance is strongly influenced by the values assumed by the Packet Delay Variation.

IV. RESULTS

The detail analysis in Section III demonstrated that QoS metrics can significantly affect the behavior of network flows, which may disturb the accuracy of DFI-based network traffic identification model. In order to obtain a robust ma- chine learning model that works well under various network conditions, the QoS metrics are also considered to be part of input features of machine learning model.

To illustrate the effectiveness of additional machine learning features consisting of QoS metrics, we compare the traffic classification accuracy of models before and after adding these additional features.

A. Effects of QoS Metric to Classification Accuracy

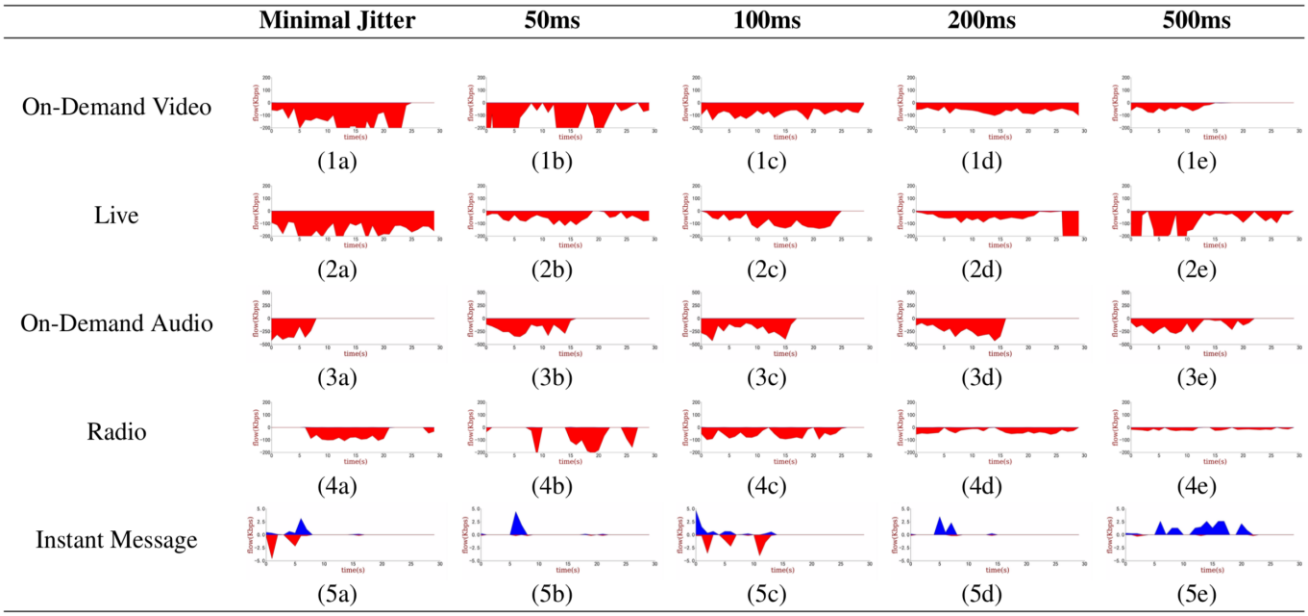


Fig. 1. Flow behavior changes of different network traffics with transmission delay varying.

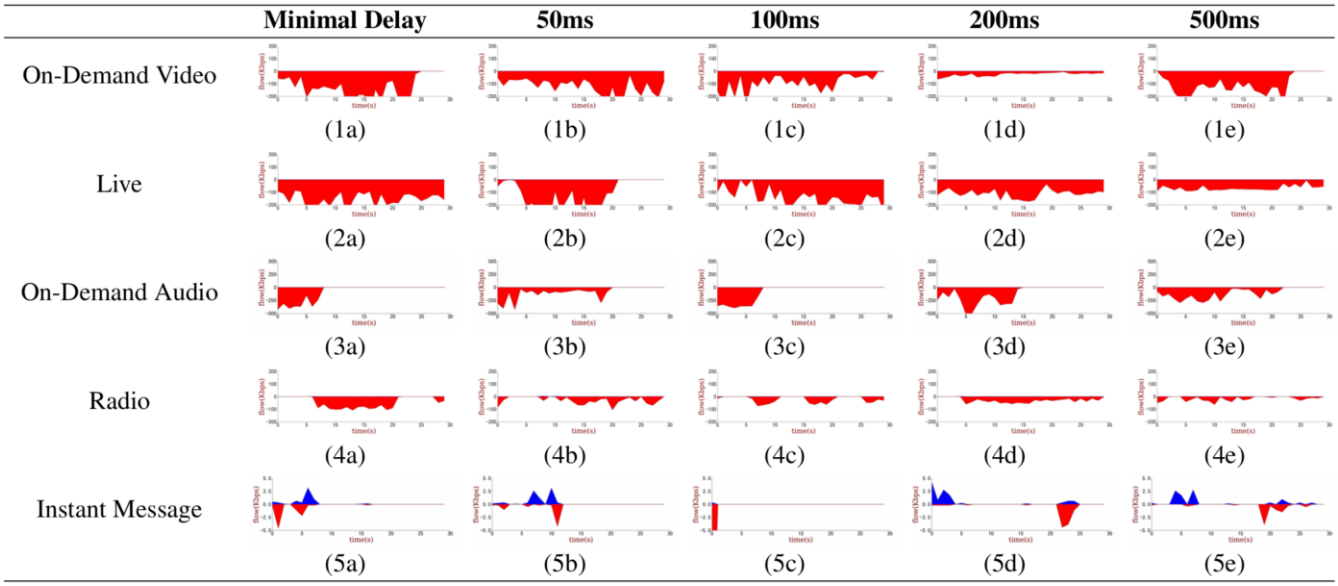


Fig. 2. Flow behavior changes of different network traffics with loss rate varying.

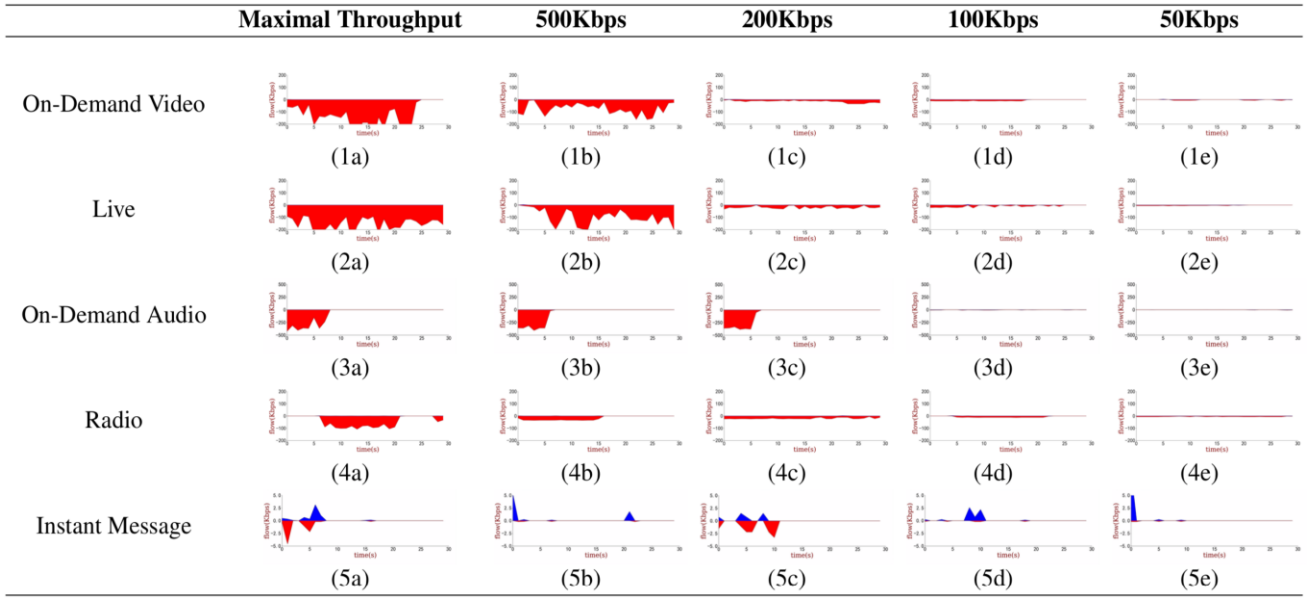


Fig. 3. Flow behavior changes of different network traffics with throughput rate varying.

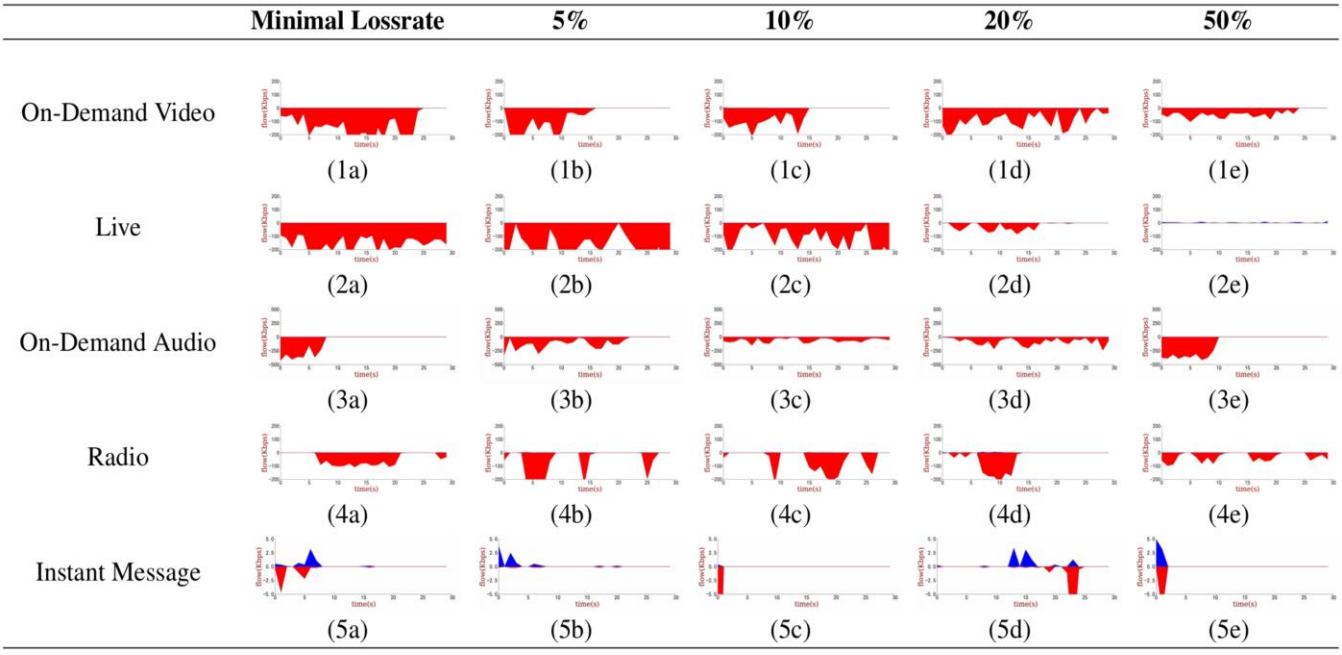


Fig. 4. Flow behavior changes of different network traffics with jitter varying.

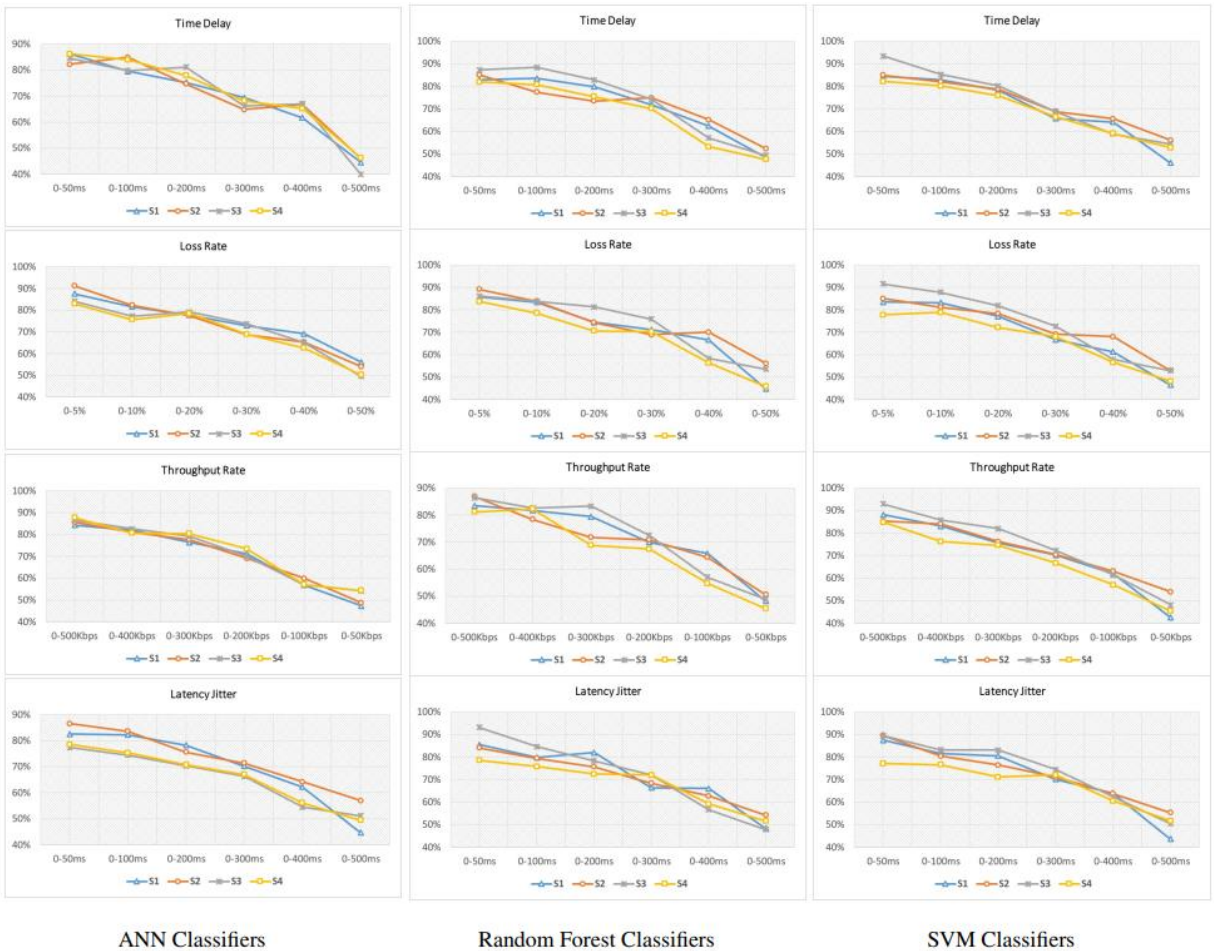


Fig. 5. The traffic classification accuracies of the four network data collection points (marked as S1, S2, S3, and S4) vary with the fluctuation of the network performance.

From Fig. 1 to Fig. 4, we can visually see that different traffic flow behaviors will change differently when the network performance changes. Correspondingly, in Fig. 5, we test the classification accuracy of network traffics of which data is mixed with different network performance metrics (i.e., transmission delay, loss rate, throughput rate, and network jitter). In order to more reliably verify the impact of network performance jitter on traffic classification accuracy, we selected four popular machine learning models as the classifiers, including artificial neural network (ANN), random forest, support vector machine (SVM) and K-nearest neighbor (KNN). It should be pointed out that the

classification strategies of these four machine learning models are also representative, where ANN uses the connectionism between the nodes among the previous and latter layers, random forest adopts the idea of ensemble learning and uses in- formation entropy to construct subtree as the basic classifier- s, SVM uses the high-dimensional plane to perform hyper- plane division on the multi-dimensional feature data, while KNN uses the distance between samples to classify.

Fig. 5 clearly shows that at all the four data collection points (S1, S2, S3 and S4), the classification accuracy will be reduced due to network performance fluctuation (or inconsistent transmission performance of the network data stream), and the greater the range of network performance fluctuation, the more the classification accuracy is reduced. This indicates that network performance fluctuation can cause a greater degree of damage to the accuracy of traffic classification based on the DFI method, which also confirms our concern.

B. Accuracy Comparison

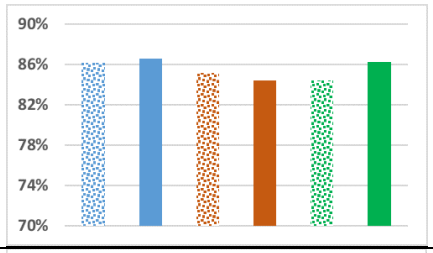
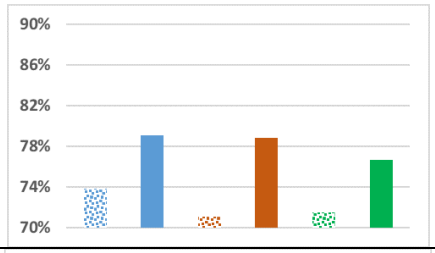
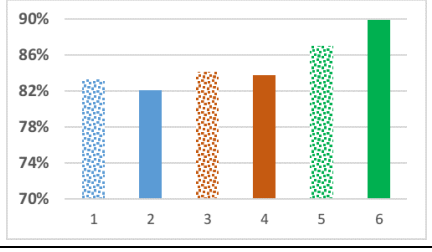
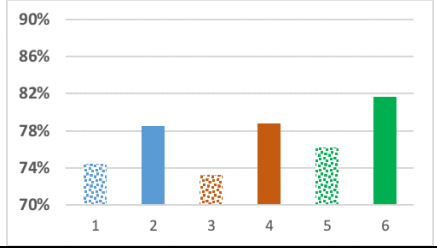
In Section II, we introduced the different selections of DFI-based feature sets in some other works. Here, we compared the traffic classification accuracy of these feature sets with QoS metrics (transmission delay, loss rate, throughput rate, and network jitter) added in the communication networks with performance fluctuations. Similarly, we also selected ANN, SVM and random forest (RF) as the basic classification models. The measurement ranges of transmission delay, loss rate, throughput rate and network jitter are de- signed as [0, 500ms], [0, 1], [0, 500Kbps] and [0, 500ms], respectively. When the actual value of these QoS features exceeds the upper bound of the corresponding measurement range, it is directly set to the upper bound value. Then, the values of these additional QoS features are normalized to between 0 and 1.

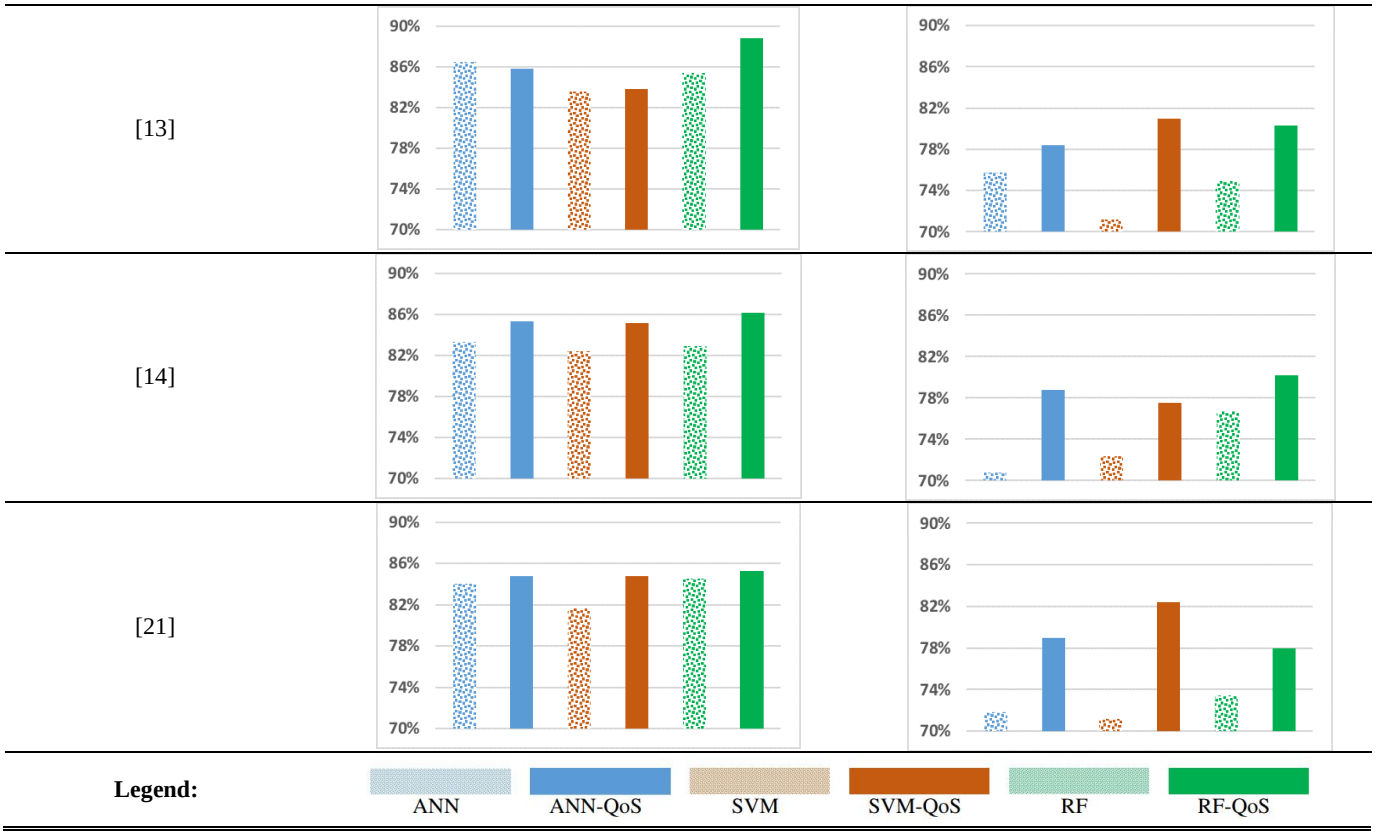
Table VI shows the classification accuracy of the original proposed features of network traffic flows that proposed by [8]-[15] and the QoS parameter features added to the ANN, SVM and RF models, respectively. We can see that under normal network conditions, the various QoS parameters of the network do not change too much, and the overall performance indexes tend to be stable. Adding additional QoS parameters as classification features at this condition does not improve the classification accuracy obviously. However, under poor network conditions, the QoS performance parameters of the network fluctuate greatly, then the additional QoS parameter features can significantly improve the classification accuracy.

V. Conclusions

In this paper, we took a deep insight into the changes of network dataflow behavior under different network QoS metrics, and it was found that network QoS metrics significantly change the dataflow pattern, which have a large influence on the output of DFI-based network traffic classification methods. Our studies demonstrated a regretful result that many existing DFI-based classification methods may fail for a changed network of QoS metrics, which caused by the inconsistence of the training dataset and test dataset that collected under different network performance. To address this problem, we added some network key performance indexes (KPIs) including transmission delay, packet loss rate, flow throughput and jitter, as additional input features of machine learning model, experiment results showed that comparing with the "pure" DFI-based classification method that considering no network KPIs as a part of the machine learning model input features, our approach can significantly improve the classification robustness under unstable network situation. Future work can be focused on the inherent correlation between the dataflow behavior of retransmission packets and network KPIs, and use the flow behavior of retransmission packets instead of network KPIs as the additional input features of machine learning model.

TABLE VI: We Compared the Classification Accuracy Difference between the Features Proposed in [8–15] and the Additional Network QoS Features Added under both Normal Network Conditions and Poor Network Conditions. The Legends "ANN", "SVM" and "RF" Mean that Using the Original Proposed Features of those References to the Models of Artificial Neural Network, Support Vector Machine and Random Forest, while "ANN-QoS", "SVM-QoS" and "RF-QoS" Mean Adding Additional Four QoS Parameters (Time Delay, Loss Rate, Throughput and Network Jitter) to the Feature Sets

Main features from	Normal network conditions	poor network conditions
[8]-[11]		
[12]		



Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Yongcheng Zhou conducted the research; Anguo Zhang analyzed the data and wrote the paper; all authors had approved the final version.

References

[1] B. Wang, J. Zhang, Z. Zhang, L. Pan, Y. Xiang, and D. Xia. "Noise-resistant statistical traffic classification," *IEEE Transactions on Big Data*, p. 1, 2017.

[2] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, "Blinc: Multilevel traffic classification in the dark," *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 4, pp. 229-240, 2005.

[3] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, "Malware traffic classification using convolutional neural network for representation learning," in *Proc. International Conference on Information Network*, pp. 712-717, 2017.

[4] P. Wang, F. Ye, X. Chen, and Y. Qian, "Datanet: Deep learning based encrypted network traffic classification in SDN home gateway," *IEEE Access*, vol. 6, pp. 55380-55391, 2018.

[5] Z. Wang, "The applications of deep learning on traffic I- dentification," *Black Hat*, 2015.

[6] D. Tong, Y. R. Qu, and V. K. Prasanna, "Accelerating decision tree based traffic classification on FPGA and multi- core platforms," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 11, pp. 3046-3059, 2017.

[7] V. F. Taylor, R. Spolaor, M. Conti, and I. Martinovic, "Robust smartphone App identification via encrypted network traffic analysis," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 1, pp. 63-78, 2018.

[8] J. Zhang, C. Chen, Y. Xiang, W. Zhou, and A. V. Vasilakos. "An effective network traffic classification method with unknown flow detection," *IEEE Transactions on Network and Service Management*, 2013.

[9] J. Zhang, Y. Xiang, Y. Wang, W. Zhou, Y. Xiang, and Y. Guan. "Network traffic classification using correlation information," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 1, pp. 104-117, 2013.

[10] J. Zhang, X. Chen, Y. Xiang, and J. Wu, "Robust network traffic classification," *IEEE/ACM Transactions on Networking*, vol. 23, no. 4, pp. 1257-1270, 2015.

[11] S. Zhao, Y. Zhang, and P. Chang, "Network traffic classification using tri-training based on statistical flow characteristics," *IEEE Trustcom/BigDataSE/ICESS*, 2017.

[12] C. Wang, T. Xu, and X. Qin, "Network traffic classification with improved random forest," in *Proc. International Conference on Computational Intelligence and Security*, 2015.

[13] Z. Fan and R. Liu, "Investigation of machine learning based network traffic classification," *International Symposium on Wireless Communication Systems*, 2017.

[14] 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*, 2017.

[15] X. Zeng, X. Chen, G. Shao, *et. al.*, "Flowcontext and host behavior based shadowsocks's traffic identification," *IEEE Access*, vol. 7, pp. 41017-41032, 2019.

[16] B. Yamansavascilar, M. Amac Guvensan, A. Gokhan Yavuz, and M. E.Karsligil, "Application identification via network traffic classification," *Workshop on Computing, Networking and Communications*, 2017.

[17] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*, New York: Addison-Wesley, 2010, p. 36.

[18] G. Miao, J. Zander, K. W. Sung, and B. Slimane, *Fundamentals of Mobile Data Networks*, Cambridge University Press, 2016.

[19] L. Angrisani, D. Capriglione, L. Ferrigno, and G. Miele, "Measurement of the IP packet delay variation for a reliable estimation of the mean opinion score in VoIP services," in *Proc. 2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, 2016, pp. 1-6.

[20] J. Kampeas, A. Cohen, and O. Gurewitz, "Traffic classification based on zero-length packets," *IEEE Transactions on Network and Service Management*, vol. 15, no. 3, pp. 1049-1062, 2018.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).

Zhou Yongcheng was born in Hunan province, China in 1989. He received M.S. degrees in control engineering from the Chongqing University, Chongqing in 2016.

He was a control engineer at CRRC in 2016. Since 2018, he has been a researcher and senior engineer with the Research Institute of Ruijie, Ruijie Networks Co., Ltd. His research interest includes machine learning, deep learning, cooperation control and applications.

Zhang Anguo was born in Hefei city, Anhui province, China in 1990. He received his bachelor's degree and master's degree in control engineering from Chongqing University, Chongqing in 2012 and 2016, respectively. Since 2018, he has been a researcher and senior engineer in the Research Institute of Ruijie, Ruijie Networks Co., Ltd. He is also pursuing a Ph.D in communication and information systems at Fuzhou University. His research interest includes machine learning, artificial neural networks, control theory and applications.